R.K. Selander

# DNA Sequence Analysis of the Genetic Structure and Evolution of *Salmonella enterica*

## Abstract

Comparative sequence analysis indicates that the effective rate of horizontal transfer and recombination in *S. enterica* is low for most housekeeping and invasion (*inv/spa*) genes. Consequently, the species is able to maintain a basically clonal population structure under which lineages are differentially adapted in host-range, disease specificity, and virulence and individual clones may achieve long-term global distribution. But for the hypervariable flagellin *fliC* locus and the *rfb* genes that determine O-antigen structure, genetic exchange among strains is a major source of allelic diversity.

## Introduction

The basic goal of bacterial population genetics is to elucidate the factors that determine the genetic structure of natural populations and mediate evolutionary change. In application to pathogenic species, the analysis of genetic variation at the molecular level within and among populations has provided high-resolution methods of strain and species discrimination for clinical and epidemiological microbiology, but the primary contribution of population genetics has been to introduce an essential evolutionary dimension to an understanding of the genetic basis of pathogenesis, host-adaptation, and the origin of new pathogenic forms (Selander and Musser, 1990).

Among the pathogenic bacteria, *Salmonella enterica* is an unusually diverse pathogenic species that presents excellent opportunities for studies of the origin, function, and evolutionary elaboration of both structures and processes. This chapter reviews the recent findings of research on the genetic diversity, population structure, and evolutionary relationships of the salmonellae, with emphasis on information obtained by the comparative nucleotide sequencing of housekeeping, invasion, and flagellin genes.

## Subspecific Relationships

For the more than 2,300 serovars of *S. enterica*, seven subspecies are formally recognized on the basis of biochemical variation and genomic DNA hybridiza-

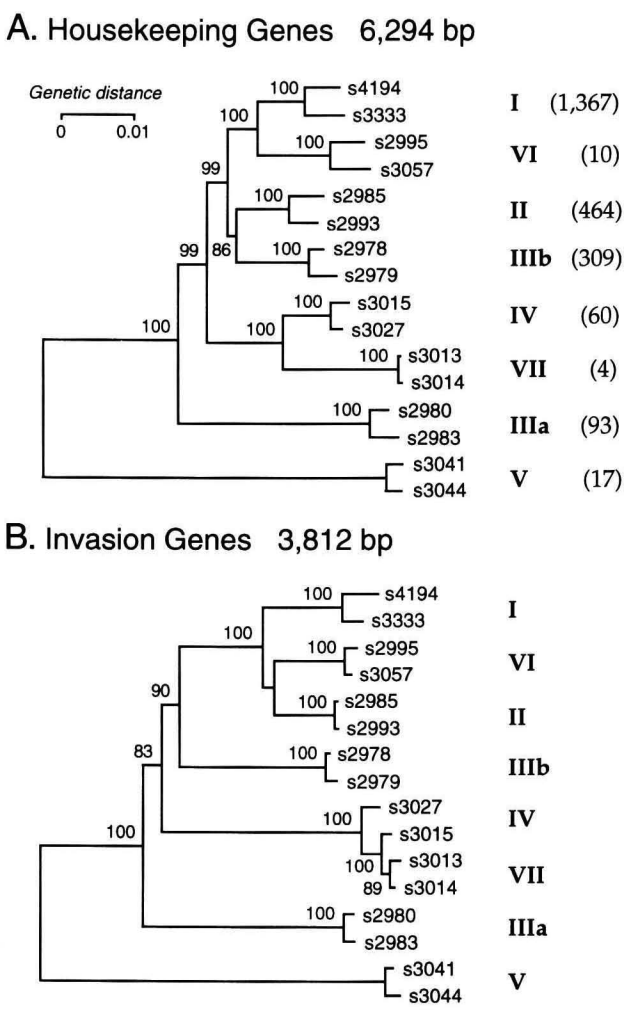


Fig. 1. A. Neighbor-joining tree for 16 strains of the eight subspecies of *S. enterica*, based on variation in the combined coding sequences of five housekeeping genes, *putP*, *gapA*, *mdh*, *gnd*, and *aceK*. The tree was constructed from a matrix of pairwise genetic distance estimated from the number of substitutions per site (Jukes and Cantor, 1969; Nei and Gojobori, 1986). Subspecies are designated by roman numerals, and the number of serovars assigned to each subspecies (Popoff and Le Minor, 1992) is indicated in parentheses. Bootstrap values based on 1,000 computer-generated trees are indicated at the nodes. B. Neighbor-joining trees for four invasion genes, based on variation in the combined coding sequences. [Adapted from Li *et al.* (1995).]

tion (Le Minor *et al.*, 1986; Popoff and Le Minor, 1992). Additionally, an eighth subspecies (designated as VII) has recently been distinguished for certain strains formerly assigned to subspecies IV (Selander *et al.*, 1991, 1995).

Estimates of the genetic relationships among strains of the eight subspecies, as indexed by variation in the combined nucleotide sequences of five housekeeping genes, proline permease (*putP*) (Nelson and Selander, 1992), glyceraldehyde-3-

phosphate dehydrogenase (*gapA*) (Nelson *et al.*, 1991a), malate dehydrogenase (*mdh*) (Boyd *et al.*, 1994), 6-phosphogluconate dehydrogenase (*gnd*) (Nelson and Selander, 1994), and isocitrate dehydrogenase kinase/phosphatase (*aceK*) (K. Nelson, unpublished data) are shown in Fig. 1A. The topology of the tree is consistent with evidence from DNA hybridization and may, therefore, be considered indicative of the actual phylogenetic relationships of the subspecies, notwithstanding the occurrence of a low level of horizontal transfer of gene segments among them (see beyond). That the subspecies I, II, IIIb, and VI - the serovars of which are exclusively or predominantly diphasic in flagellar expression - cluster apart from the monophasic subspecies IIIa, IV, V, and VII suggests the following evolutionary scenario.

Following the divergence of *S. enterica* and *Escherichia coli* from a common ancestor 120–160 million years ago, coincident with the origin of the mammals (Ochman and Wilson, 1987), *E. coli* evolved as a commensal and opportunistic pathogen of mammals and birds. Perhaps 80 million years ago, *E. coli* produced the lineages of the four nominal species of *Shigella* (Tominaga *et al.*, 1994), which, notwithstanding their taxonomic classification, are actually clonal lineages of *E. coli* (Ochman *et al.*, 1983; Karaolis *et al.*, 1994). Meanwhile, the *S. enterica* lineage remained associated with reptiles (which are still the primary hosts of the monophasic subspecies) and evolved as intracellular pathogens through acquisition of the invasion (*inv/spa*) genes and other loci that mediate invasion of host epithelial cells and otherwise distinguish *S. enterica* from *E. coli.* Subsequently, by providing increased ability to circumvent host immune systems, the invention of the mechanism of flagellar antigen phase shifting (diphasic condition) in the *S. enterica* lineage ancestral to subspecies I, II, IIIb, and VI may have been a critical factor permitting an expansion of ecological range to mammals and birds, but as a pathogen rather than a commensal - a niche already long occupied by *E. coli.* Subspecies I became highly specialized for mammals and birds, with some serovars adapting to single host species. Secondarily and inexplicably, given the adaptive advantage that phase shifting would seem to provide, 10% of the serovars of subspecies I and II have reverted to the monophasic condition, usually by loss of expression of phase 2 flagella.

## Housekeeping Genes

In the absence of horizontal genetic exchange among strains, bifurcating evolutionary trees based on variation in nucleotide sequences may be interpreted as estimated phylogenies. But even when recombination has occurred and a strict phylogenetic interpretation is, therefore, precluded, tree construction may be a useful method of presenting information on degrees of genetic distance among strains. And discordant features of the topologies of trees for different loci may provide information on the frequency and extent of genetic exchange, which is a major focus of current research in bacterial population genetics

because of the important effects of recombination on the genetic structure and mode of evolution of natural populations (Maynard Smith *et al.*, 1993; Whittam, 1995).

*Levels of Sequence Diversity*

For a sample of 16 strains representing the eight subspecies of *S. enterica*, information on sequence variation in five genes encoding housekeeping proteins are presented in Fig. 2, together with comparable data for four invasion genes. In the figure, $d_S$ is the estimated mean number of synonymous substitutions per synonymous site ( $\times 100$) that have occurred between the sequences of pairs of strains, and $d_N$ is the comparable estimate for nonsynonymous (replacement) sites. Among the housekeeping genes, *aceK* is unusually polymorphic with respect to both types of changes.

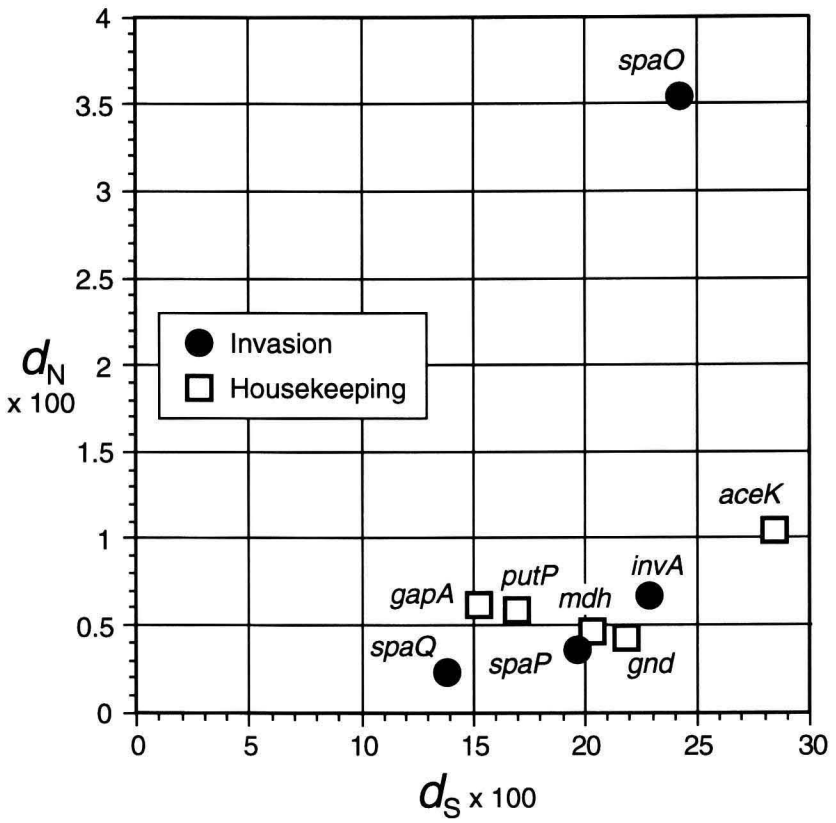


Fig. 2. Sequence variation in nine genes among 16 strains representing all eight subspecies of *S. enterica*. The plotted points indicate mean estimated numbers of substitutions between pairs of strains. See text for definitions of $d_S$ and $d_N$.

*Evidence of Recombination*

Comparisons of individual trees based on the nucleotide sequences of *putP*, *gapA*, *mdh*, and *gnd* (Fig. 3) have revealed several cases in which the order of branching of lineages is discordant. Some of these topological differences are attributable to intragenic recombination events, while others reflect the exchange

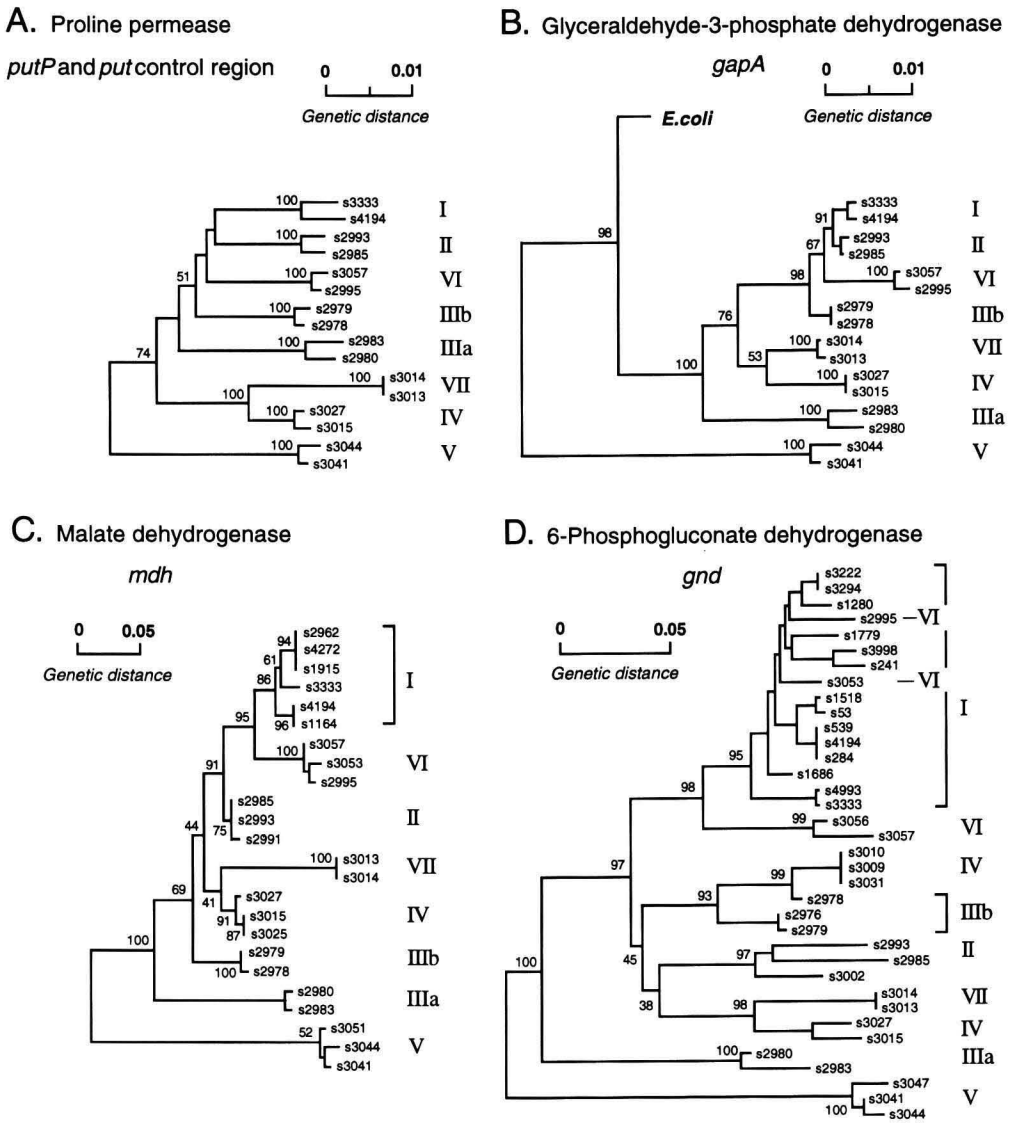


Fig. 3. Neighbor-joining trees for four genes encoding housekeeping proteins, based on variation at synonymous nucleotide sites. Subspecies are indicated by roman numerals; and bootstrap values based on 1,000 computer-generated trees are indicated at the nodes. [Adapted from Boyd *et al.*, (1994), Nelson *et al.* (1991), and Nelson and Selander (1992, 1994).]

of entire genes (assortative recombination). The DNA segments involved in intragenic recombination, which were identified with the guidance of statistical tests for nonrandom clustering of polymorphic sites (Stephens, 1985; Sawyer, 1989), varied in length from 6 to 1,073 base pairs (bp), but several of them are in the 200–400 bp range.

In the *putP*, *mdh*, and *gnd* trees (Fig. 3A, C, and D), subspecies V clusters with the other seven subspecies; but in the *gapA* tree (Fig. 3B), it forms a branch apart from both the other salmonellae and *E. coli* as a consequence of the presence of a segment of the gene that is almost identical in sequence to that of *Klebsiella pneumoniae* (Nelson and Selander, 1992). Except for this feature, the topologies of the *putP* and *gapA* trees are generally similar, with subspecies I, II, IIIb, and VI in the same relationships. But the positions of the branch leading to IV and VII and that of IIIa are reversed in the two trees. This difference in branching order is attributable to the occurrence of a group of 25 unique polymorphic sites that defines a large segment in the central part of the *putP* sequence in strains of subspecies VII.

The only distinctive feature of the topology of the *mdh* tree (Fig. 3C) is that subspecies I is more similar to VI than to II, whereas I and II cluster together in the trees for *putP* and *gapA*. This difference reflects a recombination event, which was detected by the nonrandom clustering of polymorphic sites in the 5′ region of the *mdh* gene.

That segments of substantial length in *putP*, *mdh*, and *gapA* can be clearly identified as recombinant elements in contemporary lineages of *S. enterica* indicates that episodes of intragenic recombination involving these loci are rare. And the simplest explanation for the observation that strains of a given subspecies share the same recombined segments is that these events antedate the time of divergence of the contemporary cell lineages within the subspecies.

Because 6-phosphogluconate dehydrogenase functions in an essential metabolic pathway and the *gnd* locus exhibits a moderate degree of codon bias, the expectation from evolutionary theory is that both nucleotide and amino acid sequences should be relatively conserved. However, in *S. enterica*, the number of allozyme alleles is larger than expected from estimates of total genic diversity (Nelson and Selander, 1994).

For most of 36 strains studied, evolutionary relationships deduced from *gnd* sequences (Fig. 3D) are similar to those indicated by other housekeeping genes. Thus, subspecies V is the most divergent, followed by IIIa. However, assortative recombination has occurred between strains of several subspecies (notably I and VI), and intragenic recombination has also produced positional changes for some strains. For example, the *gnd* sequences indicate a close relationship between the Paratyphi A (s4993) and Typhi (s3333), which is inconsistent with evidence from MLEE (Boyd *et al.*,1993) and the sequences of other genes.

In *E. coli*, recombination at the *gnd* locus—both intragenic and assortative (Bisercic *et al.*, 1991; Dykhuizen and Green, 1991)—has been much more extensive than in *S. enterica* (Nelson and Selander, 1994). Some strains have acquired alleles from species of *Citrobacter and Klebsiella*, and recombination has

occurred so frequently that a tree derived from *gnd* sequences bears little resemblance to other gene trees or to a tree based on the results of multilocus enzyme electrophoresis (MLEE) (Selander *et al.*, 1991).

*Linkage Interactions between* gnd *and rfb*

The most plausible explanation for the unusually high effective rate of recombination in *gnd* is its close linkage to loci that determine the structure of cell-surface macromolecules, including genes of the *rfb* cluster that mediate biosynthesis of the highly antigenic polysaccharide domain (O antigen) of the cell-surface lipopolysaccharide (Bisercic *et al.*, 1991; Dykhuizen and Green, 1991; Nelson and Selander, 1994). Genes of the *rfb* region are believed to be subject to strong frequency dependent selection (Reeves, 1992, 1993); and there is evidence that recombination occurs frequently in the *rfb* region and that some or all of these genes in *S. enterica* and *E. coli* have been recruited from other, distantly related bacteria (Stevenson *et al.*, 1994; Xiang *et al.*, 1993, 1994). The inference is that the frequency of recombination at *gnd* is increased by the action of selection for allelic diversity at neighboring loci, as *gnd* sequences occasionally hitchhike with adaptive recombinants of *rfb* and, perhaps, other genes.

The anomalous position of subspecies IV strains s3010, s3009, and s3031 (Fig. 3D), which are all but identical in *gnd* sequence and also share O antigen 50, may be explained by the transfer of both *gnd* and *rfb* from a strain of IIIb such as s2978, which also expresses O50. And the presence of nearly identical *gnd* and *rfb* sequences in Typhi and Paratyphi A (Nelson and Selander, 1994; Reeves, 1993) clearly indicates co-transfer. For strains of *E. coli*, several examples of the horizontal co-transfer of *rfb* and *gnd* alleles have been identified (Selander *et al.*, 1987; Whittam and Ake, 1993).

Thampapapillai *et al.* (1994) have recently reported the results of an extensive study of sequence diversity in *gnd* among strains of *S. enterica* in which multiple recombination events, some involving co-transfer of parts or all of *gnd* and the *rfb* region, were identified. Several of these events appear to have been mediated by chi-like sequences located near recombination junctions.

## Invasion Genes

The invasion of host cells by *S. enterica* is mediated by a large number of genes that map to several chromosomal locations. Homologues of some of these genes occur in *E. coli* (Groisman and Ochman, 1994); but there is a 40-kb segment near 59 min on the *S. enterica* chromosome that is not present in *E. coli* K-12 and contains 15 or more loci, the *inv/spa* genes, whose products are required for the invasion of epithelial cells (Mills *et al.*, 1995). Homologues of these genes, which apparently are involved in the secretion of antigens that promote cell entry, have been identified in a variety of animal and plant pathogens (Galán *et al.*, 1992; Bergman *et al.*, 1994; Eichelberg *et al.*, 1994).

Sequences of four invasion genes—*invA*, *spaO*, *spaP*, *and spaQ*—have recently been obtained for multiple strains of *S. enterica* (Li *et al.*, 1995). Levels of sequence diversity are shown in Fig. 2. The range of variation in $d_S$ and $d_N$ among the invasion genes exceeds that shown by housekeeping genes. The SpaO protein is hypervariable, with 21% of its amino acid positions polymorphic, which is consistent with evidence that it is an exported antigen. However, unlike the flagellin *fliC* gene (see beyond), the *spaO* gene diversity appears to have been generated almost entirely by point mutation, rather than by intragenic recombination. In contrast, the SpaQ protein is unusually well conserved, with only 2.3% of its amino acids polymorphic. Levels of variation in *invA* and *spaP* are relatively normal, although both $d_S$ and $d_N$ are slightly inflated in *invA* by the presence of a recombinant segment, imported from an unidentified source, in the sequences of subspecies IV and VII.

The topology of a tree constructed from the combined sequences of the four *inv/spa* genes (Fig. 1B) is generally similar to the comparable tree for five housekeeping genes (Fig. 1A), with the exception of the position of subspecies II relative to subspecies IIIb and VI and an absence of substantial differentiation between strains of subspecies IV and VII. Because the overall degree of diversification of the invasion genes among the subspecies of *S. enterica* is roughly equivalent to that of the housekeeping genes, the inference is that they were already present in the ancestral form of the species. But the lack of strong differentiation between subspecies IV and VII points to the occurrence of at least one inter-subspecific exchange of part or all of the invasion gene segment. This and other lines of evidence indicates that the chromosome of subspecies VII is a complex mosaic of large segments, some similar in sequence to those of subspecies IV and others highly distinctive in character.

## Flagellar Filament Genes

### *Recombinational Basis of Serovar Diversity*

The expression of the same flagellin and/or polysaccharide serotypes by distantly related strains, even those belonging to different subspecies (Table 1), theoretically could reflect the retention of alleles from ancestral populations, convergence in epitope structure, or recombination of horizontally transferred phase 1 flagellin (*fliC*), phase 2 flagellin (*fljB*), or *rfb* genes. On the basis of the discovery, by MLEE analysis, that Enteritidis, Derby, Newport, and some other serovars are polyphyletic assemblages of distantly related strains, horizontal transfer and recombination events involving these genes were postulated to be relatively frequent (Beltran *et al.*, 1988; Selander *et al.*, 1990a, b); and for *fliC*, this hypothesis subsequently was supported by partial sequencing of the gene in strains of Typhimurium (Smith and Selander, 1990) and several other serovars (Smith *et al.*, 1990) and by the discovery of a plasmid-borne *fliC*-like gene (*flpA*) in a triphasic strain of a normally diphasic serovar (Smith and Selander, 1991).

Table 1. Distribution of serovars with phase 1 (*fliC*) serotypes of the g complex among the subspecies of *S. enterica*

| Phase 1 serotype[a] | Total no. serovars | No. serovars in indicated subspecies | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | I | II | IIIa | IIIb | IV | V | VI | VII |
| f, g | 14 | 14 | | | | | | | |
| f, g, m, t | 1 | 1 | | | | | | | |
| f, g, s | 4 | 4 | | | | | | | |
| f, g, t | 7 | 6 | 1 | | | | | | |
| g, m | 13 | 13 | | | | | | | |
| g, m, q | 1 | 1 | | | | | | | |
| g, m, [p], s | 1 | 1 | | | | | | | |
| g, m, s | 13 | 12 | 1 | | | | | | |
| g, m, s, t | 30 | 2 | 28 | | | | | | |
| g, m, t | 16 | 7 | 9 | | | | | | |
| g, p | 3 | 3 | | | | | | | |
| g, p, s | 1 | 1 | | | | | | | |
| g, p, u | 1 | 1 | | | | | | | |
| g, q | 1 | 1 | | | | | | | |
| g, s, t | 24 | 20 | 4 | | | | | | |
| g, t | 27 | 5 | 22 | | | | | | |
| g, $z_{51}$ | 35 | 7 | | 16 | | 10 | | | 2 |
| g, $z_{62}$ | 5 | | 5 | | | | | | |
| g, $z_{63}$ | 1 | 1 | | | | | | | |
| m, p, t, [u] | 1 | 1 | | | | | | | |
| m, t | 52 | 23 | 28 | | | | | | 1 |

[Prepared from data in Popoff and Le Minor (1992).]
[a]Antigenic factors in parentheses are not expressed by all isolates.

As a test of the generality of the horizontal transfer/recombination hypothesis, Li *et al.* (1994) obtained the complete *fliC* sequences of 15 strains of several serovars of subspecies I, II, IV, and VII that express seven combinations of six phase 1 flagellar antigenic factors of the g complex (f, g, m, s, t, and $z_{51}$). In *S. enterica*, as in other bacteria (Wilson and Beveridge, 1993), the terminal regions of the flagellin molecule (C1 and C2), which are involved in secretion and polymerization, are strongly conserved in both length and amino acid sequence, whereas the central region (V), which is the site of the epitopic variation assayed in serotyping (Joys, 1988; Newton *et al.*, 1991), is hypervariable.

Individual evolutionary trees based on MLEE (indexing the overall genomic relatedness of the strains), the nucleotide sequence of the combined C1 and C2 regions of *fliC*, and the sequence of the V region of the gene are shown in Fig. 4. If the evolution of *fliC* has involved little or no recombination, trees for the V region and the C1 + C2 regions should be topologically concordant. In contrast, horizontal exchange of the V region among strains would be indicated by a clustering of sequences specifying the same flagellin serotype, regardless of the overall genetic relatedness of the strains in which they occur.

Although strains of Enteritidis (En 1) and Othmarschen (Ot 1), which express serotype g,m, are divergent in chromosomal character (Fig. 4A), their *fliC*

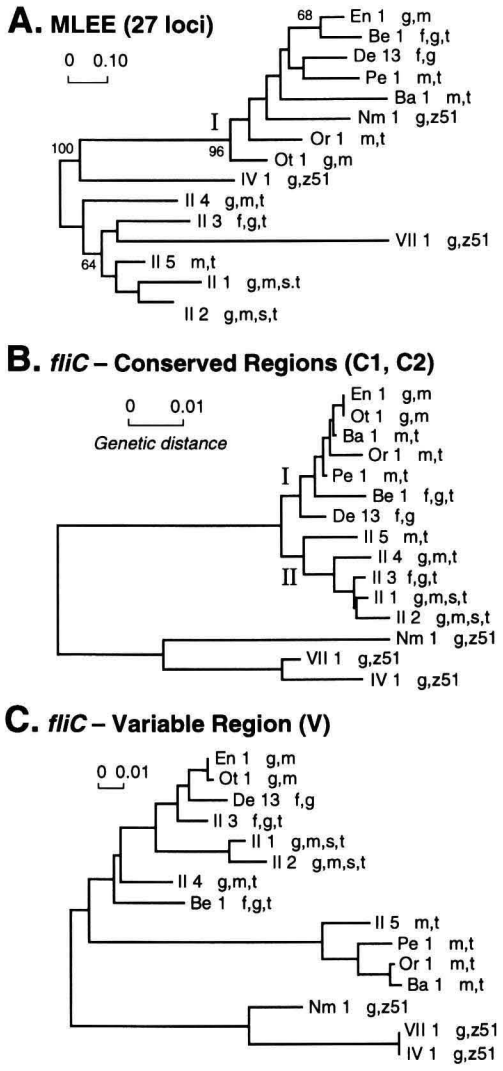


Fig. 4. Evolutionary trees for 15 strains of *S. enterica*. For each strain, the phase 1 flagellin serotype is indicated. A. MLEE tree (based on 27 enzyme loci). Bootstrap values greater than 50% are shown at the nodes. B. Conserved (C1 + C2) regions of the *fliC* gene. Clusters of ETs of subspecies I and II are labeled. Bootstrap values for all nodes are greater than 50%. C. Central variable (V) region of the *fliC* gene. Note that the scale is approximately half that of the conserved-sequence tree (B). Subspecies I serovar abbreviations are Ba, Banana; Be, Berta; De, Derby; En, Enteritidis; Nm, Newmexico; Or, Oranienburg; Ot, Othmarschen; and Pe, Pensacola. Roman numerals identify ETs of other subspecies. [Adapted from Li *et al.* (1994).]

sequences are nearly identical (Fig. 4B and C). Strains Nm 1, IV 1, and VII 1 (serotype g,$z_{51}$) represent three subspecies, but both their V and C1 + C2 sequences cluster together. ETs Pe 1, Ba 1, Or 1, and II 5 are of serotype m,t. The three subspecies I ETs are not closely allied, and all three are distantly related to II 5. Their C1 + C2 sequences reflect these relationships, but all four ETs form a distinct, tight cluster in the V region tree. This region and part of C2 have been exchanged among the four ETs; and part of this distinctive sequence has also been transferred to ETs Be 1 and II 4.

In sum, the occurrence of each of three flagellin serotypes of the g complex in distantly related strains is clearly attributable to horizontal exchange rather than to convergence in epitope-determining amino acid sequences or retention of ancestral sequences. If antigens of this complex are typical of flagellar antigens in general—and there is no reason to believe they are not—interstrain exchange and recombination is clearly a major evolutionary mechanism generating both allelic variation in *fliC* and serovar diversity in natural populations of the salmonellae.

*Evolutionary Loss of Flagella*

The serovars Gallinarum and Pullorum, which are invasive pathogens of fowl, are unique among the salmonellae in being obligatorily nonflagellate. MLEE analysis has identified Enteritidis as a close relative, and the sharing of distinctive allozyme alleles at three metabolic enzyme loci and a premature stop codon in the *flgK* gene encoding the first hook-filament junction protein by the ETs of Gallinarum and Pullorum indicates that they are monophyletic and that their most recent common ancestor was nonmotile (Li *et al.*, 1993).

Li *et al.* (1993) found that intact, but silenced, *fliC* genes are present in the nonmotile salmonellae. The genes of strains of some ETs of Gallinarum are identical in sequence to that of Enteritidis, but a nucleotide substitution has created a shared premature stop codon in strains of several ETs. The *fliC* sequences of the Pullorum ETs differ from the standard Enteritidis and Gallinarum sequence in having nonsynonymous substitutions in two or three codons in the V region.

These findings indicate that loss of motility in the avian-adapted salmonellae occurred relatively recently in evolutionary time. Flagella are not expressed *in vivo* by strains of Enteritidis infecting chickens (Chart *et al.*, 1993), and their permanent loss presumably evolved in the Enteritidis-like common ancestor of Gallinarum and Pullorum because motility was for some reason no longer required as the lineage became increasingly restricted in host-range to fowl. Nonsense mutations in *fliC* and *flgK* then began to accumulate when, as a consequence of a mutation in another gene of the flagellar regulon that inhibited flagellar biosynthesis, they were no longer active and subject to purifying selection.

Many isolates of Dublin recovered in recent years from invasive infections in cattle have been nonmotile, although the condition apparently is reversible

(Selander *et al.*, 1992). Like Gallinarum and Pullorum, Dublin is strongly host-adapted and is also a close relative of Enteritidis. It is also of interest that the strongly host-adapted shigellae, as well as most enteroinvasive strains of *E. coli.* are nonmotile, although at least two of the four *Shigella* species carry intact *fliC* genes (Tominaga *et al.*, 1994). Modulation of flagellar biosynthesis may be common strategy of invasive bacterial pathogens, with motility being advantageous under certain environmental conditions (C. Li *et al.*, 1993; Mahenthiralingam *et al.*, 1994) and during colonization but disadvantageous in later stages of infection.

## Genetic Structure Of Populations

### *The Concept of Clonal Structure*

For the salmonellae, a basically clonal population structure is evidenced by the presence of strong linkage disequilibrium among alleles at enzyme loci, the association of specific O and H serotypes with only one or a small number of multilocus enzyme genotypes, and the global distribution of certain genotypes (Beltran *et al.*, 1988; Reeves *et al.*, 1989; Maynard Smith *et al.*, 1992; Selander *et al.*, 1990a,b, 1992). *S. enterica* populations are clonal in the sense that the effective (realized) rates of recombination for most chromosomal genes are sufficiently low to permit the mutational diversification of cell lineages in terms of biochemical characteristics and ecological niche relationships, including host distribution, disease specificity, and virulence, and the long-term, if not permanent, maintenance of differentially adapted, widely distributed chromosomal genotypes in populations (Selander *et al.*, 1995). Clonality explains why serotyping is a powerful marker system for recognizing groups of strains with distinct host ranges and pathogenicities, including those of closely related populations such as Dublin and Enteritidis (Selander *et al.*, 1992). At the level of population structure resolved by MLEE, most of the serovars, including most of the common pathogens of humans and domesticated animals, are single ETs or families of closely related ETs; and for many of the medically important serovars, only a single ET is globally predominant at any one time (Table 2).

An evolutionary genetic rational for the prevalence of clonal population structure among pathogenic bacteria has been noted by Falkow (1990). Because pathogenesis is a complex, multifactorial process involving the coordinated action of a large number of virulence-associated loci and genes that permit the pathogen to survive in a variety of habitats, including those that enable it to recognize its hosts and then to avoid, subvert, or nullify their defense systems, there could be little adaptive radiation among pathogenic bacteria if frequent and indiscriminate horizontal genetic transfer and recombination occurred for genes throughout the genome.

Table 2. Relative abundance of the commonest ET of *S. enterica* serovars

| Serovar | No. of isolates | No. of ETs | Commonest ET | % of isolates of commonest ET |
|---|---|---|---|---|
| Choleraesuis | 85 | 6 | Cs 1 | 88 |
| Derby[a] | | | | |
| I | 71 | 2 | De 1 | 93 |
| II | 267 | 3 | De 13 | 61 |
| III | 11 | 1 | De 31 | 100 |
| Dublin | 117 | 3 | Du 1 | 95 |
| Enteritidis | 257 | 14 | En 1 | 93 |
| Gallinarum | 56 | 5 | Ga 2 | 95 |
| Heidelberg | 204 | 8 | He | 87 |
| Infantis | 113 | 4 | In 1 | 96 |
| Newport[a] | | | | |
| I | 72 | 7 | Np 1 | 38 |
| II | 32 | 5 | Np 11 | 84 |
| Panama | 94 | 11 | Pn 1 | 77 |
| Paratyphi A | 135 | 6 | Pa 1 | 74 |
| Paratyphi B | 123 | 13 | Pb 1 | 61 |
| Paratyphi C | 100 | 9 | Pc 1 | 60 |
| Pullorum | 75 | 7 | Pu 3 | 51 |
| Typhi | 334 | 7 | Tp 1 | 82 |
| Typhimurium | 299 | 17 | Tm 1 | 83 |

[Sources of data: Beltran *et al.* (1988), Selander *et al.* (1990a, b, 1992), and Li *et al.* (1993).]
[a]Polyphyletic serovars; see Beltran *et al.* (1988).

### *Clones of Serovar Typhi*

Clonal aspects of the genetic structure of populations of *S. enterica* are well illustrated by the serovar Typhi, which is the agent of human typhoid fever. Both biochemically and genotypically, Typhi is an unusually distinctive and homogenous serovar (Faundez *et al.*, 1990; Franco *et al.*, 1992; Moshitch *et al.*, 1992; Thelfall *et al.*, 1994; Thong *et al.*, 1994). By MLEE analysis, Selander *et al.* (1990b) found that 82% of worldwide isolates were of one genotype (ET 1), with a second clone (ET 2) being represented by 16% of strains, all from western Africa.

Throughout most of the world, Typhi is monomorphic for the d allele at the *fliC* locus encoding the phase 1 flagellin protein. But Indonesian populations are polymorphic for the d allele and a variant j allele (Maher *et al.*, 1986; Franco *et al.*, 1992), which was derived from d by the deletion of a 261-bp segment in the central part of the *fliC* gene (Frankel *et al.*, 1989). Additionally, some Indonesian strains of either phase 1 d or j uniquely express a $z_{66}$ flagellar antigen (Guinée *et al.*, 1981), which presumably is encoded by a phase 2 locus, although they lack the genes that normally mediate phase shifting in *S. enterica* (Moshitch *et al.*, 1992). Inasmuch as the specific deletion that changes d to j may be experimentally produced when Typhi is grown in anti-d serum (Smith *et al.*, 1991), it is likely that it occurs repeatedly in natural populations, where, however, it may normally be disadvantageous because motility is decreased

(Frankel *et al.*, 1989). If so, the problem is to explain why it has managed to reach a frequency as high as 15% in the Indonesian Typhi population (Grossman *et al.*, 1994). An explanation is suggested by the fact that flagellar antigens are the major factors eliciting immune responses after infection with Typhi or vaccine administration (He *et al.*, 1994).

It is probably not coincidental that the frequency of typhoid fever in Indonesia is the highest in the world, with as many as 1,300,000 cases and 20,000 deaths per year (Simanjuntak *et al.*, 1992). Moreover, it is primarily a disease of children, which suggests that the Indonesian population acquires immunity through exposure (Sudarmono and Radji, 1992). Because immunity is serotype-specific, people infected with d allele strains would remain relatively unprotected against infection by strains carrying the j allele, thus accounting for the decrease in the ratio of d/j infections with age (Grossman *et al.*, 1994). Flagellar antigen diversity in Indonesia may well be an example of the maintenance of polymorphism by balancing selection for avoidance of immune responses in the host population.

*Inter-locus Variation in Recombination Rate*

A generalization emerging from the analysis of gene sequences in populations of *S. enterica* and other bacteria is that the contribution to allelic diversity made by recombination varies markedly among loci encoding proteins of different functional types (Achtman, 1994; Nelson *et al.*, 1991a; Nelson and Selander, 1992; Sibold *et al.*, 1992; Selander *et al.*, 1995; Vásquez *et al.*, 1993), as well as among species and subdivisions of species. As shown in this review, the effective rate of recombination in *S. enterica* is low for genes encoding most metabolic enzymes and other types of housekeeping proteins, as well as for the *inv/spa* virulence genes. In contrast, horizontal transfer and recombination is a major source of allelic diversity for both the highly polymorphic flagellin *fliC* locus and the *rfb* genes that determine O-antigen structure. Entire *fliC* genes and parts or all of the epitope-determining central region have been frequently exchanged within and between subspecies, and many alleles are mosaics of segments derived from several sources. Because flagellin is a highly antigenic and interacts directly with the external environment, recombinant alleles may confer an immediate adaptive advantage to a bacterial cell and, in consequence, be brought to high frequency in local populations by natural selection and then transferred to other lineages. The prevailing view is that the extensive flagellar antigenic polymorphism in *S. enterica* is adaptive in permitting the reinfection of hosts (Brunham *et al.*, 1993); and Reeves (1992) has suggested that genes mediating antigenic variation in flagellin and the cell-surface polysaccharide are subject to niche-specific selection. The observation that sensitivity to flagellotropic bacteriophages may be serotype dependent (Iino, 1977) suggests another possible adaptive basis for flagellin polymorphism; and Skurnik and Toivanen (1994) have proposed that resistance to bacteriophages is also an adaptive basis for O antigen polymorphism.

Table 3. Proteins and polysaccharides for which there is evidence that the encoding or mediating genes are subject to horizontal transfer and intragenic or assortative recombination for which an adaptive basis has been suggested

| Protein or structure (gene) | Species | References |
|---|---|---|
| Class 1 OMP (*porA*) | *Neisseria meningitidis* | Feavers *et al.* (1992) |
| Opacity proteins (*opa* family) | *Neisseria meningitidis* | Hobbs *et al.* (1994) |
| M proteins (*emm* cluster) | *Streptococcus pyogenes* | Harbaugh *et al.* (1993); Whatmore and Kehoe (1994); Bessen and Hollingshead (1994) |
| Pilin (*pilE* and *pilS*) | *Neisseria gonorrhoeae* | Seifert and So (1988); Seifert *et al.* (1988) |
| Pili (*pap* and *prs* clusters) | *Escherichia coli* | Plos *et al.* (1989); Arthur *et al.* (1990); Marklund *et al.* (1992) |
| Capsular polysaccharide (*cap*) | *Haemophilus influenzae* | Kroll and Moxon (1990) |
| IgA protease (*iga*) | *Neisseria gonorrhoeae*<br>*Neisseria meningitidis*<br>*Haemophilus influenzae* | Halter *et al.* (1989)<br>Lomholt *et al.* (1992); Morelli *et al.* (1994)<br>Poulsen *et al.* (1992) |
| Penicillin-binding proteins | *Neisseria* spp.<br>*Streptococcus* spp. | Spratt *et al.* (1992); Bowler *et al.* (1994)<br>Dowson *et al.* (1990); Laible *et al.* (1991); Martin *et al.* (1992); Sibold *et al.* (1992) |
| O antigen (*rfb* cluster) | *Salmonella enterica* | Reeves (1993); Xiang *et al.* (1993, 1994) |
| Flagellin (*fliC*) | *Salmonella enterica* | Smith *et al.* (1990); Li *et al.* (1994) |
| R-M enzymes (*hsd*)[a] | *Escherichia coli* | Sharp *et al.* (1992); Murray *et al.* (1993) |
| Colicin immunity proteins | *Escherichia coli* | Riley (1993) |
| Pyrogenic exotoxins (*speA*, *speC*) | *Streptococcus pyogenes* | Nelson *et al.* (1991b); Kapur *et al.* (1992); Reda *et al.* (1994) |

[Modified from Selander *et al.* (1995).]
[a]R-M, restriction and modification system.

Similar explanations in terms of environmental adaptation apply to a number of genes in other bacteria for which evidence of frequent horizontal transfer and recombination is available (Table 3). Such genes almost without exception encode or mediate the expression of products for which there would seem to be a premium on structural diversity or which confer adaptive traits such as antibiotic resistance (Selander *et al.*, 1995). But for housekeeping or virulence genes that encode polypeptides for which there is no premium on diversity in amino acid sequence per se, it is unlikely that either intragenic or assortative recombination would result in a selective advantage to the recipient cell. And the probable fate of deleterious or selectively neutral recombinants is loss from the population through purifying selection and genetic drift. By comparing the degree of amino acid divergence in 179 homologous proteins in *S. enterica* and *E. coli*, Whittam (1995) has shown that cell-surface proteins, such as flagellins, porins, and pilins, are evolving at nearly three times the rate of cytoplasmic enzymes and proteins that function in transport, DNA replication, and the regulation of transcription, and 10 times faster than ribosomal proteins.

Among the double-stranded DNA bacteriophages of enteric bacteria, recombination, mediated by site-specific invertases, has been a major factor in the evolution of tail-fiber genes, with relatively frequent exchange occurring between otherwise unrelated phage groups (quasispecies) (Sandmeier, 1994). Tail-fiber genes of various phages and certain defective prophages share homologous segments, and individual loci may have a mosaic structure. This locus-specific exchange is adaptive in generating diversity in host-range determinants in the face of strong selective pressure on phages to adapt to mutated host surfaces, which, in turn, are under selective pressure to avoid phage infection. Given the great variation in rate of recombination observed among the genes of *S. enterica* and other bacteria, it would not be surprising to find comparable mechanisms that function to increase the frequency of exchange of certain genes or segments of the chromosome for which selection for diversity is strong.

## Acknowledgments

Research leading to this review was supported by Grant AI22144 from the National Institutes of Health.

## References

Achtman, M., 1994 - Clonal spread of serogroup A meningococci: a paradigm for the analysis of microevolution in bacteria. *Mol. Microbiol.* **11**, 15–22.

Arthur, M., R.D. Arbeit, C. Kim, P. Beltran, H. Crowe, S. Steinbach, C. Campanelli, R.A. Wilson, R.K. Selander and R. Goldstein, 1990 - Restriction fragment length polymorphisms among uropathogenic *Escherichia coli* isolates: *pap*-related sequences compared with *rrn* operons. *Infect. Immun.* **58**, 471–479.

Beltran, P., J.M. Musser, R. Helmuth, J.J. Farmer III, W.M. Frerichs, I.K. Wachsmuth, K. Ferris, A.C. McWhorter, J.G. Wells, A. Cravioto and R.K. Selander, 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. heidelbekrg*, *S. infantis*, *S. newport*, *and S. typhimurium*. *Proc. Natl. Acad. Sci. USA* **85**, 7753–7757.

Bergman, T., K. Erickson, E. Galyov, C. Persson and H. Wolf-Watz, 1994 - The *lcrB* (*yscN/U*)) gene cluster of *Yersinia pseudotuberculosis* is involved in Yop secretion and shows high homology to the *spa* gene clusters of *Shigella flexneri* and *Salmonella typhimurium*. *J. Bacteriol.* **176**, 2619–2626.

Bessen, D.E. and S.K. Hollingshead, 1994 - Allelic polymorphism of *emm* loci provides evidence for horizontal gene spread in group A streptococci. *Proc. Natl. Acad. Sci. USA* **91**, 3280–3284.

Bisercic, M., J.Y. Feutrier and P.R. Reeves, 1991 - Nucleotide sequences of the *gnd* genes from nine natural isolates of *Escherichia coli*: evidence of intragenic recombination as a contributing factor in the evolution of the polymorphic *gnd* locus. *J. Bacteriol.* **173**, 3894–3900.

Bowler, L.D., Q.-Y. Zhang, J.-Y. Riou and B.G. Spratt, 1994 - Interspecies recombination between the *penA* genes of *Neisseria meningitidis* and commensal *Neisseria* species during the emergence of penicillin resistance in *N. meningitidis*: natural events and laboratory simulation. *J. Bacteriol.* **176**, 333–337.

Boyd, E.F., K. Nelson, F.-S. Wang, T.S. Whittam and R.K. Selander, 1994 - Molecular genetic basis of allelic polymorphism in malate dehydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*. *Proc. Natl. Acad. Sci. USA* **91**, 1280–1284.

Boyd, E.F., F.-S. Wang, P. Beltran, S.A. Plock, K. Nelson and R.K. Selander, 1993 - *Salmonella* reference collection B (SARB): strains of 37 serovars of subspecies I. *J. Gen. Microbiol.* **139**, 1125–1132.

Brunham, R.C., F.A. Plummer and R.S. Stephens, 1993 - Bacterial antigenic variation, host immune response, and pathogen-host coevolution. *Infect. Immun.* **61**, 2273–2276.

Chart, H., D. Conway and B. Rowe, 1993 - Outer membrane characteristics of *Salmonella enteritidis* phage type-4 growing in chickens. *Epidemiol. Infect.* **111**, 449–454.

Dowson, C.G., A. Hutchison, N. Woodford, A.P. Johnson, R.C. George and B.G. Spratt, 1990 - Penicillin-resistant viridans streptococci have obtained altered penicillin-binding protein genes from penicillin-resistant strains of *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. USA* **87**, 5858–5862.

Dykhuizen, D.E. and L. Green, 1991 - Recombination in *Escherichia coli* and the definition of biological species. *J. Bacteriol.* **173**, 7257–7268.

Eichelberg, K., C.C. Ginocchio and J.E. Galán, 1994 - Molecular and functional characterization of the *Salmonella typhimurium* invasion genes *invB* and *invC*: homology of InvC to the $F_0F_1$ ATPase family of proteins. *J. Bacteriol.* **176**, 4501–4510.

Falkow, S., 1990 - The 'Zen' of bacterial pathogenicity, p. 3–9. *In* B. H. Iglewski and V.L. Clark (ed.), Molecular basis of bacterial pathogenesis. Academic Press, San Diego.

Faundez, G., L. Aron and F.C. Cabello, 1990 - Chromosomal DNA, iron-transport systems, outer membrane proteins, and enterotoxin (heat labile) production in *Salmonella typhi* strains. *J. Clin. Microbiol.* **28**, 894–897.

Feavers, I.M., A.B. Heath, J.A. Bygraves and M.C.J. Maiden, 1992 - Role of horizontal genetic exchange in the antigenic variation of the class 1 outer membrane protein of *Neisseria meningitidis*. *Mol. Microbiol.* **6**, 489–495.

Franco, A., C. Gonzalez, O.S. Levine, R. Lagos, R.H. Hall, S.L. Hoffman, M.A. Moechtar, E. Gotuzzo, M.L. Levine, D.M. Hone and J.G. Morris Jr., 1992 - Further consideration of the clonal nature of *Salmonella typhi*: evaluation of molecular and clinical characteristics of strains from Indonesia and Peru. *J. Clin. Microbiol.* **30**, 2187–2190.

Frankel, G., S.M.C. Newton, G.K. Schooknik and B.A.D. Stocker, 1989 - Intragenic recombination in a flagellin gene: characterization of the *H1*:j gene of *Salmonella typhi*. *EMBO J.* **8**, 3149–3152.

Galán, J.E., C. Ginocchio and P. Costas, 1992 - Molecular and functional characterization of the *Salmonella* invasion gene *invA*: homology of InvA to members of a new protein family. *J. Bacteriol.* **174**, 4338–4349.

Groisman, E.A. and H. Ochman, 1994 - How to become a pathogen. *Trends Microbiol.* **2**, 289–294.

Grossman, D.A., N.D. Witham, D.H. Burr, M. Lesmana, F.A. Rubin, G.K. Schoolnik and J. Parsonnet, 1994 - Flagellar serotypes of *Salmonella typhi* in Indonesia: relationship among mobility, invasiveness, and clinical illness. *J. Infect. Dis.* **171**, 212–216.

Guinée, P.A.M., W.H. Jensen, H.M.E. Maas, L. Le Minor and R. Beaud, 1981 - An unusual H antigen ($z_{66}$) in strains of *Salmonella typhi*. *Ann. Microbiol. (Inst. Pasteur)* **132A**, 331–334.

Halter, R., J. Pohlner and T.F. Meyer, 1989 - Mosaic-like organization of IgA protease genes in *Neisseria gonorrhoeae* generated by horizontal genetic exchange *in vivo*. *EMBO J.* **9**, 2737–2744.

Harbaugh, M.P., A. Podbielski, S. Hügl and P.P. Cleary, 1993 - Nucleotide substitutions and small-scale insertion produce size and antigenic variation in group A streptococcal M1 protein. *Mol. Microbiol.* **8**, 981–991.

He, X.-S., M. Rivkina, B.A.D. Stocker and W.S. Robinson, 1994 - Hypervariable region IV of *Salmonella* gene *flic*$^{d}$ encodes a dominant surface epitope and a stabilizing factor for functional flagella. *J. Bacteriol.* **176**, 2406–2414.

Hobbs, M.M., A. Seiler, M. Achtman and J.G. Cannon, 1994 - Microevolution within a clonal population of pathogenic bacteria - recombination, gene duplication and horizontal genetic exchange in the *opa* gene family of *Neisseria meningitidis*. *Mol. Microbiol.* **12**, 171–180.

Iino, T., 1977 - Genetics of structure and function of bacterial flagella. *Annu. Rev. Genet.* **11**, 161–182.

Joys, T.M., 1988 - The flagellar filament protein. *Can. J. Microbiol.* **34**, 452–458.
Jukes, T.H. and C.R. Cantor, 1969 - Evolution of protein molecules, p. 21–132. *In* Munro, H.N. (ed.), Mammalian protein metabolism. Academic Press, New York.
Kapur, V., K. Nelson, P.M. Schlievert, R.K. Selander and J.M. Musser, 1992 - Molecular population genetic evidence of horizontal spread of two alleles of the pyrogenic exotoxin C gene (*speC*) among pathogenic clones of *Streptococcus pyogenes*. *Infect. Immun.* **60**, 3513–3517.
Karaolis, D.K.R., R. Lan and P.R. Reeves, 1994 - Sequence variation in *Shigella sonnei*, a pathogenic clone of *Escherichia coli*, over four continents and 41 years. *J. Clin. Microbiol.* **32**, 796–802.
Kroll, J.S. and E.R. Moxon, 1990 - Capsulation in distantly related strains of *Haemophilus influenzae* type b: genetic drift and gene transfer at the capsulation locus. *J. Bacteriol.* **172**, 1374–1379.
Laible, G., B.G. Spratt and R. Hakenbeck, 1991 - Interspecies recombinational events during the evolution of altered PBP 2x genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. *Mol. Microbiol.* **5**, 1993–2002.
Le Minor, L., M.Y. Popoff, B. Laurent and D. Hermant, 1986 - Individualisation d'une septième sous-espèce de *Salmonella*: *S. choleraesuis* subsp. *indica* subsp. *nov. Ann. Inst. Pasteur/Microbiol.* **137B**, 211–217.
Li, C., C.J. Louise, W. Shi and J. Adler, 1993 - Adverse conditions which cause lack of flagella in *Escherichia coli*. *J. Bacteriol.* **175**, 2229–2235.
Li, J., N.H. Smith, K. Nelson, P.B. Crichton, D.C. Old, T.S. Whittam and R.K. Selander, 1993 - Evolutionary origin and radiation of the avian-adapted nonmotile salmonellae. *J. Med. Microbiol.* **38**, 129–139.
Li, J., K. Nelson, A.C. McWhorter, T.S. Whittam and R.K. Selander, 1994 - Recombinational basis of serovar diversity in *Salmonella enterica*. *Proc. Natl. Acad. Sci. USA* **91**, 2552–2556.
Li, J., H. Ochman, E.A. Groisman, E.F. Boyd, K. Nelson and R.K. Selander, 1995 - Relationship between evolutionary rate and cellular location among the Inv/Spa invasion proteins of *Salmonella enterica*. *Proc. Natl. Acad. Sci. USA* **92**, 7252–7256.
Lomholt, H., K. Poulsen, D.A. Caugant and M. Kilian, 1992 - Molecular polymorphism and epidemiology of *Neisseria meningitidis* immunoglobulin A1 proteases. *Proc. Natl. Acad. Sci. USA* **89**, 2120–2124.
Mahenthiralingam, E., M.E. Campbell and D.P. Speert, 1994 - Nonmotility and phagocytic resistance of *Pseudomonas aeruginosa* isolates from chronically colonized patients with cystic fibrosis. *Infect. Immun.* **62**, 596–605.
Maher, K.O., J.G. Morris Jr., E. Gotuzzo, C. Ferreccio, L.R. Ward, L. Benavente, R.E. Black, B. Rowe and M.M. Levine, 1986 - Molecular techniques in the study of *Salmonella typhi* in epidemiologic studies in endemic areas: comparison with Vi phage typing. *Am. J. Trop. Med. Hyg.* **35**, 831–835.
Marklund, B.-I., J.M. Tennent, E. Garcia, A. Hamers, M. Båga, F. Lindberg, W. Gaastra and S. Normark, 1992 - Horizontal gene transfer of the *Escherichia coli pap* and *prs* pili operons as a mechanism for the development of tissue-specific adhesive properties. *Mol. Microbiol.* **6**, 2225–2242.

Martin, C., C. Sibold and R. Hakenbeck, 1992 - Relatedness of penicillin-binding protein 1a genes from different clones of penicillin-resistant *Streptococcus pneumoniae* isolated in South Africa and Spain. *EMBO J.* **11**, 3831–3836.

Maynard Smith, J., N.H. Smith, M. O'Rourke and B.G. Spratt, 1993 - How clonal are bacteria? *Proc. Natl. Acad. Sci. USA* **90**, 4384–4388.

Mills, D.M., V. Bajaj and C.A. Lee, 1995 - A 40 kilobase chromosomal fragment encoding *Salmonella typhimurium* invasion genes is absent from the corresponding region of the *Escherichia coli* K-12 chromosome. *Mol. Microbiol.* **15**, 749–759.

Morelli, G., J. del Valle, C.J. Lammel, J. Pohlner, K. Müller, M. Blake, G.F. Brooks, T.F. Meyer, B. Koumaré, N. Brieske and M. Achtman, 1994 - Immunogenicity and evolutionary variability of epitopes within IgA1 protease from serogroup A *Neisseria meningitidis. Mol. Microbiol.* **11**, 175–187.

Moshitch, S., L. Doll, B.-Z. Rubinfeld, B.A.D. Stocker, G.K. Schoolnik, Y. Gafni and G. Frankel, 1992 - Mono- and bi-phasic *Salmonella typhi*: genetic homogeneity and distinguishing characteristics. *Mol. Microbiol.* **6**, 2589–2597.

Murray, N.E., A.S. Daniel, G.M. Cowan and P.M. Sharp, 1993 - Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes. *Mol. Microbiol.* **9**, 133–143.

Nei, M. and T. Gojobori, 1986 - Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426.

Nelson, K. and R.K. Selander, 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.* **174**, 6886–6895.

Nelson, K. and R.K. Selander, 1994 - Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. *Proc. Natl. Acad. Sci. USA* **91**, 10227–10231.

Nelson, K., T.S. Whittam and R.K. Selander, 1991a - 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.

Nelson, K., P.M. Schlievert, R.K. Selander and J.M. Musser, 1991b - Characterization and clonal distribution of four alleles of the *speA* gene encoding pyrogenic exotoxin A (scarlet fever toxin) in *Streptococcus pyogenes. J. Exp. Med.* **174**, 1271–1274.

Newton, S.M.C., R.D. Wasley, A. Wilson, L.T. Rosenberg, J.F. Miller and B.A.D. Stocker, 1991 - Segment IV of a *Salmonella* flagellin gene specifies flagellar antigen epitopes. *Mol. Microbiol.* **5**, 419–425.

Ochman, H. and A.C. Wilson, 1987 - Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* **26**, 74–86.

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.

Plos, K., S.I. Hull, R.A. Hull, B.R. Levin, I. Ørskov, F. Ørskov and C. Svandborg-Edén, 1989 - Distribution of the P-associated-pilus (*pap*) region among *Escherichia coli* from natural sources: evidence for horizontal gene transfer. *Infect. Immun.* **57**, 1604–1611.

Popoff, M.Y. and L. Le Minor, 1992 - Antigenic formulas of the *Salmonella* serovars, 6th ed. WHO Collaborating Centre for Reference and Research on *Salmonella*, Institut Pasteur, Paris.

Poulsen, K., J. Reinholdt and M. Kilian, 1992 - A comparative genetic study of serologically distinct *Haemophilus influenzae* type 1 immunoglobulin A1 proteases. *J. Bacteriol.* **174**, 2913–2921.

Reda, K.B.,V. Kapur, J.A. Mollick, J.G. Lamphear, J.M. Musser and R.R. Rich, 1994 - Molecular characterization and phylogenetic distribution of the streptococcal superantigen gene (*ssa*) from *Streptococcus pyogenes. Infect. Immun.* **62**, 1867–1874.

Reeves, P.R., 1992 - Variation in O-antigens, niche-specific selection and bacterial populations. *FEMS Microbiol. Lett.* **100**, 509–516.

Reeves, P.R., 1993 - Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. *Trends Genet.* **9**, 17–22.

Reeves, M.W., G.M. Evins, A.A. Heiba, B.D. Plikaytis and J.J. Farmer 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. J. Clin. Microbiol.* **27**, 311–320.

Riley, M.A., 1993 - Positive selection for colicin diversity in bacteria. *Mol. Biol. Evol.* **10**, 1048–1059.

Sandmeier, H., 1994 - Acquisition and rearrangement of sequence motifs in the evolution of bacteriophage tail fibers. *Mol. Microbiol.* **12**, 343–350.

Sawyer, S., 1989 - Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **6**, 526–538.

Seifert, H.S. and M. So, 1988 - Genetic mechanisms of bacterial antigenic variation. *Microbiol. Rev.* **52**, 327–336.

Seifert, H.S., R.S. Ajioka, C. Marchal, P.F. Sparling and M. So, 1988 - DNA transformation leads to pilin antigenic variation in *Neisseria gonorrhoeae. Nature* **336**, 392–395.

Selander, R.K. and J.M. Musser, 1990 - Population genetics of bacterial pathogenesis, p. 11–36. *In* B.H. Iglewski and V.L. Clark (ed.), Molecular basis of bacterial pathogenesis. Academic Press, San Diego.

Selander, R.K., D.A. Caugant and T.S. Whittam, 1987 - Genetic structure and variation in natural populations of *Escherichia coli*, p. 1625–1648. *In* F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter, and H.E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2., American Society for Microbiology, Washington, D.C.

Selander, R.K., P. Beltran, N.H. Smith, R.M. Barker, P.B. Crichton, D.C. Old, J.M. Musser and T.S. Whittam, 1990a - Genetic population structure, clonal phylogeny, and pathogenicity of *Salmonella paratyphi* B. *Infect. Immun.* **58**, 1891–1901.

Selander, R.K., P. Beltran, N.H. Smith, R. Helmuth, F.A. Rubin, D.J. Kopecko, K. Ferris, B.D. Tall, A. Cravioto and J.M. Musser, 1990b - Evolutionary genetic relationships of clones of *Salmonella* serovars that cause human typhoid and other enteric fevers. *Infect. Immun.* **58**, 2262–2275.

Selander, R.K., P. Beltran and N.H. Smith, 1991 - Evolutionary genetics of *Salmonella*, p. 25–57. *In* R.K. Selander, A.G. Clark, and T.S. Whittam (ed.), Evolution at the molecular level. Sinauer Associates, Sunderland, Mass.

Selander, R.K., N.H. Smith, J. Li, P. Beltran, K. Ferris, D.J. Kopecko and F.A. Rubin, 1992 - Molecular evolutionary genetics of the cattle-adapted serovar *Salmonella dublin*. *J. Bacteriol.* **174**, 3587–3592.

Selander, R.K., J. Li, E.F. Boyd, F.-S. Wang and K. Nelson, 1995 - DNA sequence analysis of the genetic structure of populations of *Salmonella enterica* and *Escherichia coli*, p. 1–36. *In* F.G. Priest, A. Ramos-Cormenzana, and R. Tindall (ed.), Bacterial systematics and diversity. Plenum, New York.

Sharp, P.M., J.E. Kelleher, A.S. Daniel, G.M. Cowan and N.E. Murray, 1992 - Roles of selection and recombination in the evolution of type 1 restriction-modification systems in enterobacteria. *Proc. Natl. Acad. Sci. USA* **89**, 9836–9840.

Sibold, C., J. Wang, J. Henrichsen and R. Hakenbeck, 1992 - Genetic relationships of penicillin-susceptible and -resistant *Streptococcus pneumoniae* isolated on different continents. *Infect. Immun.* **60**, 4119–4126.

Simanjuntak, C.H., N.H. Punjabi, F.P. Paleologo, H. Totosudirjo, P. Haryanto, R. Darmowigoto, Soeprawoto [no initials], N.D. Witham, and S.L. Hoffman, 1992 - Vaccine trials for control of typhoid fever in Indonesia, p. 235–241. *In* T. Pang, C.L. Koh, and S.D. Puthucheary (ed.), Typhoid fever. Strategies of the 90's. World Scientific, Singapore.

Skurnik, M. and P. Toivanen, 1994 - *Yersinia enterocolitica* lipopolysaccharide: genetics and virulence. *Trends Microbiol.* **4**, 148–152.

Smith, N.H. and R.K. Selander, 1990 - Sequence invariance of the antigen-coding central region of the phase 1 flagellar filament gene (*fliC*) among strains of *Salmonella typhimurium*. *J. Bacteriol.* **172**, 603–609.

Smith, N.H. and R.K. Selander, 1991 - Molecular genetic basis for complex flagellar antigen expression in a triphasic serovar of *Salmonella*. *Proc. Natl. Acad. Sci. USA* **88**, 956–960.

Smith, N.H., P. Beltran and R.K. Selander, 1990 - Recombination of *Salmonella* phase 1 flagellin genes generates new serovars. *J. Bacteriol.* **172**, 2209–2216.

Spratt, B.G., L.D. Bowler, Q.-Y. Zhang, J. Zhou and J. Maynard Smith, 1992 - Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal *Neisseria* species. *J. Mol. Evol.* **34**, 115–125.

Stephens, J.C., 1985 - Statistical methods of DNA sequence analysis: detection of intragenic recombination or gene conversion. *Mol. Biol. Evol.* **2**, 539–556.

Stevenson, G., B. Neal, D. Liu, M. Hobbs, N.H. Packer, M. Batley, J.W. Redmond, L. Linquist and P. Reeves, 1994 - Structure of the O antigen of

*Escherichia coli* K-12 and the sequence of its *rfb* gene cluster. *J. Bacteriol.* **176**, 4144–4156.

Sudarmono, P. and M. Radji, 1992 - Features of typhoid fever in Indonesia, p. 11–16. *In* T. Pang, C.L. Koh, and S.D. Puthucheary (ed.), Typhoid fever. Strategies of the 90's. World Scientific, Singapore.

Thampapapillai, G., R. Lan and P.R. Reeves, 1994 - Molecular evolution in the *gnd* locus of *Salmonella enterica*. *Mol. Biol. Evol.* **11**, 813–828.

Thong, K.L., Y.M. Cheong, S. Puthucheary, C.L. Koh and T. Pang, 1994 - Epidemiologic analysis of sporadic *Salmonella typhi* isolates and those from outbreaks by pulsed-field gel electrophoresis. *J. Clin. Microbiol.* **32**, 1135–1141.

Threlfall, E.J., E. Torre, L.R. Ward, A. Dávalos-Pérez, B. Rowe and I. Gibert, 1994 - Insertion sequence IS*200* fingerprinting of *Salmonella typhi*: an assessment of epidemiological applicability. *Epidemiol. Infect.* **112**, 253–261.

Tominaga, A., M.A.-H. Mahmoud, T. Mukaihara and M. Enomoto, 1994 - Molecular characterization of intact, but cryptic, flagellin genes in the genus *Shigella*. *Mol. Microbiol.* **12**, 277–285.

Vázquez, J.A., L. de la Fuente, S. Berron, M. O'Rourke, N.H. Smith, J. Zhou and B.G. Spratt, 1993 - Ecological separation and genetic isolation of *Neisseria gonorrhoeae* and *Neisseria meningitidis*. *Curr. Biol.* **3**, 567–572.

Whatmore, A.M. and M.A. Kehoe, 1994 - Horizontal gene transfer in the evolution of group A streptococcal *emm*-like genes: gene mosaics and variation in Vir regulons. *Mol. Microbiol.* **11**, 363–374.

Whittam, T.S., 1995 - Genetic population structure and pathogenicity in enteric bacteria, p. 217–245. *In* S. Baumberg, J.P.W. Young, E.M.H. Wellington, and J.R. Saunders (ed.), Population genetics of bacteria. Cambridge University Press, Cambridge.

Whittam, T.S. and S.E. Ake, 1993 - Genetic polymorphisms and recombination in natural populations of *Escherichia coli*, p. 223–245. *In* N. Takahata and A.G. Clark (ed.), Mechanisms of molecular evolution. Sinauer Associates, Sunderland, Massachusetts.

Wilson, D.R. and T.J. Beveridge, 1993 - Bacterial flagellar filaments and their component flagellins. *Can. J. Microbiol.* **39**, 451–472.

Xiang, S.-H., A.M. Haase and P.R. Reeves, 1993 - Variation of the *rfb* gene clusters in *Salmonella enterica*. *J. Bacteriol.* **175**, 4877–4884.

Xiang, S.-H., M. Hobbs and P.R. Reeves, 1994 - Molecular analysis of the *rfb* gene cluster of a group D2 *Salmonella enterica* strain: evidence for its origin from an insertion sequence-mediated recombination event between group E and D1 strains. *J. Bacteriol.* **176**, 4357–4365.

Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, Pennsylvania 16802