The gnd gene, the structural gene for 6-phosphogluconate dehydrogenase, was sequenced and analyzed in 34 isolates from different serovars of the seven subspecies of Salmonella enterica to provide comparative information on the evolution in this gene, which has been studied extensively in Escherichia coli. The gene tree obtained by the neighbor-joining method in general gave separate branches for each subspecies, with the few exceptions readily explained by recombination. There is evidence of recombination involving transfer of long (more than 400 bp) and short (30–150 bp) segments of DNA. Four of the six long-segment transfers detected are at the 5’ end of the gene, and in all four cases a variant of the chi sequence is located close to the recombination junction and appears to have mediated the recombination events. We suggest that in these four cases and in a fifth case with intersubspecies transfer of the whole gnd gene, the adjacent rfb (O antigen) locus may have been transferred in the same event.

The estimates of the number of synonymous substitutions per synonymous site, Ks, and the number of nonsynonymous substitutions per nonsynonymous site, Ka, within the E. coli and S. enterica gnd genes, and also between the two species show an interesting distribution, with Ks being lower toward the ends of the gene and Ka in particular being lower in the first than in the second domain. In S. enterica, synonymous sites also seem to be subjected to negative selection. The ratio of Ka to Ks was higher within S. enterica and E. coli than between them, which may indicate that intraspecies variation is essentially between clones and that mildly deleterious mutations can be fixed within clones, which would thus raise Ka within species.

Introduction

We are undertaking several studies of variation within and between bacterial species with the aim of understanding the structure and evolution of bacterial populations. They include several studies of variation in the O antigen of Salmonella enterica, which is present in the outer membrane of gram-negative bacteria and is very variable, with about 60 and 160 forms detected in S. enterica and Escherichia coli, respectively (Ewing 1986). The structural variation in O antigen depends on genetic variation in the rfb gene cluster, which comprises a cluster of genes involved in O antigen synthesis (Mäkelä and Stocker 1984; Reeves 1992). The analysis of the rfb locus in S. enterica has shown that interspecies transfer has been a major force in generating the diversity (Reeves 1993).

Key words: S. enterica gnd gene, lateral gene transfer, chi sites, synonymous and nonsynonymous substitutions, selection and evolution.

Address for correspondence and reprints: Peter R. Reeves, Department of Microbiology, G08, University of Sydney, New South Wales 2006, Australia.
acidic, but alignment of amino acid sequences of 6-PGDH from *E. coli*, *S. enterica*, *Bacillus subtilis*, *Synechococcus sp.*, sheep and pig, has shown that these proteins are homologous and share conserved regions that are functionally important (Reizer et al. 1991). The structure of sheep 6-PGDH is known, and from this data the structure of *E. coli* and *S. enterica* 6-PGDHs can be deduced (Adams et al. 1991; M. Adams, personal communication). The enzyme 6-PGDH consists of three domains: an N terminal α/β NADP binding domain (residues 1–176), a second all α domain extending from residues 177–434, and a tail comprising residues 435–468. The amino acid sequence of the *gnd* gene has two putative nucleotide binding sequences, one in residues 10–15 with a consensus sequence GXAXXXG (Adams et al. 1991) and another, GXGXGXXXXG, for the βαβ fold in residues 124–133 and a strongly conserved putative substrate binding sequence ILDXANKGTGK in residues 253–264 (Reizer et al. 1991).

In this article we present an analysis of the DNA sequence variation in the *gnd* gene of *S. enterica* and show that the evolution of this gene involves recombination at both ends of the coding region and that, at least at the 5' end, these recombinations seem to be mediated and perhaps involve the adjacent rfb locus. Comparison of the number of nucleotide substitutions in the synonymous and nonsynonymous sites in the *E. coli* and *S. enterica* *gnd* genes indicates that both nonsynonymous and synonymous sites are subjected to negative selection in *S. enterica*.

**Material and Methods**

Bacterial Strains and DNA Sequences

Details of the *Salmonella enterica*, *Citrobacter freundii*, and *Yersinia pseudotuberculosis* strains used are given in table 1. In the data presented, we have added to the *S. enterica* strain names a suffix s1, s2, and so forth, to indicate the subspecies classification of the strain. For example, M318s4 indicates that strain M318 is a subspecies IV strain. We have used 16 published *Escherichia coli* *gnd* sequences for comparison with the *S. enterica* sequences. Details of the *E. coli* sequences, with GenBank numbers in parentheses, are K-12 (K02072) (Nasoff et al., 1984); ECOR 4 (M64324), ECOR 16 (M64325), ECOR 65 (M64331), ECOR 68 (M64330), ECOR 69 (M64328), and ECOR 70 (M64329) (Dykhuizen and Green 1991); ECOR 10 (M63821), ECOR 11 (M63822), ECOR 18 (M63823), ECOR 20 (M63824), ECOR 21 (M63825), ECOR 23 (M63826), ECOR 25 (M63827), ECOR 47 (M63828), and ECOR 56 (M63829) (Bisercic et al. 1991).

The *gnd* DNA sequences of 34 natural isolates of *S. enterica* representing several O antigen forms and the seven documented subspecies (Crosa et al. 1973; Le Minor et al. 1986) have been analyzed in this work. The 34 *S. enterica* *gnd* sequences include 33 new *S. enterica* *gnd* genes sequenced for this study and the already published *S. enterica* LT2 *gnd* sequence (Reeves and Stevenson 1989). The *gnd* gene of a *C. freundii* strain 396 was also sequenced at the same region and used as an outgroup in the phylogenetic analysis. The analysis is based on 1,329 bp of DNA sequence bases from 16 to 1344 in the coding region of the *S. enterica* *gnd* gene (coding sequence 1,404 bp). Bases 16 to 400 of *gnd* DNA of a *Y. pseudotuberculosis* strain were also sequenced and used as an outgroup in the phylogenetic analysis of *E. coli* and *C. freundii* strains.

In some strains, the *gnd* sequence differed from that expected for the particular species or subspecies, and in such cases sufficient properties were tested to confirm the species or subspecies designation. We acknowledge the help of Robert Chiew, Westmead Hospital, Sydney, for confirmation of the subspecies status of *S. enterica* M130s2, M38s2, M298s1, and M318s4, and Dr. K. A. Bettelheim and Dr. D. E. Leslie, Fairfield Hospital, Victoria, for confirmation that our stock of *C. freundii* 396 is indeed *C. freundii*.

**PCR and Sequencing**

Chromosomal DNA was extracted by the method devised by Ardeshir et al. (1981). The sequences for oligonucleotide primers were chosen taking into account the segments conserved in the published *Escherichia coli* K-12 (Nasoff et al. 1984) and *Salmonella enterica* LT2 (Reeves and Stevenson 1989) *gnd* gene sequences. The following oligonucleotides were used for PCR amplification and sequencing:

No. 218, 5' tgcgaaagcagcgccagt-
CCAAAGCAACAGATCGG 3',
No. 239, 5' tggaaaacgagccgagt-
TCTGATTCGCTGAAACC 3',
No. 261, 5' tggaaaacgagccgagt-
GAATATGGCGATATGCA 3',
No. 279, 5' tggaaaacgagccgagt-
TCTTGGAAGATCGT 3',
No. 219, 5' caggaacagctagcc-
TATTAGGTGCGAC 3', and
No. 260, 5' caggaacagctagc-
ATCGGCGTITTCTGCGTA 3'.

The segments shown in upper case are *gnd* gene sequences, and those in lower case are either the universal forward or reverse M13 primer sequences. DNA from the coding region was amplified by the PCR method described by Saiki et al. (1988). The amplified PCR product was purified with prep A gene matrix (BioRad) according to the manufacturer's instructions in order to remove excess PCR primers, precipitated by ethanol and resuspended in 8 ml of TE (pH 8). Cycle sequencing
Table 1
Information on Strains Used

<table>
<thead>
<tr>
<th>Serovar (for S. enterica) or Species</th>
<th>Strain a</th>
<th>Subspecies</th>
<th>O Antigen</th>
<th>Source b</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typhimurium LT2</td>
<td>SARA 2</td>
<td>I</td>
<td>B</td>
<td>a</td>
</tr>
<tr>
<td>Paratyphi B</td>
<td>SARA 41</td>
<td>I</td>
<td>B</td>
<td>a</td>
</tr>
<tr>
<td>Choleraesuis</td>
<td>M36</td>
<td>I</td>
<td>C1</td>
<td>d</td>
</tr>
<tr>
<td>Tennessee</td>
<td>M55</td>
<td>I</td>
<td>C1</td>
<td>d</td>
</tr>
<tr>
<td>Giostrup</td>
<td>M46</td>
<td>I</td>
<td>C2</td>
<td>d</td>
</tr>
<tr>
<td>Muenchen</td>
<td>SARA 71</td>
<td>I</td>
<td>C2</td>
<td>a</td>
</tr>
<tr>
<td>Typhi</td>
<td>Ty21a (M229)</td>
<td>I</td>
<td>D1</td>
<td>m</td>
</tr>
<tr>
<td>Strasbourg</td>
<td>M13</td>
<td>I</td>
<td>D2</td>
<td>d</td>
</tr>
<tr>
<td>Canoga</td>
<td>M35</td>
<td>I</td>
<td>E3</td>
<td>i</td>
</tr>
<tr>
<td>Senftenberg</td>
<td>M73</td>
<td>I</td>
<td>E4</td>
<td>j</td>
</tr>
<tr>
<td>Berkeley</td>
<td>M295</td>
<td>I</td>
<td>43</td>
<td>i</td>
</tr>
<tr>
<td>Dahlem</td>
<td>M298</td>
<td>I</td>
<td>48</td>
<td>i</td>
</tr>
<tr>
<td>Sofia</td>
<td>M494</td>
<td>II</td>
<td>B</td>
<td>w</td>
</tr>
<tr>
<td>9,12: mt: enx</td>
<td>M130</td>
<td>II</td>
<td>D1</td>
<td>w</td>
</tr>
<tr>
<td>Lindrick</td>
<td>M495</td>
<td>II</td>
<td>D1</td>
<td>w</td>
</tr>
<tr>
<td>1,9,12: g,mst: 1.5z42</td>
<td>M496</td>
<td>II</td>
<td>D1</td>
<td>w</td>
</tr>
<tr>
<td>Haarlem</td>
<td>M38</td>
<td>II</td>
<td>D2</td>
<td>w</td>
</tr>
<tr>
<td>Westpark</td>
<td>M497</td>
<td>II</td>
<td>E1</td>
<td>w</td>
</tr>
<tr>
<td>Springs</td>
<td>M261</td>
<td>II</td>
<td>40</td>
<td>i</td>
</tr>
<tr>
<td>Freemantle</td>
<td>M287</td>
<td>II</td>
<td>42</td>
<td>i</td>
</tr>
<tr>
<td>Phoenix</td>
<td>M311</td>
<td>II</td>
<td>47</td>
<td>p</td>
</tr>
<tr>
<td>Ar 10a10b: 17,20</td>
<td>M314</td>
<td>IIIa</td>
<td>40</td>
<td>p</td>
</tr>
<tr>
<td>Ar 1,3,1,3,11</td>
<td>M313</td>
<td>IIIa</td>
<td>44</td>
<td>p</td>
</tr>
<tr>
<td>Ar 38: 1, v: z53, z54</td>
<td>M498</td>
<td>IIIb</td>
<td>38</td>
<td>w</td>
</tr>
<tr>
<td>Ar 28:27–21</td>
<td>M316</td>
<td>IIIb</td>
<td>47</td>
<td>p</td>
</tr>
<tr>
<td>Ar 5, 29: 33–31</td>
<td>M317</td>
<td>IIIb</td>
<td>48</td>
<td>p</td>
</tr>
<tr>
<td>38: z4, z23</td>
<td>M318</td>
<td>IV</td>
<td>38</td>
<td>p</td>
</tr>
<tr>
<td>Houten</td>
<td>M319</td>
<td>IV</td>
<td>43</td>
<td>p</td>
</tr>
<tr>
<td>43. z4 . z23</td>
<td>M320</td>
<td>IV</td>
<td>43</td>
<td>p</td>
</tr>
<tr>
<td>Brookfield</td>
<td>M322</td>
<td>V</td>
<td>66</td>
<td>p</td>
</tr>
<tr>
<td>Balboa</td>
<td>M321</td>
<td>V</td>
<td>48</td>
<td>p</td>
</tr>
<tr>
<td>Marseille</td>
<td>M324</td>
<td>VI</td>
<td>F</td>
<td>p</td>
</tr>
<tr>
<td>41 : b:1,7</td>
<td>M325</td>
<td>VI</td>
<td>41</td>
<td>p</td>
</tr>
<tr>
<td>Vrindaban</td>
<td>M326</td>
<td>VI</td>
<td>45</td>
<td>p</td>
</tr>
<tr>
<td><em>Citrobacter freundii</em></td>
<td>396 (M132)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Yersinia pseudotuberculosis</em> (Grp. V)</td>
<td>M89</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a Laboratory name is given in parentheses when strain name is not available.

b The sources of the strains are as follows: a, SARA Collection, R. K. Selander (Beltran et al., 1991); p, L. Le Minor, Institut Pasteur, Paris, France; w, R. Chiew, Westmead Hospital, Sydney, Australia; i, C. Murray, Institute of Medical and Veterinary Science, Adelaide, Australia; j, K. Jann, Max-Plank-Institut für Immunobiologie, Freiberg; m, J. Taplin, Department of Microbiology, University of Melbourne, Australia; d, This laboratory; y, Yersinia Reference Centre, Public Health Laboratory, Leicester Royal Infirmary, Leicester U.K.

Reactions were carried out according to the method described by Applied Biosystems using a dye-labeled primer complementary to the universal primer sequence present in the primer used for PCR amplification. The extended products from the cycle sequencing reactions were run on a denaturing polyacrylamide gel and read by an ABI373A sequencer.

Computer Analysis

The sequences were edited and analyzed by programs made available through the Australian National Genomic Information Service (ANGIS), at the University of Sydney (Reisner et al. 1993). The DNA sequences were analyzed using MULTICOMP (Reeves et al. 1994), which gives pairwise comparisons of DNA and derived amino acid sequences and facilitates the use of programs such as PAUP, MACLADE and those within the PHYLIP package (version 3.4 written by Joseph Felsenstein, Department of Genetics, University of Washington, Seattle). It also uses Sawyer’s algorithm to detect intragenic recombination (Sawyer 1989), calculates nucleotide diversity $\pi$, using the method given by Nei and Miller (1990) and estimates $K_s$, the number of synonymous substitutions per synonymous site, and $K_A$, the number...
of nonsynonymous substitutions per nonsynonymous site, using the program kindly provided by Wen-Hsiung Li of the University of Texas, Houston (Li et al. 1985; Li 1993). The values for $K_S$ and $K_A$ in the interspecies comparison were derived by estimating the values for pairwise comparisons of each *E. coli* strain against each *S. enterica* strain and taking the average. Phylogenetic trees were constructed using the neighbor-joining (NJ) method described by Saitou and Nei (1987). The program, METREE (Rzhetsky and Nei 1992), which identifies the minimum evolution tree and other trees similar to the minimum evolution tree, was kindly provided by A. Rzhetsky, Pennsylvania State University, and was used on our NJ trees. In each case these ME trees were only minor variants of the NJ tree, and the NJ trees are presented with the bootstrap values given by the METREE program.

**Results and Discussion**

**Phylogenetic Analysis**

In the 34 *Salmonella enterica gnd* sequences there are 314 polymorphic sites, and of these, 201 are phylogenetically informative, such sites having at least two bases present in two or more strains (fig. 1). These informative bases have been used to construct a phylogenetic (NJ) tree using the *gnd* sequences of *Citrobacter freundii* strain 396 as outgroup. The NJ tree (fig. 2a) showed that, in general, DNA polymorphism at the *gnd* locus of *S. enterica* corresponded to subspecies variation within the species. However, there are exceptions. The sequences of two strains, those of M298s1 and M326s6, were displaced from the branches that contain the other members of the same subspecies, which shows that in these strains the whole or most of the *gnd* gene has been replaced by lateral transfer.

Three other recombination events, in addition to those in M298s1 and M326s6, appear to affect the topology of the gene tree. In the placement of subspecies IV, IIIa, and V, the *gnd* gene tree does not correspond to the trees based on the *gapA* or *putP* genes (Nelson et al. 1991; Nelson and Selander 1992). However, if the strains M298s1 and M318s4 and the segment from positions 916 to 1344 in the two subspecies V strains are excluded from the analysis (as they contain DNA atypical for the subspecies), the tree obtained is similar to those for *gapA* and *putP*, with the exception of M326s6 (fig. 2b). Indeed, to subspecies level, there is a consensus tree for the three genes, with the *putP* tree differing from the *gapA* tree only in the order of subspecies IIIa and IV and the *gnd* tree (fig. 2b) differing from *gapA* tree only in the placement of subspecies II. We conclude that this consensus tree is the best estimate for the relationship between subspecies and that recombination on a scale that affects the tree for the *gnd* gene has occurred in only the few strains referred to above. The MLEE trees obtained by Reeves et al. (1989) and Selander et al. (1991) differ in the relationships of the subspecies in several ways from each other and from this consensus gene tree, which suggests that for determining the relationships between the very divergent subspecies of *S. enterica*, a tree derived from MLEE data may not be as good as one based on DNA sequence of a few genes.

**Recombination in the *Salmonella enterica gnd* Gene**

Intersubspecies transfers involving segments of about 400-750 bp or more in size were detected in a total of 6 of 34 strains analyzed. All six recombinant *gnd* sequences, those of strains M318s4, M298s1, M38s2, M130s2, M321s5, and M322s5, have the same characteristic feature, in that the deduced donor segments are located at either end of the gene, which indicates the possibility that these transfers may involve regions adjacent to the *gnd* gene on either side. In addition to these large segment transfers, lateral transfers of small segments of DNA (30-150 bp) are also evident in many strains.

**Detection of Recombination**

The conclusions on recombination events were derived from both phylogenetic and statistical analysis. In general all methods of phylogenetic analysis infer that DNA sequence similarity is due to common ancestry. Phylogenetic analysis based on protein polymorphisms at 24 loci (Selander et al. 1991) and DNA sequence variation at *gapA* and *putP* loci have confirmed that *S. enterica* genome is essentially clonal, and the genetic relatedness in *S. enterica* strains strongly follows the subspecies variation.

Given this background, in a phylogenetic reconstruction, one would expect strains belonging to the same subspecies of *S. enterica* to cluster in the same branch. When a gene tree places a *S. enterica* strain in a branch different to that in which the members of its assigned subspecies cluster, one could infer that the erroneous or unusual placement of the strain in a different branch is due to lateral gene transfer. Such observation is further confirmed when independent analysis such as biotyping, serotyping, or DNA sequence analysis at a different locus confirms the original subspecies status of this strain.

In the same manner, if part of the sequence places a strain in the branch that contains the other members of the same subspecies, while another part of the same gene places the strain elsewhere, it can be inferred that the latter segment has been acquired by partial gene transfer. In such chimeric DNA segments of an essentially clonal bacterium such as *S. enterica*, one could not only detect recombination but also deduce the donor
FIG. 1.—Alignment of informative bases of Salmonella enterica gnd gene sequences showing only those bases that differ from the consensus sequence shown above. Asterisks indicate nonsynonymous substitutions, and the position of the base in its codon is indicated in the last line. Suffix s1, a2, and so forth, indicates subspecies.
and recipient parts of the DNA segments with some degree of accuracy. This rationale has been used previously for detecting recombination in other species (Gyllensten et al. 1991).

The statistical test for detecting gene conversion devised by Sawyer (1989) was also applied to confirm the observed recombination events. A subset of eight strains, S41s1, M322s5, M298s1, M130s2, M495s2, M311s2, M318s4, and M320s4, was selected, and the distribution of the silent informative sites in the 1329-bp gnd DNA was analyzed. The $P$ values for the sum of the squares of the condensed fragment lengths (SSCF) and maximum condensed fragment length (MCF) were highly significant. The extremely low $P$ values of $<0.0001$ for both parameters show that the type of clustering of silent sites present in our real data set cannot be reproduced by permuted data sets. One set of results of Sawyer's test is presented in table 2. One point to note in the four condensed fragments shown in table 2 is that unlike in the other three pairs of strains, the similarity between strains M130s2 and M495s2 from region 429 to 984 bp is not due to recombination but to common descent, and the dissimilarity between these two strains in the segment spanning 16-429 bp is due to a recombination event at this region in strain M130s2.

Several small segment transfers of DNA of about 30-150 base pairs were also observed in many of the 34 gnd sequences. Sawyer's test was carried out on several sets of strains to provide statistical evidence for these gene conversion events, and the $P$ values obtained were highly significant.

*Recombination in Strain M318s4*

Strain M318s4 is a subspecies IV strain that has a chimeric gnd sequence with segments derived from subspecies I and IV. The first 771 bp of its sequence is of subspecies I type and identical to the sequences of strains S41s1 and M46s1 (figs. 1 and 3a). From position 819, it resembles other subspecies IV strains, M319s4 and M320s4, and not the subspecies I strains, S41s1 or M46s1. As allelic variation in this gene in general corresponds to subspecies variation, the subspecies I DNA present up to position 771 in strain M318s4 must have been acquired by recombination. The junction of the subspecies I and subspecies IV DNA could be anywhere between positions 771 and 819 (or even 881). A single-base variant, 5' CCTGGTGG 3', of the general recombination stimulating sequence chi, 5' GCTGGTGG 3', is located at positions 744-751 bp. This chi-like sequence is near the 3' end of the deduced donor sequence thought to be derived from a subspecies I strain. Sequences of M318s4 and related strains illustrating the recombination event and showing the chi variant are given in figure 3a.

*Recombination in Strains M38s2 and M130s2*

Although the NJ tree constructed with informative bases present in the entire region sequenced places the sequences of strains M38s2 and M130s2 with other subspecies II strains, a phylogenetic tree constructed with informative bases in the first 429 bases places these two strains in a branch away from that of subspecies II (fig. 2c). We suggest that the segment to 429 base, unique to strains M38s2 and M130s2, was acquired by recombination. We set the junction in the vicinity of position 429-477 bp. Alignments of M38s2 and M130s2 with related strains M495s2 and M497s2 are shown in figure 3b.
The same single-base variant of chi reported above is also present in strains M38s2 and M130s2 at positions 424–431 bp, very close to the end of the distinctive segment (fig. 3b).

**Recombination in Strain M298s1**

Twelve subspecies I strains have been analyzed at the *gnd* locus. With the exception of the sequence of M298s1, they show very little variation, particularly up to position 960 bp. The sequence of M298s1 is atypical for this subspecies; irrespective of the region used in phylogenetic construction, it is placed in a branch different to the one in which other subspecies I strains cluster. Therefore, transfer of the whole gene is invoked. The *gnd* sequence of M298s1 has segments representative of subspecies II at the 5' end and segments representative of subspecies IIIb at the 3' end, which suggests multiple recombination events at this locus.

The subspecies II *gnd* gene is highly variable, and those analyzed fall into two types (fig. 2), which we refer as type IIa and IIb simply for this discussion. A major difference between the two types is the presence of the chi-like sequence at positions 744–751 bp in type IIb strains M38s2, M130s2, M497s2, and M495s2. The type IIa strains M311s2, M494s2, M261s2, and M287s2 have clustered substitutions in and around this chi-like region, with the exception of type IIa strain M496s2. The *gnd* sequence of M298s1 is almost identical to those of M311s2 and M494s2 (type IIa) from the start of the gene to around position 705 but differs from them by possessing the chi-like sequence in position 744–751 bp as in type IIb strains. Beyond this region, sequence of M298s1 resembles that of subspecies IIIb (figs. 2c, 2d, and 3b). We suggest that there could have been two consecutive transfers of *gnd* DNA, whole-gene transfer with a subspecies IIIb strain as donor initially and a subsequent chi-mediated transfer of a partial segment at the 5' end from a subspecies II strain.

**Recombination in Subspecies V Strains, M321s5 and M322s5**

Our analysis involved 12 subspecies I and two subspecies V strains. There is little variation within either subspecies in the 5' end, other than in strain M298s1 discussed above, but considerable differences exist between the two subspecies, as expected from the MLEE analysis (Reeves et al. 1989; Selander et al. 1991). However, in the 3' end, from about position 930 bp, the subspecies V DNA is almost identical to that of subspecies I (fig. 3d). An NJ tree constructed with the informative bases present in the region 915–1344 bp (fig. 2d) places the two subspecies V strains with subspecies I strains. This is an unusual placement for subspecies V strains that could only be possible if they have acquired a subspecies I segment by lateral transfer. A third subspecies V strain sequenced in part also had subspecies I DNA in this region (data not shown). The subspecies I strains are very similar to each other and do not have any distinct groups within them up to position 1089 bp. From position 1089 to 1221 bp they form two distinct groups,
and the sequence of one subspecies V strain, M322s5, is similar to that of LT2s1, which differs from the consensus sequence at positions 1148, 1152, 1155, and 1158 bp. The sequence of the other subspecies V strain, M321s5, is similar in this region to that of the other group, eg M55s1, which differs from the consensus sequence in positions 1089, 1119, 1146, 1215, and 1221 bp. This suggests that the differences within subspecies I strains in this region must have evolved before the transfer of subspecies I DNA into subspecies V gene.

Interspecies Recombination in the \textit{gnd} Gene

Evidence for interspecies transfers involving large segments of DNA is absent in the \textit{Salmonella enterica gnd} gene. A similar observation was made by Nelson and Seelander (1992). However, the region between 960 and 1200 bp shows a high level of amino acid diversity in both \textit{S. enterica} and \textit{Escherichia coli gnd} genes, and transfers of small segments from divergent species cannot be discounted in this region. In the \textit{gapA} gene, Nelson and Seelander (1992) observed a transfer of a short segment of DNA from \textit{Klebsiella pneumoniae} to subspecies V strains of \textit{S. enterica}. This transfer resulted in six amino acid changes within 48 bases. Similar clustered amino acid changes have been observed in the 960- to 1200-bp region in the \textit{S. enterica gnd} gene, a particularly striking case being the segment 989-1029 of M38s2, which includes four amino acid substitutions within 40 bases (fig. 4). However, as no donor species is known, it is difficult to confirm interspecies recombination.

We have used the 16 published \textit{E. coli gnd} sequences to compare with \textit{S. enterica gnd} sequences; among them, those of ECOR 4 and ECOR 16 are very divergent. At the DNA level they are more than 15\% divergent from each other and from the other \textit{E. coli gnd} sequences. This is near the average level of DNA divergence between \textit{E. coli} and \textit{S. enterica} seen for many genes and is unusually high for alleles within a species,
as previously noted by Dykhuizen and Green (1991), who sequenced these two divergent E. coli gnd alleles.

An NJ tree of the E. coli, S. enterica, and Citrobacter freundii 396 gnd sequences using the partial gnd sequence of Yersinia pseudotuberculosis as an outgroup (fig. 5a) places the ECOR 4 and C. freundii 396 gnd genes in the same branch between the major E. coli and S. enterica branches, an unusual location for a C. freundii strain. The placement of ECOR 4 and C. freundii 396 was the same when the gnd genes of Synechococcus sp. or Trypanosoma brucei were used as the outgroup. The similarity of the DNA sequences of the ECOR 4 and C. freundii 396 gnd genes is shown in figure 6. Observation of interspecies recombination at the gnd locus of E. coli was mentioned by Nelson and Selander (1992), although details were not given.

Escherichia coli and S. enterica are more closely related to each other than either is to C. freundii, a view supported by trees based on the gapA and ompA genes (Lawrence et al., 1991). At the DNA level, the divergence of the gnd sequence of C. freundii 396 from those of both S. enterica LT2 and E. coli K-12 is similar (15%) to the divergence of the E. coli K-12 and S. enterica LT2 sequences, but at the amino acid level E. coli K-12 and S. enterica LT2 are closer to each other than either is to C. freundii 396 (fig. 5b). The unusual placement of the C. freundii 396 gnd DNA sequence perhaps reflects a recombination event involving the gnd gene. Citrobacter freundii 396 has an O antigen that resembles a hybrid of S. enterica groups, C1 and B (Jann and Jann 1984) and perhaps the rfb and gnd genes were jointly involved in an interspecies recombination event. In this context it is interesting that the G+C content at the third base of the E. coli ECOR 4 gnd gene is 0.59, much higher than that of the other E. coli strains in which it ranges from 0.52 to 0.56. In this regard the ECOR 4 gene also resembles that of the S. enterica gene, which has an average value of 0.59, while the range is 0.57–0.62. The corresponding value for C. freundii 396 gene is 0.58.

The C. freundii 396 sequence is very different to those of S. enterica to which its O antigen is related, but examination of the sequences shows that at some sites an allele shared by C. freundii 396 and ECOR 4 is also present in most S. enterica strains but not in other E. coli strains. It appears that there may have been some complex interactions.

Relationship between the rfb Locus and Recombination in the gnd Gene

The deduced donor segments in recombinant strains M318s4, M298s1, M130s2, and M38s2 extend

![Figure 4](image-url)

**Fig. 4.**—Alignment of a segment of gnd gene sequences of strains M130s2, M38s2, and M497s2, with LT2s1 as the reference strain to show a short, highly divergent segment in M38s2.
into the \textit{gnd} gene from the beginning of the sequence. The \textit{rfb} locus is immediately upstream of the \textit{gnd} gene, and we suggest that these recombination events could have also involved the \textit{rfb} gene cluster (fig. 7).

Strain M318s4 is a subspecies IV strain carrying O antigen 38, predominantly present in subspecies I and IIIb (table 3). We suggest that at least part of the \textit{rfb} region and 5' end of \textit{gnd} were transferred together to strain M318s4 from a subspecies I strain by a chi-stimulated recombination event. A similar conclusion can be drawn for strain M298s1, which carries O antigen 48, prevalent in subspecies IIIb, and to a lesser extent in subspecies II and has a \textit{gnd} gene with segments from both of these subspecies.

In strain M326s6 the entire \textit{gnd} gene appears to be derived from subspecies I: it carries O antigen 45, which is predominantly found in subspecies I, again suggests the possibility of a recombination event involving \textit{rfb} and in this case the whole of the \textit{gnd} gene. Strains M130s2 and M38s2 carry O antigens D1 and D2, respectively. These are predominantly subspecies I epitopes but also present to a significant degree in subspecies II. It is possible that these \textit{O} antigens were transferred to subspecies II together with the 5' end of the \textit{gnd} gene. However, although the sequences of strains M130s2 and M38s2 are similar in the 5' end, especially up to the junction site, and would appear to derive from a single recombination event, they have different \textit{O} antigens D1 and D2. The \textit{rfb} regions of groups D1 and D2 differ only in the central region of the gene cluster (Xiang and Reeves 1994), and substitution of one for the other in the central region could have occurred in a separate event involving only \textit{rfb} locus. Recombination involving the \textit{rfb} cluster and the \textit{gnd} gene occurs in relatively few strains, and in general \textit{gnd} variation corresponds to the subspecies relationship. The overall variation in \textit{rfb} is more extensive, and the distribution of \textit{O} antigen types is widespread, with most forms present in more than one subspecies (Ewing 1986). Intersubspecies recombination involving the \textit{rfb} genes must be relatively common, and it appears that in many cases recombination affecting \textit{rfb} has no effect on \textit{gnd} genes.

Chi Activity in the \textit{Salmonella enterica} \textit{gnd} Gene

In the recombinant \textit{gnd} sequences of strains M318s4, M298s1, M38s2, and M130s2, a chi-like sequence 5' CCTGGTGG 3' is located at the 3' end of the deduced donor segments. Chi, an octamer DNA sequence element, 5' GCTGGTGG 3', has been shown to stimulate recombination in \textit{Escherichia coli} (Stahl 1979) and \textit{Salmonella enterica} (Smith et al. 1986). The chi

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|c|c|}
\hline
\textbf{O Antigen} & \textbf{I} & \textbf{II} & \textbf{IIIa} & \textbf{IIlb} & \textbf{IV} & \textbf{V} & \textbf{Total} \\
\hline
O38 & 23 & 3 & 1 & 18 & 2 & 0 & 47 \\
O48 & 5 & 10 & 5 & 19 & 3 & 2 & 44 \\
D1 & 55 & 24 & 0 & 0 & 0 & 79 \\
D2 & 51 & 11 & 0 & 0 & 0 & 62 \\
O45 & 15 & 9 & 4 & 0 & 2 & 30 \\
Others & 1,103 & 339 & 83 & 231 & 35 & 5 & 1,796 \\
\hline
\textbf{Total} & 1,252 & 396 & 93 & 268 & 42 & 7 & 2,058 \\
\hline
\end{tabular}
\caption{Distribution of Selected O Antigens among Subspecies*}
\end{table}

* Number of serovars with given O antigen. From Ewing (1986).
sequence is recognized by the multifunctional enzyme RecBCD, which mediates recombination by the RecBC pathway (Smith 1987; West 1992). The substrate for RecBCD enzyme is a linear double-stranded DNA that the enzyme unwinds and degrades as it moves along (Ganesan and Smith 1993). Recognition of the chi sequence terminates the nuclease activity of the enzyme (Dixon and Kowalczykowski 1993). The enzyme continues to unwind the DNA, releasing a single-stranded DNA with a 3' tail that is the ideal substrate for RecA protein, which, together with SSB protein, initiates strand invasion, the first step in the complex recombination process. The resulting recombinant would therefore have the donor DNA at the 5' end, with the chi sequence almost defining the 3' end of the donor DNA (Smith 1991). The location of the chi sequence and its polarity with respect to donor and recipient DNA segments in the S. enterica gnd gene recombinants M318s4, M298s1, M38s2, and M130s2 is in agreement with this model for chi-dependent recombination. The chi-like sequence in these strains is also oriented in the directions of replication and transcription as found for most chi sites in E. coli (Burland et al. 1993).

However, there are two discrepancies. The chi element associated with these junctions is a single-base variant of the E. coli chi sequence. Single-base mutations in chi are known to reduce the level of chi activity in E. coli to varying degrees depending on the base and position within chi (Smith et al. 1984), but we suggest that this variant 5’ CCTGGTGG 3’ is associated with recombination functions under natural conditions in S. enterica. In strains M318s4, M38s2, and M130s2 the junction appears to be at least 20 bp 3’ of the chi sequence rather than 3–10 bp as observed in in vitro experiments (Ponticelli et al. 1985). However, it should be noted that there are other aspects of the in vivo effects of chi, which remain unexplained (Smith 1987). We estimated the probability (Krowczynska et al. 1990) that the proximity of a chi site or any of the 24 single-base variants being within 50 bp of the recombination junction by chance to be 0.00069 for three of five junctions (for this calculation we conservatively treat M38s2 and M130s2 as deriving from a single recombination event).

Chi Activity and Recombination between Subspecies I and V in putP

As chi-like sequences are associated with several recombination events in the gnd gene of Salmonella enterica, we looked at the gapA and putP data in both Escherichia coli and S. enterica for evidence of chi activity. An additional example of possible chi activity was observed in the putP sequences of S. enterica (Nelson and Selander 1992). The putP sequences of subspecies V are the most divergent among the S. enterica putP genes in this set, yet in the 1467-bp coding region of this gene, the sequence of subspecies V, from about positions 400 to 1145 (about 750 bp), is very similar to that of subspecies I (fig. 8). This similarity between the two subspecies ends at the region at about 1145 bp, and the DNA beyond this region in subspecies V has the expected level of divergence characteristic for this subspecies. This unexpected similarity between subspecies I and V could have been due to the acquisition of subspecies I DNA by subspecies V; the possibility of recombination between the two subspecies was suggested by Nelson and Selander (1992). There is an authentic chi site between positions 1137 and 1144 immediately adjacent to the recombination junction in subspecies V putP genes. This chi site, for those in the gnd gene, is in the same orientation as replication and transcription. It is remarkable that for both gnd and putP, the genes of subspecies V strains have substantial segments from subspecies I, which suggests that, despite the great divergence between the subspecies, this may be a widespread situation.

Nelson and Selander (1992) have also proposed a recombination event in the putP genes of subspecies VII. The same central region described in the recombination of subspecies V putP sequences is involved and the same chi site is also present in subspecies VII putP sequences. We suggest that the recombination in subspecies VII of

---

**Fig. 8.—Alignment of the putP genes of subspecies I, II, V, and VII on both sides of a putative chi-mediated recombination event. Data from Nelson et al. (1992). Presentation as in fig. 3.**
putP could also have been stimulated by the same chi element.

Comparison of Polymorphism in the gnd Genes of Salmonella enterica and Escherichia coli

The gnd gene of Escherichia coli has been reported to be extremely variable, and a major objective of this study was to investigate the level of polymorphism in Salmonella enterica. Among the 16 E. coli sequences available, there are 378 polymorphic sites compared to the 318 polymorphic sites among the 34 S. enterica sequences. Nucleotide diversity, π, the average number of nucleotide substitutions per site for a set of alleles sampled from a population (Nei 1987), was used to estimate the level of variation in the gnd gene and other genes of these two species (table 4). We are limited by the number of genes sequenced in multiple strains of S. enterica (data for only gnd, putP, and gapA are available). The data presented in table 4 show that the π values for E. coli genes range from 0.002 for gapA to 0.074 for the gnd gene with the other four in the range 0.011 to 0.024. The π values for the three genes of S. enterica show much less variation, and the gnd gene with a π value of 0.048, about two-thirds the value for the gnd gene of E. coli, seems typical for S. enterica, although the number of genes studied is very small.

The E. coli gnd sequences used in this analysis include two very divergent sequences: those of ECOR 4 and ECOR 16, of which ECOR 4 shows evidence of interspecies transfer (see the section on interspecies recombination). If these two aberrant strains are omitted from the analysis, the nucleotide diversity seen among the other 14 E. coli gnd sequences is almost the same (0.046) as that observed for S. enterica (0.048).

The small number of genes studied and the variation between genes makes it difficult to establish norms at this stage and hence to evaluate the situation for the gnd gene. The E. coli gapA and putP genes have less variation than their homologues in S. enterica genes. For putP this correlates reasonably well with the depth in their respective MLEE trees (Reeves et al. 1989; Herzer et al. 1990; Selander et al. 1991). The remarkable difference in the variation in gapA in the two species shows that species specific effects can be very marked, and there is no explanation as yet for the gapA situation. The gnd gene of E. coli is still seen to have a relatively high level of variation for the species, even if two highly divergent strains are excluded. This is not the case in S. enterica, which suggests that for two of three genes for which we have comparative information, different circumstances may apply in the two species.

Synonymous and Nonsynonymous Substitutions in the gnd Gene of Salmonella enterica and Escherichia coli

We computed Ks, the number of synonymous substitutions per synonymous site, and KA, the number of nonsynonymous substitutions per nonsynonymous site (see Material and Methods), and compared the values for the Salmonella enterica and Escherichia coli gnd sequences. We have presented two sets of Ks and KA data for E. coli. One, the E. coli (16) set, includes the two aberrant strains, while the E. coli (14) set excludes them. We also examined Ks and KA in comparisons between E. coli and S. enterica and between subspecies 1 and II of S. enterica (table 5).

Values for Ks and KA vary along the gene (table 5 and fig. 9), and in all three groups the variation in KA is more marked than that in Ks. This is expected as structural and functional constraints limit substitutions at nonsynonymous sites, and in the gnd gene we observe strong purifying selection to about position 900. In all three groups, purifying selection is strongest in the N terminal domain (fig. 9a). A similar level of purifying selection is observed in the N terminal region of the second domain in S. enterica and E. coli (14) sets, while

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nucleotide Diversity, π, of Escherichia coli and Salmonella enterica Genes</strong></td>
</tr>
<tr>
<td>Gene</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>gnd</td>
</tr>
<tr>
<td>gapA</td>
</tr>
<tr>
<td>putP</td>
</tr>
<tr>
<td>phoA</td>
</tr>
<tr>
<td>celC</td>
</tr>
<tr>
<td>crr</td>
</tr>
<tr>
<td>gutB</td>
</tr>
</tbody>
</table>

* No homologue of phoA gene in S. enterica.
in the *E. coli* (16) set recombination has introduced a slightly higher level of amino acid substitutions. The $K_A$ values in all three groups suggest that the region at about 735–880 bp is highly conserved, and this is the region where a putative substrate binding sequence is located (Reizer et al. 1991). The region from 900 to 1200 bp has the highest amino acid diversity and therefore the lowest level of purifying selection, presumably associated with reduced functional and structural constraints in this region.

The $K_A$ values beyond 900 bp in the *S. enterica* and *E. coli* (16) sets are similar and twice that in the *E. coli* (14) set. In the *E. coli* (16) set, the increase is due to the inclusion of highly divergent *gnd* sequences of ECOR 4 and ECOR 16 that we have shown above to result from interspecies recombination. In *S. enterica*, although 34 sequences from seven different subspecies have been analyzed, direct evidence for interspecies recombination has not been forthcoming. Interspecies recombination in the adjacent *rfb* locus (Reeves 1993) and cotransfer of the *gnd* gene with the *rfb* cluster (Achtman and Pluschke 1986) has been previously shown, and the apparent lack of interspecies recombination in the *S. enterica gnd* gene was initially puzzling. Further analysis of the *gnd* gene has shown that the region up to 900 bp could be subjected to efficient dam methylase directed mismatch repair, which would act like an antirecombinant during the processing of the recombinant intermediates and reduce the possibility of interspecies recombination in *S. enterica gnd* gene (G. Thampapillai and P. R. Reeves, unpublished data).

However, in the region from 900 to 1200 bp there are clustered amino acid changes in *S. enterica* (fig. 4). These changes may imply that lateral transfer of small segments of DNA from other related species does occur in *S. enterica* but could be limited to a small region. If interspecies recombination has contributed to the increase in $K_A$ in this region in *S. enterica*, as in the *E. coli* (16) set, then the dissimilarity of $K_S$ values in the two needs to be explained, as $K_S$ for *E. coli* (16) is twice that for *S. enterica*. In the *E. coli* (16) set, while selection pressure has reduced $K_A$ in positions up to about 900 bp, $K_S$ appears to be unaffected throughout and has increased the nucleotide diversity in the *E. coli* (16) set to twice that in the *E. coli* (14) set. The low $K_S$ values in both regions in the *gnd* gene in *S. enterica* therefore implies that unlike in the *E. coli* (16) set, synonymous sites in *S. enterica* are also subjected to negative selection in at least the first 900 bp. The constraint on synonymous sites in the *S. enterica gnd* gene could be related to the conservation of DNA sequences, which are recognized by proteins involved in recombinational repair, including the chi site. The possible reasons for increase in synonymous sites in the *E. coli gnd* genes will be discussed elsewhere.

Although the synonymous sites in general are not constrained in the *E. coli gnd* gene, the synonymous sites in nearly the first hundred base pairs are conserved both in *E. coli* and *S. enterica* (fig. 9b), and a reduction in $K_S$ at the beginning of the gene has also been observed in other enterobacterial genes (Eyre-Walker and Bulmer 1993). Gene expression studies in *E. coli* K-12 have

---

**Table 5**

$K_S$, $K_A$, and $K_A/K_S$ for *Salmonella enterica*, *Escherichia coli*, and Subspecies of *S. enterica* and between Species and Subspecies

<table>
<thead>
<tr>
<th>Species or Subspecies</th>
<th>$K_S$</th>
<th>$K_A$</th>
<th>$K_A/K_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>16–1344 bp:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. enterica</em></td>
<td>0.153 ± 0.054</td>
<td>0.011 ± 0.004</td>
<td>0.071</td>
</tr>
<tr>
<td><em>E. coli</em> (16)</td>
<td>0.313 ± 0.271</td>
<td>0.011 ± 0.009</td>
<td>0.035</td>
</tr>
<tr>
<td><em>E. coli</em> (14)</td>
<td>0.163 ± 0.038</td>
<td>0.006 ± 0.002</td>
<td>0.036</td>
</tr>
<tr>
<td><em>E. coli</em> (16)/S. ea*</td>
<td>0.872</td>
<td>0.023</td>
<td>0.027</td>
</tr>
<tr>
<td><em>S. enterica</em> ssp. I</td>
<td>0.058 ± 0.016</td>
<td>0.005 ± 0.003</td>
<td>0.086</td>
</tr>
<tr>
<td><em>S. enterica</em> ssp. II</td>
<td>0.103 ± 0.010</td>
<td>0.011 ± 0.003</td>
<td>0.106</td>
</tr>
<tr>
<td><em>S. enterica</em> ssp. I/II*</td>
<td>0.159</td>
<td>0.011</td>
<td>0.065</td>
</tr>
<tr>
<td>16–900 bp:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. enterica</em></td>
<td>0.140 ± 0.059</td>
<td>0.006 ± 0.003</td>
<td>0.042</td>
</tr>
<tr>
<td><em>E. coli</em> (16)</td>
<td>0.304 ± 0.249</td>
<td>0.008 ± 0.007</td>
<td>0.026</td>
</tr>
<tr>
<td><em>E. coli</em> (14)</td>
<td>0.168 ± 0.004</td>
<td>0.004 ± 0.002</td>
<td>0.023</td>
</tr>
<tr>
<td>901–1344 bp:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>S. enterica</em></td>
<td>0.182 ± 0.077</td>
<td>0.021 ± 0.011</td>
<td>0.113</td>
</tr>
<tr>
<td><em>E. coli</em> (16)</td>
<td>0.341 ± 0.333</td>
<td>0.018 ± 0.015</td>
<td>0.052</td>
</tr>
<tr>
<td><em>E. coli</em> (14)</td>
<td>0.156 ± 0.054</td>
<td>0.009 ± 0.004</td>
<td>0.057</td>
</tr>
</tbody>
</table>

* Comparison between *E. coli* (14 strains) and *S. enterica*.
* Comparison between subspecies I and II of *S. enterica*. 

---

Molecular Variation at the *gnd* Locus of *Salmonella enterica* 825
shown the presence of a 16-bp internal negative control element at codons 69–74 (Carter-Muenchau and Wolf 1989), which may also contribute toward additional selection pressure at the synonymous sites in the gnd gene.

We next compare the ratio of $K_A$ to $K_S$ for interspecies comparisons. The ratio $K_A/K_S$ within $S. enterica$ has about twice the value found within $E. coli$, and the value for the interspecies comparison between $E. coli$ and $S. enterica$ is even lower than that for $E. coli$. The high level of $K_A$ within each species could be due to the clonal nature of bacterial populations. It has been argued elsewhere (Reeves 1992) that these clones can be niche adapted and long-lived and that $N_e$, the effective population size is low for such clones. Under these circumstances most sequence polymorphisms in bacteria arise by fixation of new alleles within clones. The low value of $N_e$ in individual clones will increase the proportion of replacement substitutions that are fixed, as with low population size mildly deleterious mutations can behave like neutral alleles (Ohta 1973). There is indeed evidence from the distribution of replacement substitutions in the gnd gene of $E. coli$ that some replacement substitutions are mildly deleterious (Sawyer et al. 1987). However, while low population sizes of clones will lead to fixation of mildly deleterious alleles and increase the ratio of $K_A/K_S$, this effect would not be expected for fixation of substitutions between species, as that requires that a mildly deleterious form be fixed in all clones, after transfer between clones. Low levels of selective disadvantage, which would not prevent fixation within a clone, would nonetheless prevent it being fixed in all clones, $N_e$ for the species being much larger than for a clone.

In the comparison of the ratio of $K_A$ to $K_S$ between subspecies, the same trend is observed as between $E. coli$ and $S. enterica$. The ratios within subspecies I and II are 0.086 and 0.106, respectively, yet the ratio between subspecies I and II of $S. enterica$ is 0.065 (table 5). This implies that the mechanism operating to give a lower interspecies than intraspecies $K_A/K_S$ also operates at the subspecies level.

Conclusions

The phylogenetic tree constructed with the gnd gene sequences of 34 strains of Salmonella enterica agrees closely with the gene trees generated for the gapA and putP genes (Nelson et al. 1991; Nelson and Selander 1992). There has been considerable recombination in the gnd gene, and these events did have an effect on the gene tree. The DNA sequences of some of the recombinants show the involvement of a variant of the chi sequence in stimulating these recombinations. To our knowledge, this is the first time the chi site has been associated with natural recombination events in $S. enterica$ in a manner that matches the chi-dependent recombination events observed under laboratory conditions in $S. enterica$ or Escherichia coli (Smith 1991).

The polymorphism observed in the rfb cluster has major effects on antigenicity. This cluster is also highly mobile, presumably under selection to generate diversity. The presence of two chi sites in the gnd gene of $S. enterica$ (424–451 and 744–751 bp), which is located immediately 3' to the rfb locus, may reflect occasional cotransfer of parts of gnd gene with segments downstream of rfb gene cluster.

The pattern of recombination seems to be rather different in the gnd genes of $E. coli$ and will be discussed elsewhere. Chi activity is not detectable in the $E. coli$ gnd gene; a similar observation was made by Dykhuizen and Green (1991). A biotype-based subspecies structure as defined for $S. enterica$ is not available for $E. coli$. This would make detection of whole and partial gene transfers in $E. coli$ genes more difficult than in $S. enterica$. The transfer of segments observed in the $S. enterica$ gnd gene seems to be preferentially from subspecies I to other subspecies and not vice versa. The presence of subspecies I DNA in subspecies V in gnd and perhaps putP genes is interesting and makes the pair good candidates for
population studies, especially in the areas of selection and origin of subspecies.

**Sequence Availability**

The sequences have been deposited in GenBank with accession numbers U14336 to U14369.

**Acknowledgments**

This work was supported by a grant from the Australian Research Council. G.T. was a recipient of an Australian postgraduate research scholarship. We thank those referred to in table 1 for their kind donation of strains listed in that table.

**LITERATURE CITED**


JULIAN P. ADAMS, reviewing editor

Received May 31, 1994

Accepted June 16, 1994