Gene Transfer is a Major Factor in Bacterial Evolution

Ruiting Lan and Peter R. Reeves

Department of Microbiology, University of Sydney

Lateral gene transfer in four strains of *Salmonella enterica* has been assessed using genomic subtraction. Strain LT2 (subspecies I serovar Typhimurium) chromosomal DNA was used as target and subtracted by three subspecies I strains of serovars Typhimurium (S21), Muenchen (S71), Typhi (M229), and a subspecies V strain (M321). Data from probing random cosmids of LT2 DNA with preparations of the residual LT2 DNA after subtraction were used to estimate the amounts of LT2 DNA not able to hybridize to strains S21, S71, M229, and M321 to be in the range of 84–106, 191–355, 305–629, and 778–1,286 kb, respectively. Several lines of evidence indicate that most of this DNA is from genes not present in strain M321 and not from genes that have diverged in sequence. The amounts correlate with the divergence of the four strains as revealed by multilocus enzyme electrophoresis and sequence variation of housekeeping genes. Sequence of 39 of the fragments from the M321 subtracted residual LT2 DNA revealed only six inserts of known gene function with evidence of both gain and loss of genes during the development of *S. enterica* clones. Sixteen of the 39 segments have 45% or lower G+C content, below the species average, but over half are within the normal range for the species. We conclude that even within a species, clones may differ by up to 20% of chromosomal DNA, indicating a major role for lateral transfer, and that on the basis of G+C content, a significant proportion of the DNA is from distantly related species.

Introduction

Bacterial populations of many species, and in particular those of *Escherichia coli* and *Salmonella enterica,* are clonal as shown by multilocus enzyme electrophoresis (MLEE) (Selander, Caugant, and Whittam 1987; Selander, Beltran, and Smith 1991). Sequencing of housekeeping genes has revealed limited transfer of segments of DNA between clones (Bisercic, Feutrier, and Reeves 1991; Nelson and Selander 1992; Boyd et al. 1994; Thampapillai, Lan, and Reeves 1994). However, studies of genes involved in clone adaptation revealed far more extensive gene transfers, examples being the *rfb* gene cluster which determines the O antigen variation (Reeves 1993) and *fliC*, a flagellar antigen gene (Smith, Beltran, and Selander 1990). Comparative studies of *E. coli* K-12 and *S. enterica* LT2 chromosomes have indicated that large changes have occurred since divergence of the two species (Riley and Sanderson 1990). Lateral transfer in *S. enterica* is exemplified by the study of Fitts (1985) who systematically screened an *S. enterica* serovar (sv) Typhimurium *Bam*HI plasmid library and found that 54 of 200 clones probed to *S. enterica* LT2 DNA but not *E. coli* K-12 DNA. From a limited test of *S. enterica* and non-*S. enterica* isolates 10 of the 54 clones were found to be unique to *S. enterica*. Groisman et al. (1993) studied five of those fragments in detail. Three of them had G+C contents con-

siderably lower than that of the *S. enterica* genome suggesting horizontal transfer.

The above observations suggest that lateral gene transfer may play a major role in the evolution of bacteria. There is, however, no comprehensive assessment of the extent of lateral gene transfer but this can now be made easily by genomic subtraction, a procedure developed by several groups to enrich sequences that are present in one genome but not in another (Straus and Ausubel 1990; Wieland et al. 1990). Although the methods differ in detail, they are based on the same principles. In the method of Straus and Ausubel, restriction enzyme-digested target DNA is mixed with excess sheared, biotinylated subtracter DNA. The mixture is denatured and hybridized. Biotinylated DNA and with it any target DNA which hybridized to it is removed using avidin beads. Several cycles of hybridization with newly added biotinylated subtractor DNA removes target DNA with sequences present in both target and subtracter strains. The remaining unbound target DNA is enriched in sequences absent in the subtracter DNA. This DNA is then ligated to an adaptor and amplified by polymerase chain reaction (PCR) for further analysis.

In this paper, we report on the extent of lateral gene transfer in the evolution of *S. enterica* as studied by genomic subtraction. *Salmonella enterica* has a well-defined subspecies structure and a clonal population structure with deep branch lengths as revealed by MLEE (Reeves et al. 1989; Selander and Smith 1990) and sequencing of housekeeping genes (Nelson, Whittam, and Selander 1991; Nelson and Selander 1992; Boyd et al. 1994; Thampapillai, Lan, and Reeves 1994). We selected four strains with various levels of divergence from Typhimurium LT2. We estimated the amount of non-

homologous DNA by probing LT2 cosmids with resid-
ual DNA and sequenced some randomly selected clones
of residual DNA to study the nature of the DNA in-
volved.

Materials and Methods
Strains

Salmonella eneterica sv Typhimurium strain LT2
(SGSC1412) was obtained from the Salmonella Genetic
Stock Center, Canada. This is the wild type originally
isolated by Lilleengen (1948). The 90-kb pSLT plasmid
present in the strain was removed using a pSLT removal
kit provided by Dr. K. E. Sanderson of the Salmonella
Genetic Stock Center. The derived plasmidless strain
(P9336) was used as the target strain for subtraction.

Other strains: S21: sv Typhimurium (SARA 21 of
S. enterica reference A set [Beltran et al. 1991]); S71:
sv Muenchen (SARA71 of *S. enterica* reference A set
[Beltran et al. 1991]); m229: sv Typhi (laboratory col-
lection); M321: subspecies V (sv Balboa, L. le Minor,
Institut Pasteur, France).

General DNA Manipulations

Chromosomal DNA and plasmid DNA were iso-
lated by standard procedures (Sambrook, Fritsch, and
Maniatis 1989). The LT2 cosmid bank was constructed
as described by Bastin, Romana, and Reeves (1991).
The sizes of cosmid clones were determined from *Eco*RI
digests and large fragments were resolved using pulse
field gel electrophoresis.

Subtractive Hybridization

The method of Straus and Ausubel (1990) was used
with the following minor modifications. Hybridization
was done in phenol emulsion at room temperature (Koh-
ne, Levison, and Byers 1977). Removal of biotinylated
DNA after hybridization was achieved using phenol–
chloroform extraction (Sive and St John 1988).

Subtracter DNA was sonicated and labeled with
photobiotin (Bresatec). In the first cycle of subtraction,
0.3 μg of *Sau*3AI-digested target DNA was mixed with
10 μg of biotinylated DNA in a 0.5-ml tube and the
mixture denatured by boiling for 4 min. Twenty-five mi-
croliters of 2 M NaOAc and 8 μl of Tris-EDTA (TE)-
saturated phenol were added. Hybridization was at room
temperature for 24 h with constant shaking using a vor-
texer (Vortex-genie, Scientific Industries, Mass.) at set-
ting 6. Biotinylated DNA was removed using streptav-
idin and phenol extraction (Sive and St John 1988). Sub-
sequent rounds of subtraction were done as for the first
cycle except that residual DNA from the preceding cycle
rather than original target DNA was added. The final
residual DNA was dissolved in 10 μl TE.

PCR Amplification after Subtractive Hybridization

Unbound DNA fragments (0.5 μl) from the final
subtractive hybridization cycle was ligated to a *Sau*3AI
adaptor to add a segment to act as template for subse-
quent PCR. The adaptor sequences and PCR conditions
were as described by Straus and Ausubel (1990). The
PCR product was purified by Sephadex G50 column
chromotography (Pharmacia) and used for probing or
cloning. For cloning, the DNA was end filled using the
Klenow fragment of DNA polymerase I and ligated into
the *Sma* I site of pUC18.

Southern Hybridization

Two micrograms of chromosomal or cosmid DNA
was digested with appropriate restriction enzymes
(Boehringer Mannhein), fractionated by electrophoresis
on 1% agarose, and transferred to nylon membranes
(Hybond+, Amersham) by vacuum blotting using
VacuGene (Pharmacia) according to manufacturers' in-
structions. Probes were labeled by random priming. Hy-
bridization was done with 50% formamide, 6× SSPE,
0.5% sodium dodecyl sulfate (SDS), 5× Denhardt's so-
lution, and 0.1 μg salmon sperm DNA at 42°C over-
night. Lower stringency hybridization was done by re-
ducing the amount of formamide. Reduced stringency,
as referred to later in the text, was using 40% formamide
while low stringency was using 30% formamide.

Sequence Analysis

Purified plasmid DNA was used as template and
the sequence was determined in one direction by the dye
primer method using a thermal cycler (Perkin-Elmer Ce-
tus) and an automated 373A DNA sequencer (Applied
Biosystems) according to the manufacturers' instruc-
tions.

Searching of DNA and protein databases was done
using the FASTA (Pearson and Lipman 1988) and
BLAST (Altschul et al. 1990) programs in the Austra-
lian Genomic Information Service (ANGIS).

Results
Efficacy of Genomic Subtraction in Removing Common DNA

To assess the extent of lateral transfer in the evo-
lution of *S. enterica,* we chose four strains, one from
each of four serovars: Typhimurium (S21), Muenchen
(S71), Typhi (M229), and a subspecies V strain (M321)
to subtract DNA from *S. enterica* Typhimurium LT2, a
strain that has been widely used for genetic analysis.
The 90-kb plasmid in LT2 was removed in order to
avoid the plasmid being a major contributor of residual
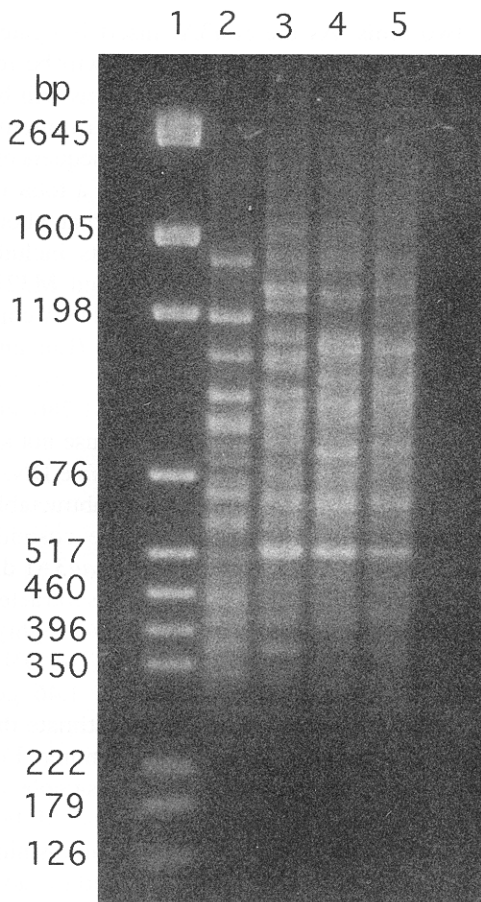DNA after subtraction.

Fig. 1.—*Salmonella enterica* LT2 chromosomal DNA remaining after subtraction by other *S. enterica* strains and amplified by PCR. Lane 1, pGEM® molecular weight marker; lanes 2, 3, 4, and 5: subtraction by DNA from strains S21, S71, M229, and M321, respectively.
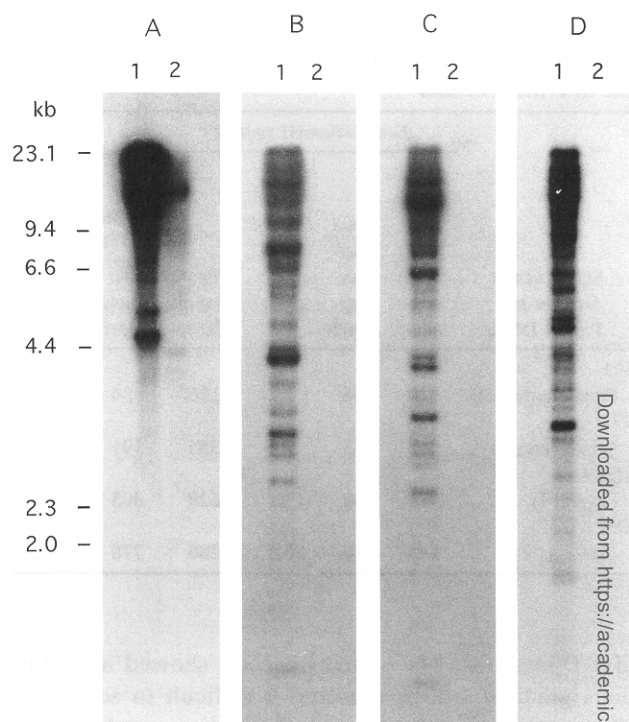

Fig. 2.—Probing of *Eco*RI-digested LT2 and subtracter strain DNA by LT2 DNA remaining after subtraction. DNA in parts A, B, C, and D was probed by S21-, S71-, M229-, and M321-subtracted LT2 DNA, respectively. Lane 1, LT2; lane 2, respective subtracter strain. The sizes and positions of λ *Hind*III molecular weight markers are indicated at the left.

After four rounds of subtraction, remaining LT2 DNA was amplified by PCR and an agarose gel fractionation of the PCR product is shown in figure 1. In all four cases, fragments showing apparent enrichment of nonsubtractable DNA are obvious on the gel. From inspection of figure 1, the size of DNA fragments ranges from 250 bp to 2 kb. To confirm that residual DNAs were unique, total residual DNA was labeled and shown to hybridize to LT2 DNA strongly but very weakly or not at all to the respective subtracter DNA even at reduced stringency (fig. 2). This indicates that subtraction has effectively removed most, if not all, DNA common to the two strains.

As a test for bias in PCR, we used a fragment in the *rfb* region. The LT2 *rfb* cluster (O group B) has no DNA in common with M321, the subspecies V strain (O group 48) (Xiang, Haase, and Reeves 1993). A cloned 9.75-kb *Eco*RI fragment from the LT2 *rfb* region was digested with *Sau*3AI and probed with total M321 subtracted residual LT2 DNA. As *Sau*3AI digests were used for subtraction and for blotting, there should be a one-to-one relationship, i.e., all fragments present in the probe DNA should be observed in the residual DNA by probing. The 13 *Sau*3AI fragments of 250 bp or greater expected from sequence information (Jiang et al. 1991) were observed in the autoradiogram (data not shown), indicating that the residual DNA has been amplified without undue bias.

Estimation of the Proportion of the Genome not Subtractable

A cosmid bank of LT2 DNA was constructed and used to estimate the proportion of the genome represented by the residual DNA. The average size of insert was 35.5 kb as determined from a random sample of clones. Six-hundred randomly picked cosmids were screened with the four residual DNA preparations by colony blotting. Unfortunately, except for Typhimurium S21, the number of positives was very difficult to determine because the heterogeneity of residual DNA gave a low signal-to-background ratio. An objective measure was attempted by quantifying the radioactivity associated with each colony using a phosphorimager (Molec-

Table 1
Estimation of Amount of *S. enterica* LT2 DNA not Present in Other Strains

| | EcoRI–BamHI DIGEST | | | | Sau3 AI DIGEST | |
SUBTRACTER STRAIN FOR PROBE DNA	Total positive cosmids	Total positive fragments	Percentage fragments positive	Per Genome (kb)	Per genome (kb)	Percentage
S21 (Typhimurium, I)	10	34	2.2	106	84	1.8
S71 (Muenchen, I) . . .	48	113	7.4	355	191	4.0
M229 (Typhi, I)	75	200	13.1	629	305	6.4
M321 (Balboa, V)	143	410	26.8	1286	778	16.2

ular Dynamics). The signal intensity showed a continuous gradient, which rendered it difficult to set a cut-off point between positives and negatives, and this approach was abandoned. Instead DNA from 200 randomly selected cosmids was extracted and probed by Southern blotting. This slightly compromised the coverage of the genome as 200 cosmids only cover 77% of the genome while 600 cover 99% (Sambrook, Fritsch, and Maniatis 1989). Each of the 200 cosmid DNA preparations was double digested with *Eco*RI and *Bam*HI and probed with the four residual DNAs. The results were much clearer than those from colony screening. One-hundred ninety-seven cosmids were scored for positive fragments as three preparations had too little DNA. There were cases in which signal was present but faint and such fragments were not counted as positive. Similarly, cosmids containing faint fragments only (usually one or two bands) were not counted as positive. The numbers of positive cosmids and positive fragments are summarized in table 1. In no case were all fragments of a cosmid positive. To estimate the proportion of the genome involved, we need to know the size of the positive fragments. As it is difficult to determine the size of each fragment, we simply divided the number of positive fragments by the total number of fragments to determine the proportion of positive fragments. Assuming the distribution of positive fragments is random and the number is relatively large, the estimate is unbiased. The number of fragments from *Eco*RI–*Bam*HI double digests was counted for the 197 cosmids. *Eco*RI–*Bam*HI digestion of the recombinant cosmid produces three vector-containing fragments (Jiang et al. 1987): an internal *Eco*RI fragment of less than 0.7 kb, which was not included in the count and two *Eco*RI–*Bam*HI segments

carrying the two arms. As the *Sau*3AI insert is ligated to *Bam*HI-digested vector, 25% of junctions will be reconstituted *Bam*HI sites and 75% of vector arms will be associated with insert DNA. Thus 98.5 fragments (2*0.25*197) are expected to be pure vector sequences. We subtracted 99 "vector fragments" to give a total of 1,528 insert fragments. We then estimated that about 2.2%, 7.4%, 13.1%, and 26.8% of fragments include nonsubtractable DNA in S21, S71, M229, and M321, respectively. Converting these percentages to DNA contents, assuming a genome size of 4,800 kb (Liu and Sanderson 1992; Wong and McClelland 1992), the amounts are 106 kb, 355 kb, 629 kb, and 1,286 kb, respectively. These will be overestimates because not all of each positive fragment will be nonsubtractable DNA.

We also estimated a lower limit for nonsubtractable DNA by probing *Sau*3AI digests of positive cosmids with respective residual DNA preparations. *Sau*3AI digests from all 10 cosmids positive with S21 subtracted residual LT2 DNA were probed with that DNA. Eighty-nine fragments totalling 122 kb were positive. The DNA present in the 197 cosmids is equivalent to 1.46 genomes (197 × 35.5/4,800), from which we estimate the *Sau*3AI-positive fragments to be 84 kb per genome. For the other three cases, we probed a random sample of 16–19 positive cosmids after digestion with *Sau*3AI. For S71, positive *Sau*3AI fragments in 16 positive cosmids totalled 93 kb and thus the estimate of nonsubtractable DNA in the 48 positive cosmids is 279 kb equivalent to 191 kb per genome. The 16 cosmids were also positive for M229-subtracted DNA and were probed to give 95 kb of positive fragments. This gave an estimate of 446 kb in the 75 positive cosmids or 305 kb per genome. For M321, the positive fragments totalled 151 kb from 19 cosmids to give an estimate of 778 kb per genome. In contrast to the estimates using *Eco*RI–*Bam*HI digests, use of *Sau*3AI digests gives an underestimate because *Sau*3AI fragments with only a segment of homologous DNA would have been removed by subtraction, and also fragments smaller than 250 bp appear not to be retained after subtraction (see fig. 1). The amounts of LT2 DNA not present in S21, S71, M229, and M321 are in the ranges of 84–106, 191–355, 305–629, 778–1,286 kb, respectively. In percentages they are in the range of 1.8%–2.2%, 4.0%–7.4%, 6.4%–13.1%, and 16.2%–26.8%, or approximately 2%, 5%, 9%, and 20%.

Sequence Analysis of M321 Subtracted, Residual LT2 DNA

We cloned M321-subtracted residual LT2 DNA into pUC18. Forty-six inserts, ranging from 460 bp to 1.76 kb, were found to hybridize to LT2 but not to M321 DNA. Two hundred to three hundred base pairs of sequence was obtained from each insert. Since the se-

Table 2
Properties of Genes Present in LT2 but not M321[a]

Insert	Gene	Evidence	Function	Nature of Event[b]
1	*nrfB*	13% difference to gene in K-12 (Hussain et al. 1994)	Formate dependent nitrite reductase	Loss
2	*gcl*	12% difference to gene in K-12 (Chang, Wang, and Cronan 1993)	Glyoxylate carboligase	Loss
3	Unknown	23% difference to gene in K-12 (Sofia et al. 1994)	Homology to *E. coli* ORF o155	Loss
4	*citC*	Gene in LT2 but not K-12 (Ishiguro et al. 1992)	Citrate/isocitrate carrier protein	Gain
5	*pgtB*	Gene in LT2 but not K-12 (Yang, Goldrick, and Hong 1988)	Phosphoglycerate transport regulatory protein	Gain
6	*hsdM*	Gene in LT2 and K-12 (Sharp et al. 1992)	Restriction modification	(?)
7	Prophage	Partial homology to phage T2 gene 37	(?)	(?)

[a] The six inserts with known functions were used to probe M321 at low stringency (30% formamide) to reveal any divergent homologue but no hybridization signal was detected.

[b] Genes recorded as present in *E. coli* K-12 and LT2 are assumed to have been lost in the M321 lineage; those absent in K-12 are assumed to have been gained in the LT2 lineage.

quencing was done in only one direction the accuracy is not 100% and some bases were not resolved. Seven inserts are duplicates so only 39 are unique. The sequences were used to search against GenBank using FASTA, and six frames of amino-acid translation were used to search against the nonredundant peptide databases using BLAST. Matches were found for seven sequences (table 2), only four of which matched sequences in K-12, although over 60% of its genome is now sequenced. Of these four, *nrfB* (formate-dependent nitrite reductase) and *gcl* (glyoxylate carboligase) relate to nonhousekeeping functions and are presumed be lost in M321; *hsdM* is known to be polymorphic (Sharp et al. 1992), so presumably M321 and LT2 have different forms; we do not know the function of open reading frame (ORF) o155. Two other genes, *citC* (citrate/isocitrate carrier protein gene) and *pgtB* (phosphoglycerate transport regulatory protein gene) have presumably been gained by LT2. The *citC* gene is present in several serovars (Yang, Goldrick, and Hong 1988), so it could have been gained a long time ago. *Escherichia coli* lacks the phosphoglycerate transport system (Saier et al. 1975), and the two genes were probably gained by the ancestors of LT2 after divergence from subspecies V.

Discussion

Genomic subtraction has been used for example to clone genes corresponding to deletions in plants (Sun, Goodman, and Ausubel 1992) and for isolating species-specific probes in bacteria (Bjourson, Stone, and Cooper 1992). Three to five cycles of hybridization are used in most of the experiments reported. We have used genomic subtraction for study of nonhomologous DNA and found it to be a very useful approach. The small genome size of bacteria probably facilitated enrichment of nonhomologous sequences. We arbitrarily chose four cycles after a pilot experiment of three-cycle hybridization using an *rfb* delete LT2 strain to subtract against the wild-type strain. However, we did not study efficiency of the subtraction in detail. Probing of target and subtracter DNA using the residual DNA was used as an indicator of purity of residual DNA.

Our aim is to estimate the extent of lateral transfer by measuring the amount of DNA in one strain not present in another. However, residual DNA after subtraction may not contain all such DNA, causing underestimation. A *Sau*3AI digest of LT2 DNA was used as target DNA for subtraction and any *Sau*3AI fragments with homologous DNA in addition to nonhomologous DNA would have been removed during the subtraction process. Small insertion/deletions may escape detection completely as they will not be detected if they do not contain internal *Sau*3AI fragments larger than 250 bp. We studied the distribution of *Sau*3AI sites using DNA sequences of *S. enterica* genes in GenBank. Of 130 genes of 1 kb or larger, only two (1.5%) have no fragment larger than 250 bp. Therefore the proportion of nonhomologous regions/gene segments not detected in these experiments is probably very small. Also fragments smaller than 250 bp appear to be lost during subtraction, which does not affect the upper estimation by *Eco*RI–*Bam*HI digest as long as most fragments of the digest contain at least one *Sau*3AI fragment greater than 250 bp, but causes underestimation of the lower bound of *Sau*3AI digest. The extent of underestimation is about 30% from analysis of *Sau*3AI fragments of *E. coli* and

S. enterica sequence databases. Finally bias in PCR amplification from residual DNA would affect our estimation of nonhomologous DNA. It is known that PCR amplifies preferentially smaller size fragments (Straus and Ausubel 1990; Wieland et al. 1990). Wieland et al. (1990) also reported that PCR does not amplify every fragment. In the case of the *rfb* genes in LT2 known not to be present in M321, we found that all expected *Sau*3AI fragments of 250 bp or more were present in the residual DNA, indicating no bias.

Residual DNA may be either not homologous to DNA in the driver strain or divergent in sequence such that hybridization did not occur. Our aim is to detect the amount of nonhomologous DNA and we took several approaches to eliminate the possibility that a significant part of the residual DNA is homologous but divergent. The M321-subtracted DNA is of particular concern as subspecies V and I are the most divergent. The divergence between subspecies I and V ranges from 4.3% to 9.0% for housekeeping genes sequenced in both (Nelson, Whittam, and Selander 1991; Nelson and Selander 1992; Boyd et al. 1994; Thampapillai, Lan, and Reeves 1994), so divergence is unlikely to prevent subtraction between LT2 and M321. Nonetheless we tested the M321-subtracted residual DNA by probing M321 at reduced stringency and observed no hybridization. As expected the *E. coli mdh* (malate dehydrogenase) gene probe hybridized to LT2 DNA readily at high stringency (data not shown).

We also used the six cloned fragments with known functions (table 2) to probe M321 by Southern hybridization at low stringency, but none hybridized at all, ruling out any reasonable possibility of divergence. Further, we can treat as a null hypothesis the proposal that bacterial evolution occurs only by sequence divergence (with or without duplication). In that case all genes in LT2 should be present in *E. coli* K-12, as well as M321, the nonsubtraction of 20% by M321 being due to sequence divergence: on this hypothesis we would expect 23 (60%) of the 39 sequences to match K-12 sequences in GenBank, but only four did and the three of known function probably represent gain or loss of DNA. We conclude that the residual DNA isolated is essentially nonhomologous DNA.

DNA–DNA solution reassociation has been extensively used to define relationships of strains of the same or closely related bacterial species. For within-species comparisons 60%–100% of labeled DNA of one strain hybridizes to DNA of the other, while for closely related species 20%–60% reassociates (Krieg and Holt 1984). As for residual DNA after subtraction, the DNA that does not hybridize comprises one or both of nonhomologous DNA and DNA of genes that have diverged substantially. Crosa et al. (1973) obtained reassociation values for LT2 against two other Typhimurium strains of 93% and 98%, and for LT2 against Muenchen and Typhi values of 92% and 88%, respectively. The unhybridized proportions are strikingly similar to our estimate of the percentage nonhomologous DNA between LT2 and S21 (2%), S71 (5%), and M229 (9%). We suggest that most or even all of the unhybridized DNA is nonhomologous DNA rather than DNA of low homology. Unfortunately the data available on DNA–DNA hybridization between LT2 and subspecies V (Le Minor, Veron, and Popoff 1982) are from a study using S1 nuclease and not comparable to our subtraction method. Our conclusion should be extrapolated cautiously to comparisons between more distantly related strains. DNA–DNA reassociation between *S. enterica* LT2 and *E. coli* K-12 is 46% (Crosa et al. 1973). This suggests that the *E. coli* and *S. enterica* genomes differ by 54%. Given that of 67 pairs of genes sequenced in both species 90% had more than 79% identity (Sharp 1991), it is probable that most homologous DNA would be able to associate and it is quite likely that a major portion of the 54% of DNA that does not reassociate is nonhomologous DNA. However, for between-species comparisons in general, the DNA–DNA data would represent both divergent sequences and laterally transferred sequences: the proportion that is nonhomologous DNA depending on the distance between the two species.

This study is the first attempt to assess the overall level of lateral gene transfer in bacteria by quantitation of nonhomologous DNA. There are of course many reported examples of lateral gene transfer in bacterial evolution. Comparison of *E. coli* K-12 and *S. enterica* LT2 linkage maps has indicated that there are 14 segments in K-12 but not LT2 and 15 segments in LT2 but not K-12 (Riley and Sanderson 1990). These segments harbor genes encoding metabolic properties differentiating the two species. Comparative restriction mapping of two *S. enterica* strains using rare cutting enzymes (Liu et al. 1994) also revealed large differences in their genomes: the *Xba*–*Bln*I–*Ceu*I map differs in several instances in interval length between Typhimurium LT2 and Paratyphi B, indicating insertion/deletions. For example, the region between *pncB* and *pyrD* is 45 kb in LT2 but only a few kilobases in Paratyphi B. The Paratyphi B strain they used is approximately the same genetic distance as Muenchen from LT2, as measured by MLEE (Beltran et al. 1991).

Nonhomologous DNA consists of genes gained by the target strain or lost in the subtracter strain after their divergence. Assuming gain or loss to be equally probable, of the 20% LT2 DNA not in M321, half would have undergone lateral transfer to LT2 and half lost in the subtracter strain. Six of the 39 segments sequenced were from genes homologous to genes in GenBank, and

in some cases we could infer from their distribution whether they had been lost in M321 or gained in LT2. Residual DNA may also contain prophages, IS elements, and repetitive DNA. One of the 39 segments sequenced may be of a prophage origin. It is perhaps not surprising that none of the inserts was identified as IS sequence or repetitive DNA as the two strains may harbor the same classes of IS sequence or repetitive DNA.

We also looked at the G+C content of the 39 sequences. It is known that the G+C content of bacterial genomes is relatively homogeneous within a species but varies between species as a result of directional mutation pressure (Sueoka 1992). Unusual G+C composition is taken as evidence of horizontal gene transfer (Groisman et al. 1993; Reeves 1993; Riley 1993; Whittam and Ake 1993; Ochman and Lawrence 1995). Sixteen of the 39 sequences have G+C contents from 31% to 45%, far lower than species average (52%), indicating that LT2 gained those genes from species of low G+C content. The remaining 23 sequences, including all 7 from identified genes, have a G+C content typical of *S. enterica*, which of course does not exclude the possibility of lateral transfer.

It seems clear that most of the residual LT2 DNA after subtraction by M321 represents genes present in LT2 and absent in M321. We have discussed it in the context of lateral transfer, but for transfers from within *S. enterica*, we could treat the loci as polymorphic. Given the low frequency of recombination in *S. enterica* as indicated by the congruence of individual gene trees (Nelson, Whittam, and Selander 1991; Nelson and Selander 1992; Boyd et al. 1994; Thampapillai, Lan, and Reeves 1994) the term lateral transfer is properly applied to intraspecies transfer of alleles of polymorphic loci. An example would be the *rfb* locus. Those *rfb* loci that have been studied have apparently transferred to *S. enterica* since divergence from *E. coli* (Reeves 1993), but most O antigen forms are present in several subspecies and hence have also undergone lateral transfer within the species. It will require further study to determine the intra- and interspecies dynamics of the genes that undergo lateral transfer.

Concluding Comments

Lateral transfer of nonhomologous DNA in which new properties are acquired from outside the lineage contributes to evolutionary changes in bacteria. In this study we show that lateral gene transfer plays a quantitatively important role in the development of *S. enterica* subspecies and clones. LT2 differs from four other *S. enterica* strains, Typhimurium S21, Muenchen S71, Typhi M229, and subspecies V strain M321 by nonsubtractable DNA of 2%, 5%, 9%, and 20%, respectively.

The finding that there is a very high level of lateral transfer in bacteria suggests that it plays a major role in their evolution. Consistency of trees based on 16s RNA or housekeeping proteins that are generally present (Lawrence, Ochman, and Hartl 1991; Woese 1987) shows clearly that there is a core of genes generally not subject to such transfer. The genes subject to lateral transfer must be those that are not required under all circumstances and hence may be lost or, in circumstance where there is strong selection pressure for their presence, may be gained by lateral transfer. We postulate a useful distinction between "housekeeping" genes for functions required for all or most strains and species of a group and "niche adaptive" genes relating to functions present in some strains or species only, which survive only while the function is relevant. The former will only rarely be subject to lateral transfer as there is no selective advantage in substituting one form for another, but the latter will be subject to selection for lateral transfer if circumstances of a clone or species change. Transfer of nonhomologous DNA with acquisition of new properties is probably characteristic of the evolution of bacteria, in which many of the properties that differentiate species or subspecies are due to presence or absence of specific functions and hence of the genes that encode them (see for example Ewing [1986] for tables of properties of enterobacterial species). High levels of gene transfer may facilitate rapid adaptation of bacteria to any new niche once available and the selective pressures involved have been discussed recently (Reeves 1992). This situation is in marked contrast to that which prevails at least in more complex eukaryotes where most evolution involves developmental process and is probably due to changes in regulation of gene and gene product expression rather than gain of new gene functions. One can speculate that the continued success of the bacteria is due in part to their ability to adapt by gene capture.

Acknowledgment

LITERATURE CITED

ALTSCHUL, A. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN. 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410.

BASTIN, D. A., L. K. ROMANA, and P. R. REEVES. 1991. Molecular cloning and expression in *Escherichia coli* K-12 of the *rfb* gene cluster determining the O antigen of an *E. coli* O111 strain. Mol. Microbiol. **5**:2223–2231.

BELTRAN, P., S. A. PLOCK, N. H. SMITH, T. S. WHITTAM, D. C. OLD, and R. K. SELANDER. 1991. Reference collection

of strains of the *Salmonella typhimurium* complex from natural populations. J. Gen. Microbiol. **137**:601–606.

BISERCIC, M., J. Y. FEUTRIER, and P. R. REEVES. 1991. Nucleotide sequence of the *gnd* gene 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.

BJOURSON, A. J., C. E. STONE, and J. E. COOPER. 1992. Combined subtraction hybridization and polymerase chain reaction amplification procedure for isolation of strain-specific *Rhizobium* DNA sequences. Appl. Environ. Microbiol. **58**:2296–2301.

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.

CHANG, Y.-Y., A.-Y. WANG, and J. E. CRONAN. 1993. Molecular cloning, DNA sequencing and biochemical analyses of *Escherichia coli* glyoxylate carboligase: an enzyme of the acetohydroxy acid synthase–pyruvate oxidase family. J. Biol. Chem. **268**:3911–3919.

CROSA, J. H., D. J. BRENNER, W. H. EWING, and S. FALKOW. 1973. Molecular relationships among the Salmonellae. J. Bacteriol. **115**:307–315.

EWING, W. H. 1986. Edwards and Ewing's identification of the Enterobacteriaceae. Elsevier Science Publishers, Amsterdam.

FITTS, R. 1985. Development of a DNA–DNA hybridization test for the presence of *Salmonella* in foods. Food Technol. **39**:95–102.

GROISMAN, E. A., M. A. STURMOSKI, F. R. SOLOMON, R. LIN, and H. OCHMAN. 1993. Molecular, functional and evolutionary analysis of sequences specific to *Salmonella*. Proc. Natl. Acad. Sci. USA **90**:1033–1037.

HUSSAIN, H., J. GROVE, L. GRIFFITHS, S. BUSBY, and J. COLE. 1994. A seven-gene operon essential for formate-dependent nitrite reduction to ammonia by enteric bacteria. Mol. Microbiol. **12**:153–163.

ISHIGURO, N., H. IZAWA, M. SHINAGAWA, T. SHIMAMOTO, and T. TSUCHIYA. 1992. Cloning and nucleotide sequence of the gene (*cit*C) encoding a citrate carrier from several *Salmonella* serovars. J. Biol. Chem. **267**:9559–9564.

JIANG, X.-M., H. N. BRAHMBATT, N. B. QUIGLEY, and P. R. REEVES. 1987. A low copy number cosmid. Plasmid **18**:170–172.

JIANG, X. M., B. NEAL, F. SANTIAGO, S. J. LEE, L. K. ROMANA, and P. R. REEVES. 1991. Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). Mol. Microbiol. **5**:695–713.

KOHNE, D. E., S. A. LEVISON, and M. J. BYERS. 1977. Room temperature method for increasing the rate of DNA reassociation by many thousandfold: the phenol emulsion reassociation technique. Biochemistry **16**:5329–5341.

KRIEG, N. R., and J. G. HOLT. 1984. Bergey's manual of systematic bacteriology. Williams & Wilkins, Baltimore.

LAWRENCE, J. G., H. OCHMAN, and D. L. HARTL. 1991. Molecular and evolutionary relationships among enteric bacteria. J. Gen. Microbiol. **137**:1911–1921.

LE MINOR, L., M. VERON, and M. POPOFF. 1982. Taxonomie des *Salmonella*. Ann. Microbiol. (Inst. Pasteur) **133B**:223–243.

LILLEENGEN, K. 1948. Typing *Salmonella typhimurium* by means of bacteriophage. Acta Pathol. Microbiol. Scand. Suppl **77**:11–125.

LIU, S.-L., A. HESSEL, H.-Y. M. CHENG, and K. E. SANDERSON. 1994. The *XbaI–BlnI–CeuI* genomic cleavage map of *Salmonella paratyphi* B. J. Bacteriol. **176**:1014–1024.

LIU, S.-L., and K. E. SANDERSON. 1992. A physical map of the *Salmonella typhimurium* LT2 genome made by using *XbaI* analysis. J. Bacteriol. **174**:1662–1672.

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., T. S. WHITTAM, and R. K. SELANDER. 1991. Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli*. Proc. Natl. Acad. Sci. USA **88**:6667–6671.

OCHMAN, H., and J. G. LAWRENCE. 1995. Phylogenetics and the amelioration of bacterial genomes. *In* F. C. NEIDHARDT, ed. *Escherichia coli* and *Salmonella typhimurium* cellular and molecular biology. (In press) American Society for Microbiology, Washington, D.C.

PEARSON, W. R., and D. J. LIPMAN. 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.

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*. J. Clin. Microbiol. **27**:313–320.

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.

RILEY, M. 1993. Functions of gene products of *Escherichia coli*. Microbiol. Rev. **57**:862–952.

RILEY, M., and K. E. SANDERSON. 1990. Comparative genetics of *Escherichia coli* and *Salmonella typhymurium*. Pp. 85–95 *in* K. DRLICA and M. RILEY, eds. The bacterial chromosome. American Society for Microbiology, Washington, D.C.

SAIER, M. H., D. L. WENTZEL, B. U. FEUCHT, and J. J. JUDICE. 1975. A transport system for phosphenolpyruvate, 2-phosphoglycerate, 3-phosphoglycerate in *Salmonella typhimurium*. J. Biol. Chem. **250**:5089–5096.

SAMBROOK, J., E. F. FRITSCH, and T. MANIATIS. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

SELANDER, R. K., P. BELTRAN, and N. H. SMITH. 1991. Evolutionary genetics of *Salmonella*. Pp. 25–27 *in* R. K. SELANDER, A. G. CLARK, and T. S. WHITTAM, eds. Evolution

at the molecular level. Sinauer Associates, Sunderland, Mass.

SELANDER, R. K., D. A. CAUGANT, and T. S. WHITTAM. 1987. Genetic structure and variation in natural populations of *Escherichia coli*. Pp. 1625–1648 *in* F. C. NEIDHARDT, J. L. INGRAHAM, B. MAGASANIK, M. SCHAECHTER et al., eds. *Escherichia* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.

SELANDER, R. K., and N. H. SMITH. 1990. Molecular population genetics of *Salmonella*. Rev. Med. Microbiol. **1**:219–228.

SHARP, P. M. 1991. Determinants of DNA sequence divergence between *Escherichia coli* and *Salmonella typhimurium*: codon usage, map position, and concerted evolution. J. Mol. Evol. **33**:23–33.

SHARP, P. M., J. E. KELLEHER, A. S. DANIAL, G. M. COWAN, and N. E. MURRAY. 1992. Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. Proc. Natl. Acad. Sci. USA **89**: 9836–9840.

SIVE, H. L., and T. ST JOHN. 1988. A simple subtractive hybridization technique employing photoactivable biotin and phenol extraction. Nucleic Acids Res. **16**:10937.

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.

SOFIA, H. J., V. BURLAND, D. L. DANIALS, G. PLUNKETT III, and F. R. BLATTNER. 1994. Analysis of the *Escherichia coli* genome. V. DNA sequence of the region from 76.0 to 81.5 minutes. Nucleic Acids Res. **22**:2576–2586.

STRAUS, D., and F. M. AUSUBEL. 1990. Genomic subtraction for cloning DNA corresponding to deletion mutations. Proc. Natl. Acad. Sci. USA **87**:1889–1893.

SUEOKA, N. 1992. Directional mutation pressure, selective constraints, and genetic equilibria. J. Mol. Evol. **34**:95–114.

SUN, T.-P., H. M. GOODMAN, and F. M. AUSUBEL. 1992. Cloning the *Arabidopsis GA1* locus by genomic subtraction. Plant Cell **4**:119–128.

THAMPAPILLAI, G., R. LAN, and P. R. REEVES. 1994. Molecular variation at the *gnd* locus of 34 natural isolates of *Salmonella enterica*: DNA sequence: evidence for probable chi-dependent interallelic recombination. Mol. Biol. Evol. **11**: 813–828.

WHITTAM, T. S., and S. E. AKE. 1993. Genetic polymorphisms and recombination in natural populations of *Escherichia coli*. Pp. 223–245 *in* N. TAKAHATA and A. G. CLARK, eds. Mechanisms of molecular evolution. Japan Scientific Societies Press, Tokyo.

WIELAND, I., G. BOLGER, G. ASOULINE, and M. WIGLER. 1990. A method for difference cloning: gene amplification following subtractive hybridization. Proc. Natl. Acad. Sci. USA **87**:2720–2724.

WOESE, C. R. 1987. Bacterial evolution. Microbiol. Rev. **51**: 221–271.

WONG, K. K., and M. MCCLELLAND. 1992. A *Bln*I restriction map of the *Salmonella typhimurium* LT2 genome. J. Bacteriol. **174**:1656–1661.

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.

YANG, Y.-L., D. GOLDRICK, and J.-S. HONG. 1988. Identification of the products and nucleotide sequences of the two regulatory genes involved in the exogenous induction of phosphoglycerate transport in *Salmonella typhimurium*. J. Bacteriol. **170**:4299–4303.