Molecular evolution of the GDP-mannose pathway genes (*manB* and *manC*) in *Salmonella enterica*

Slade O. Jensen and Peter R. Reeves

Author for correspondence: Peter R. Reeves. Tel: +61 2 9351 2536. Fax: +61 2 9351 4571. e-mail: reeves@angis.usyd.edu.au

Department of Microbiology (G08), University of Sydney, New South Wales 2006, Australia

The evolutionary history of the GDP-mannose pathway in *Salmonella enterica* was studied via sequencing *manB* and *manC* genes from 13 representative strains for O antigens containing mannose and/or sugar derivatives of GDP-D-mannose. In addition, colanic acid (CA) *manB* and *manC* genes were sequenced from selected strains, as the basis for a detailed comparison. Interestingly, including the eight previously characterized O antigen gene clusters, 12 of the 21 *S. enterica* strains studied in total (each representing a different O antigen structure) possess a *manB* gene which displays DNA identity, ranging from 93 to 99%, to the CA *manB* gene of *S. enterica* LT2. Furthermore, the CA-like *manB* genes (as well as the CA *manB* and *manC* genes) display subspecies specificity, and the CA and CA-like *manB* genes (for individual strains) appear to be evolving in concert via gene conversion events. In comparison, the *manC* genes were generally not CA-like, a situation also apparent in *Escherichia coli*, and therefore most strongly reflected the evolutionary history of the *S. enterica* O antigen GDP-mannose pathway. It appears that, in relatively recent times, gene capture from a distant source has occurred infrequently, and that groups of *manB* and *manC* genes have been maintained and are continuing to evolve within *S. enterica* and more closely related species.

**Keywords:** concerted evolution, GDP-D-mannose pathway, gene conversion, lateral transfer, *Salmonella enterica*

INTRODUCTION

Bacterial cell surface repeat-unit polysaccharides comprise linear repeats of identical oligosaccharide units, which commonly consist of three to five or more sugars. They include exopolysaccharides (EPSs) and the O antigen regions of lipopolysaccharides (LPSs), which often display extensive intraspecies variation. LPS replaces the outer leaflet of phospholipid in the outer membrane of Gram-negative bacteria, and the immunodominant O antigen region extends from the cell surface forming a polysaccharide layer. The polymorphisms displayed by the O antigen (variations in sugars present and the linkages between them) are the primary basis of serological classification of species in the *Enterobacteriaceae*. The O antigen gene cluster contains genes required for nucleotide sugar biosynthesis (O antigen precursors), and for assembly and processing of oligosaccharide units, and in most cases maps directly upstream of the *gnd* gene in *Escherichia coli* and *Salmonella enterica*.

Genes within a given bacterial species generally have the same GC content, which is thought to be due to directional mutation pressure (Sueoka, 1988, 1992). The sequencing of several *S. enterica* O antigen gene clusters has shown that the GC content of this region is variable and atypical in comparison to the mean GC content of *S. enterica* DNA. We have previously suggested that this indicates that many of the O antigen genes evolved in other species and were later captured by *S. enterica* via lateral gene transfer (Reeves, 1991, 1993).

*S. enterica* has seven well-defined subspecies, which were initially resolved by biotyping and confirmed by DNA hybridization (Le Minor *et al.*, 1982, 1986).
Multilocus enzyme electrophoresis and the sequencing of several housekeeping genes have also shown that strains of a given subspecies generally cluster together, indicating low levels of intersubspecies gene transfer (Boyd et al., 1994; Nelson et al., 1991; Reeves et al., 1989). However, most of the 46 O antigens occur in two or more subspecies, indicating extensive intersubspecies transfer of O antigen genes (Reeves, 1997). Cell surface interactions and niche adaptation are believed to provide the selective pressure for the distribution of O antigen genes amongst subspecies, and the generation of new O antigen gene clusters (Reeves, 1992, 1997).

The nucleotide sugar GDP-d-mannose is required for the mannosylation of many bacterial cell surface repeat-unit polysaccharides and acts as the precursor for other nucleotide sugars (GDP-l-fucose, GDP-colitose, GDP-persamine and GDP-d-rhamnose) involved in polysaccharide biosynthesis. GDP-d-mannose is synthesized from fructose 6-phosphate by products of the manA, manB and manC genes as follows: manA, fructose 6-phosphate to mannose 6-phosphate; manB, mannose 6-phosphate to mannose 1-phosphate; manC, mannose 1-phosphate to GDP-d-mannose. manA encodes type 1 phosphomannomannose isomerase (PMI), and the reversible PMI reaction also enables exogenous mannose to be catabolized via the glycolytic pathway (Neidhardt et al., 1996). The manA gene is generally present in E. coli and S. enterica due to its role in mannose catabolism, and maps as an individual gene not associated with polysaccharide gene clusters (Neidhardt et al., 1987). manB and manC encode phosphomannomutase (PMM) and GDP-mannose pyrophosphorylase (GMP), respectively. These genes are exclusively used for GDP-d-mannose synthesis, and are located within relevant polysaccharide gene clusters. In most examples found so far, both genes are transcribed as part of the same operon, proceeding from manC into manB.

Colanic acid (CA) or M antigen is an EPS produced by E. coli, S. enterica and other enteric bacteria. CA contains l-fucose, and the manB and manC genes, required for the production of GDP-l-fucose via GDP-d-mannose, are located within the CA gene cluster (Aoyama et al., 1994; Stevenson et al., 1991), which maps upstream of the O antigen gene cluster. Strains with an O antigen containing mannose and/or sugars with precursors derived from GDP-d-mannose have a separate set of manB and manC genes in their O antigen gene cluster. For the sake of brevity, the location of specific manB and manC genes will be denoted, when necessary, in subscript following the gene name. For example, the manB genes from the O antigen and CA gene clusters will be referred to as manB_{OAg} and manB_{CA}, respectively. In S. enterica, O antigen genes are generally of lower GC content (ranging between 0.30 and 0.50) than the chromosomal mean (0.52), whereas the manB_{CA} and manC_{CA} genes of S. enterica LT2 have a mean GC content of 0.62. Interestingly, the manB_{OAg} genes of S. enterica C1 (Lee et al., 1992) and E. coli O7 and O157 (Marolda & Valvano, 1993; Wang & Reeves, 1998) display a high level of sequence identity to the manB_{CA} genes of S. enterica LT2 and E. coli K-12, respectively, indicating that in these cases the manB_{OAg} gene was derived from a CA manB gene (Lee et al., 1992; Marolda & Valvano, 1993; Wang & Reeves, 1998).

Homologues of the S. enterica manB and manC genes have been identified in a broad range of species and repeat-unit polysaccharide gene clusters. Therefore, the manB and manC genes provide a basis for following lateral gene transfer, as unlike most other genes in these clusters, the distinct homologies of the GDP-mannose pathway genes enable evolutionary relationships to be studied. This initial study focuses on determining the extent of variation within S. enterica and, for the manB_{OAg} gene, the relationships with the CA isogene.

Twenty-one S. enterica O antigen structures are known to contain mannose and/or sugars with precursors derived from GDP-d-mannose, based on a chemotype study (Luderitz et al., 1966), which included 37 of the 46 currently known structures (Popoff & Minor, 1997). The manB_{OAg} and manC_{OAg} genes from 8 of these 21 structures have previously been characterized, many of them being closely related. In this study, we sequenced these genes from representative S. enterica strains for the remaining 13 O antigen gene clusters. Furthermore, manB_{CA} and manC_{CA} genes were also sequenced for comparison.

**METHODS**

**Bacterial strains, plasmids, bacteriophage and antisera.** In S. enterica, two different systems have been used to name O antigens. The O antigens initially studied in detail were given alphabetical names (A, B, etc.), with individual epitopes, identified by serum absorption studies, given numbers. More recently, specific numbered epitope/s have been used to name O antigens, and in this study we have used the numerical system except for the well recognized alphabetical O antigen names A–E.

Details of bacterial strains, plasmids and bacteriophages used in this study are given in Table 1. Bacteria were routinely grown in nutrient broth (10 g peptone l^{-1}, Amyl Media; 5 g yeast extract l^{-1}, Amyl Media; and 5 g sodium chloride l^{-1} in water) and on nutrient agar (15 g bacteriological agar l^{-1} in nutrient broth), which were supplemented with ampicillin (25 µg ml^{-1}) and chloramphenicol (25 µg ml^{-1}) when required. Antisera against the E. coli O9 antigen was supplied by the Institute of Medical and Veterinary Science (Adelaide, South Australia, Australia).

**Cloning of the manB and manC genes.** Chromosomal DNA was partially digested with SotA3A and subjected to agarose gel electrophoresis. Fragments approximately 3–5 kb in length were purified from the gel and ligated to BamHI-digested pUC19 treated with alkaline phosphatase. E. coli HU1190, containing the plasmid pXX195 (Kido et al., 1995), was transformed with the ligation mix and grown for 3 h at 37 °C. HU1190 does not express O antigen due to a chromosomal deletion of the region encompassing the CA and O antigen gene clusters. The plasmid pXX195 contains the E. coli O9 O antigen gene cluster but does not confer O antigen expression.
Table 1. Bacterial strains, plasmids and bacteriophage

<table>
<thead>
<tr>
<th>Strain*/plasmid/bacteriophage</th>
<th>Relevant properties</th>
<th>Source†/reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strains</strong>&lt;br&gt; <em>S. enterica</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1639</td>
<td>$z_4$, $z_3$: $-$</td>
<td>O11 (IIIa)</td>
</tr>
<tr>
<td>M324</td>
<td>a:1,5</td>
<td>O11 (VI)</td>
</tr>
<tr>
<td>M258</td>
<td>Raus</td>
<td>O13 (I)</td>
</tr>
<tr>
<td>M1711</td>
<td>r: $-$</td>
<td>O13 (V)</td>
</tr>
<tr>
<td>M252</td>
<td>Carrau</td>
<td>O6,14 (I)</td>
</tr>
<tr>
<td>M264</td>
<td>Hvittingfoss</td>
<td>O16 (I)</td>
</tr>
<tr>
<td>M1728</td>
<td>k: $x$</td>
<td>O17 (IIIb)</td>
</tr>
<tr>
<td>M255</td>
<td>Cerro</td>
<td>O18 (I)</td>
</tr>
<tr>
<td>M284</td>
<td>Urbana</td>
<td>O30 (I)</td>
</tr>
<tr>
<td>M274</td>
<td>Adelaide</td>
<td>O35 (I)</td>
</tr>
<tr>
<td>M1784</td>
<td>g, $z_3$: $-$</td>
<td>O38 (IIIa)</td>
</tr>
<tr>
<td>M273</td>
<td>Champaign</td>
<td>O39 (I)</td>
</tr>
<tr>
<td>M1646</td>
<td>Shikmonah</td>
<td>O40 (I)</td>
</tr>
<tr>
<td>M261</td>
<td>a:$z_3$</td>
<td>O40 (II)</td>
</tr>
<tr>
<td>M1651</td>
<td>$z_4$, $z_3$: $-$</td>
<td>O40 (IIIa)</td>
</tr>
<tr>
<td>M1653</td>
<td>i:1,5,7</td>
<td>O40 (IIIb)</td>
</tr>
<tr>
<td>M1654</td>
<td>g, t: $-$</td>
<td>O40 (IV)</td>
</tr>
<tr>
<td>M1655</td>
<td>g, $z_3$: $-$</td>
<td>O40 (IV)</td>
</tr>
<tr>
<td>M1656</td>
<td>$z_3$: $-$</td>
<td>O40 (V)</td>
</tr>
<tr>
<td>M1657</td>
<td>$z_3$: $-$</td>
<td>O40 (V)</td>
</tr>
<tr>
<td>M286</td>
<td>Vietnam</td>
<td>O41 (I)</td>
</tr>
<tr>
<td>M277</td>
<td>$z_{121}$:1,5,7</td>
<td>O43 (II)</td>
</tr>
<tr>
<td>M326</td>
<td>a: $x$,n,x</td>
<td>O45 (VI)</td>
</tr>
<tr>
<td>M1633</td>
<td>i: $-$</td>
<td>O48 (V)</td>
</tr>
<tr>
<td>M290</td>
<td>$z$: $x$,n,x</td>
<td>O50 (II)</td>
</tr>
</tbody>
</table>

**Genotype or O group**

| Strain*M1180 | O39 | I |
| M1182 | O41 | I |
| M1194 | O125 | I |
| M1204 | O6 | I |
| HU1190 | Δ(sbcB–rfb) bsdR4 recA56 srl::Tn10 | Sugiyama *et al.* (1994) |

**Plasmids**

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Relevant property</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pXX195</td>
<td>Derivative of pACYC184</td>
<td>Sugiyama <em>et al.</em> (1994)</td>
</tr>
<tr>
<td>pBHC4</td>
<td>Derivative of pBluescript II</td>
<td>Sugiyama <em>et al.</em> (1994)</td>
</tr>
</tbody>
</table>

**Bacteriophage**

<table>
<thead>
<tr>
<th>Bacteriophage</th>
<th>Relevant property</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ffm</td>
<td>Lyses <em>E. coli</em> strains with rough LPS</td>
<td>Jayaratne <em>et al.</em> (1994)</td>
</tr>
</tbody>
</table>

* M numbers denote laboratory strain names.
† I, Institute of Medical and Veterinary Science, Adelaide, Australia (C. Murray); P, Pasteur Institute, Paris, France (L. Le Minor).
‡ Numerals in parentheses denote subspecies.

due to a deletion of the *manB*<sub>ony</sub> and *manC<sub>G</sub>ory* genes. The gene bank in *E. coli* HU1190 was spread onto plates seeded with approximately 10<sup>5</sup> p.f.u. Ffm. Transforms containing both the *manB* and *manC* genes expressed the *E. coli* O9 O antigen and were positively selected for via resistance to the bacteriophage Ffm (Jayaratne *et al.*, 1994). Ffm attaches to the core of the LPS molecule and expression of the O antigen confers resistance by masking the receptor site. Sequencing of the *manB* and *manC* genes and adjacent areas revealed that many of the clones were derived from the CA
RESULTS AND DISCUSSION

We sequenced the man\textsubscript{OAg}B and man\textsubscript{OAg}C genes from representative strains for \textit{S. enterica} O antigens O11, O13, O6/14, O16, O18, O30, O35, O39, O40, O41, O43, O45 and O50, which contain mannose and/or sugars with precursors derived from GDP-D-mannose. The sequences were aligned and the results are presented diagrammatically in Fig. 1, together with the \textit{S. enterica} man\textsubscript{OAg}B and man\textsubscript{OAg}C genes previously characterized. Interestingly, most of the man\textsubscript{OAg}B genes (11 of the 13 new sequences) displayed a high level of DNA identity (ranging from 93 to 99\%) to the man\textsubscript{C}\textsubscript{CA} gene of \textit{S. enterica} LT2 (Stevenson et al., 1991). These genes are referred to as CA-like man\textsubscript{OAg}B genes.

It is now evident that CA-like man\textsubscript{OAg}B genes are common in \textit{Salmonella}, although previously only seen for group C1 (Lee et al., 1992). However, six of the previously characterized O antigen gene clusters (A, B, D1, D2, D3 and E1; Brown et al., 1992; Curd et al., 1998; Jiang et al., 1991; Liu et al., 1991; Wang et al., 1992; Xiang et al., 1994) are closely related and essentially contain the same \textit{man}\textsubscript{OAg}B and man\textsubscript{OAg}C genes. Note that the man\textsubscript{OAg}B and man\textsubscript{OAg}C genes of O antigens A, D1, D2, D3 and E1 were not fully sequenced but shown, by detailed restriction mapping and limited sequencing, to be very similar to the LT2 (B) O antigen genes. With the new data included, 12 of the 21 strains studied (each representing a different O antigen structure) possess a man\textsubscript{OAg}B gene which, in whole or part, displays a high level of identity to the LT2 man\textsubscript{C}\textsubscript{CA} gene, with six of the nine exceptions being very closely related.

In general, the man\textsubscript{OAg}C genes do not display a high level of identity to the LT2 man\textsubscript{C}\textsubscript{CA} gene (Stevenson et al., 1991). However, the man\textsubscript{OAg}C genes of M255 (O18), M286 (O41) and M324 (O11) (Fig. 1) are chimeric in structure, resembling in part both a man\textsubscript{C}\textsubscript{CA} gene (3’ end) and a man\textsubscript{OAg}C gene (low GC content) (5’ end). We
also sequenced the manB<sub>CA</sub> and manC<sub>CA</sub> genes from selected S. enterica strains as the basis for a more detailed comparison.

Phylogenetic analysis of the GDP-mannose pathway

To examine the phylogenetic relationships of the derived ManB and ManC sequences, evolutionary trees were constructed using the neighbour-joining method (Saitou & Nei, 1987). The trees, shown in Fig. 2(a, b), also include ManB and ManC sequences from several other species, and for both trees S. enterica and E. coli sequences predominantly cluster together. In considering this, we also constructed a tree based on the N-terminal end of the ManC sequences, which included the O antigen (not CA-like) segments of the chimeric manC<sub>OA</sub> genes from M255 (O18), M286 (O41), and M324 (O11) (Fig. 2c). It can be seen in Fig. 2(c) that the O18, O41 and O11' gene segments also cluster with S. enterica and E. coli sequences. Furthermore, in addition to the CA and CA-like sequences (discussed in detail later), there are a number of distinct relationships among the S. enterica ManB and ManC sequences which we will discuss.

GDP-colitose, GDP-l-fucose and GDP-perosamine are synthesized via extensions of the GDP-d-mannose pathway, and there are eight S. enterica O antigens known to contain one or more of these sugars. Six of these O antigens (O13, O30, O35, O43, O45 and O50) lack d-mannose, and with the exception of O35, they all have similar manC<sub>OA</sub> genes (DNA identity ranging from 90 to 98%). These six manC<sub>OA</sub> genes are also preceded by the gmm gene, which encodes GDP-mannose mannosyl hydrolase (Frick et al., 1995).

S. enterica O35 has the same O antigen structure as E. coli O111 and it is interesting that their manB<sub>OA</sub> and manC<sub>OA</sub> genes are closely related (Fig. 2a, b, respect-
ively). The level of divergence between their \( \text{man}B_{\text{OAg}} \) and \( \text{man}C_{\text{OAg}} \) genes (data not shown) is consistent with these genes, and presumably the whole O antigen gene cluster, having been present in the common ancestor and maintained in both extant species. Similarly, \( S. \text{enterica} \) O50 and O30 have the same O antigen structures as \( E. \text{coli} \) O55 and O157, respectively. However, in these examples, the \( \text{man}C_{\text{OAg}} \) genes are more divergent and not as closely related (Fig. 2b). The \( S. \text{enterica} \) O50 and O30 and \( E. \text{coli} \) O157 \( \text{man}B_{\text{OAg}} \) genes are \( \text{CA-like} \), and are therefore not relevant in discussing common ancestry.

The \( S. \text{enterica} \) O16 and O39 O antigen structures contain both \( \text{D-mannose} \) and \( \text{l-fucose} \), and their \( \text{man}C_{\text{OAg}} \) and \( \text{CA-like} \) \( \text{man}B_{\text{OAg}} \) genes are unique in \( S. \text{enterica} \), as they are separated by an additional gene. Their \( \text{man}C_{\text{OAg}} \) genes are almost identical (99.9% identity), and are related to those from \( E. \text{coli} \) O41 and O125 (this study). \( Yersinia \text{ pseudotuberculosis} \) 1b (Skurnik et al., 2000) and \( Yersinia \text{ enterocolitica} \) O8 (Zhang et al., 1997) (Fig. 2b). Interestingly, in \( E. \text{coli} \) O41 and O125 and \( Y. \text{ pseudotuberculosis} \) 1b, the \( \text{man}C_{\text{OAg}} \) and \( \text{man}B_{\text{OAg}} \) genes are also separated by an additional gene. In each case, the additional gene encodes a putative glycosyl transferase, which like the \( \text{man}C_{\text{OAg}} \) gene, is related amongst these species (data not shown).

Many of the well-documented O antigens (A, B, D and E) are structurally similar, based on O-unit linkages and sugar composition, and have related gene clusters (discussed above). The individual O antigen groups with related \( \text{man}C-\text{man}B \) gene assemblages, which in some cases include adjacent genes, also have structural elements in common. The relationships within these groups are only based on parts of the O antigen gene cluster; however, they indicate that there are conserved groups of O antigen genes, related to structural elements, which can cross species boundaries.

**CA-like \( \text{man}B_{\text{OAg}} \) genes**

There is a very distinct group of \( \text{man}B_{\text{OAg}} \) and \( \text{man}C_{\text{OAg}} \) genes with \( \text{CA-like} \) characteristics, based on sequence similarity and GC content. In addition to the \( \text{man}B_{\text{CA}} \) and \( \text{man}C_{\text{CA}} \) genes, the \( \text{CA-like} \) O antigen genes also have a GC content ranging from 0.60 to 0.62. In comparison, the GC content of other \( \text{man}B_{\text{OAg}} \) and \( \text{man}C_{\text{OAg}} \) genes (or gene segments) ranges from 0.34 to 0.43. As in the case of the \( S. \text{enterica} \) C1 CA-like \( \text{man}B_{\text{OAg}} \) gene, the newly sequenced CA-like \( \text{man}B_{\text{OAg}} \) genes display a high level of identity to the LT2 \( \text{man}B_{\text{CA}} \) gene. In addition, the CA-like \( \text{man}B_{\text{OAg}} \) genes of strains M324 (O11), M286 (O41) and M255 (O18) display CA similarity in the upstream intergenic region and/or \( \text{man}C \) gene (discussed in detail later). However, the 28–60 bp segment at the 3′ end exhibits a high level of divergence between strains, a low GC content, and variation in the position of stop codons over a range of 15 bp. In the case of strain M255, the divergence extends 360 bp upstream of the 3′ end.

**CA-like \( \text{man}B_{\text{OAg}} \) genes are subspecies-specific**

The phylogenetic relationships of the \( \text{man}B_{\text{CA}} \) genes and the CA-like \( \text{man}B_{\text{OAg}} \) genes are shown in a tree constructed using the neighbour-joining method (Saitou & Nei, 1987) (Fig. 3a). Only CA and CA-like sequences were used for this analysis and it became clear that these genes display subspecies specificity. The strains initially studied were not selected with subspecies in mind and are predominantly subspecies I. Therefore, to provide further data on subspecies specificity of the CA-like \( \text{man}B_{\text{OAg}} \) genes, we sequenced CA and/or O antigen \( \text{man}B \) genes from additional strains. In particular, we included both the \( \text{man}B_{\text{CA}} \) and CA-like \( \text{man}B_{\text{OAg}} \) isogenes of some strains, and representative O40 strains from subspecies I, II, IIIa, IIIb, IV and V. It can be seen in Fig. 3(a) that not only do CA and CA-like \( \text{man}B \) genes display subspecies specificity but that the topology of this tree is generally in accordance with a tree constructed from the combined sequences of five housekeeping genes (Neidhardt et al., 1996) (Fig. 3b). The \( \text{man}C_{\text{CA}} \) genes also clustered according to subspecies (data not shown).

The subspecies V \( \text{man}B_{\text{CA}} \) gene is atypical, as it appears to be partially derived from subspecies I. We constructed trees for different regions of the \( \text{man}B_{\text{CA}} \) gene and found that for nucleotide positions 1–825 subspecies V clusters apart from the other subspecies in accordance with the consensus tree for housekeeping genes. However, for nucleotide positions 826–1307, subspecies V is most closely related to subspecies I. This indicates that since the divergence of subspecies I from the other subspecies, the 3′ end of a subspecies I \( \text{man}B_{\text{CA}} \) gene replaced the 3′ end of a subspecies V \( \text{man}B_{\text{CA}} \) gene and was maintained during the continuing divergence of subspecies V.

The only other exceptions to the subspecies specificity of the CA-like \( \text{man}B_{\text{OAg}} \) genes were the sequences from strains M1653 (O40, IIIb) and M326 (O45, VI), which clustered with subspecies I sequences. This presumably reflects intersubspecies transfer of O antigen genes, as also inferred from similar observations on the \( \text{gnd} \) gene (Reeves, 1997; Thampapillai et al., 1994). Thampapillai et al. (1994) observed that the \( \text{gnd} \) genes of several \( S. \text{enterica} \) strains have a chimeric structure, and proposed that the 5′ end of these genes had transferred between subspecies in association with O antigen genes, presumably driven by natural selection for antigenic variation.

**Origins of the CA-like O antigen sequence**

The \( \text{man}B_{\text{OAg}} \) and \( \text{man}C_{\text{OAg}} \) genes identified in this study are distributed at various positions within the O antigen gene cluster (data not shown), and no definitive pattern is evident. Most of these genes are internally located, and in reference to lateral transfer between subspecies, are presumably transferred with the O antigen gene cluster. In considering this, and the subspecies specificity of the CA-like \( \text{man}B_{\text{OAg}} \) genes, we conclude that in general the CA-like \( \text{man}B_{\text{OAg}} \) and \( \text{man}C_{\text{OAg}} \) genes were acquired after transfer to the
Evolution of the manB and manC genes in S. enterica

Fig. 3. (a) Neighbour-joining tree of S. enterica CA and CA-like manB genes. Bootstrap values are percentages of 1000 computer-generated trees and are shown at the nodes (calculated without subspecies V strains). Values of less than 50 are not shown. Individual genes are designated by laboratory strain name and polysaccharide type, and numerals denote subspecies. Dashed lines do not reflect evolutionary distance and represent the position of subspecies V strains for the regions shown in brackets. (b) Neighbour-joining tree based on the combined sequences of five S. enterica housekeeping genes, putP, gapA, mdh, gnd and aceK (adapted from Neidhardt et al., 1996).

relevant subspecies. Therefore, it appears that the CA and O antigen manB genes, of individual strains, are evolving in concert via gene conversion events. These gene conversion events appear to be unidirectional, as we have not seen a manB<sub>OAg</sub> gene (low GC content), in a CA gene cluster.

The CA-like manB<sub>OAg</sub> genes from strains M277 (O43), M290 (O50), M324 (O11) and M1654 (O40) are not only subspecies-specific but most closely related to their respective CA isogenes (Fig. 3a). This indicates that relatively recent gene conversion events have occurred within these strains, supporting the above explanation for the subspecies specificity of CA-like manB<sub>OAg</sub> genes. The manB isogenes from strains M264 (O16), M273 (O39), M286 (O41), M1651 (O40) and M1656 (O40) are not as closely related. However, both genes of each pair still cluster within the same subspecies group, and presumably the CA-like manB<sub>OAg</sub> genes originated from earlier gene conversion events, after transfer of the O antigen gene cluster to the relevant subspecies.

Distribution of polymorphic nucleotide sites within the CA and CA-like manB<sub>OAg</sub> genes

Among the S. enterica CA and CA-like manB genes (22 and 17 sequences, respectively), there were 362 polymorphic nucleotide sites within the 1307 bp segment studied (codons 1–444, excluding the divergent 3’ end) (Fig. 4). The mean pairwise difference was 7-65% at the DNA level, with variation predominantly attributable to differences between subspecies.

The Stephens test (Stephens, 1985) was used to detect non-random clustering of polymorphic nucleotide sites, which may be indicative of intragenic recombination. For the manB<sub>CA</sub> and CA-like manB<sub>OAg</sub> genes of the S. enterica strains studied, the Stephens test identified two partitions with significantly non-random distribution of sites. Partitions based on subspecies specificity and/or supported by five or fewer sites are not discussed.

The first partition, from nucleotide positions 123 to 825 (5’ end of the gene), separates the subspecies V strains from other subspecies, and reflects the fact that the 3’ end of the subspecies V genes (both CA and CA-like) is similar to those of subspecies I, as discussed above. The remaining partition, from nucleotide positions 1076 to 1307, separates the M255 (O18) CA-like manB<sub>OAg</sub> sequence from all others. This region represents the divergent (low GC content) 3’ end of the M255 CA-like manB<sub>OAg</sub> gene (Fig. 4).

Visual inspection of the polymorphic nucleotide sites (Fig. 4) provides evidence for recent intragenic exchange between the manB<sub>CA</sub> and CA-like manB<sub>OAg</sub> isogenes from strain M326. M326 is a subspecies VI strain with a subspecies I CA-like manB<sub>OAg</sub> gene (discussed above). However, for nucleotide positions 403 to 798 the CA and CA-like manB isogenes are identical. This region most closely resembles subspecies VI manB<sub>CA</sub> sequence,
S. O. JENSEN and P. R. REEVES

Fig. 4. Polymorphic nucleotide sites among *S. enterica* CA and CA-like *manB* genes (nucleotide positions 1–1307). Individual sequences are specified by O antigen group, polysaccharide type and laboratory strain number. Roman numerals denote subspecies and asterisks indicate synonymous substitutions. The low GC content 3’ end of the CA-like *manB* genes is not shown.
Chimeric CA-like manC\textsubscript{OAg} genes

Strains M255 (O18), M286 (O41) and M324 (O11) possess a CA-like manB\textsubscript{OAg} gene and a chimeric manC\textsubscript{OAg} gene with a CA-like 3' end (Fig. 1). These genes (including the intergenic region) were aligned with their respective CA isogenes. Visual inspection of the polymorphic nucleotide sites (Fig. 5) reveals that, for M286 and M324, the CA-like O antigen manC–manB gene assemblage (nucleotide positions 376–2888 and 466–2888, respectively) is very similar to that of their own CA gene clusters. It is likely that a relatively recent gene conversion event involved part of the manC\textsubscript{CA} gene, in addition to the downstream intergenic region and manB\textsubscript{CA} gene. In contrast, the O antigen gene cluster of M255 appears to have acquired the CA-like part of its manC\textsubscript{OAg} gene (positions 808–1391) in a separate event to the acquisition of the CA-like manB\textsubscript{OAg} gene.

Comparison of E. coli and S. enterica

Recently, Lai \textit{et al.} (1998), in an attempt to determine the ancestral O antigen of \textit{E. coli} Sonnei, found 14 \textit{E. coli} strains (each representing a different O antigen group) to have, at least in part, a CA-like manB\textsubscript{OAg} gene. Therefore, including the \textit{E. coli} O antigens O7, O8, O9, O9\textsubscript{a}, O11 and O157, 16 out of the 20 \textit{E. coli} strains previously studied, which contain mannose and/or sugar derivatives of GDP-mannose, possess a manB\textsubscript{OAg} gene with a high level of identity to the manB\textsubscript{CA} gene of \textit{E. coli} K-12. In this study, we completely sequenced the manB\textsubscript{OAg} and manC\textsubscript{OAg} genes from four of the strains partly sequenced by Lai \textit{et al.} (1998) (see Table 1), representing O antigen groups O6, O39, O41 and O125. These strains (like in \textit{E. coli} O7 and O157) were found to possess a CA-like manB\textsubscript{OAg} gene and a low GC manC\textsubscript{OAg} gene (shown in Fig. 2b). Therefore, in parallel to \textit{S. enterica}
enterica, CA-like manB_{OAg} genes are prevalent in E. coli, and are associated with low GC manC_{OAg} genes. In addition, we also sequenced part of the O6, O39, O41 and O125 manB_{CA} genes. In contrast to the situation in S. enterica, we found that the manB isogenes of individual strains were not particularly similar (data not shown), indicating that the CA-like manB_{OAg} genes have been affected by recombination since any gene conversion events. This is not too surprising, as E. coli does not display a defined subspecies structure, which is probably due to a higher rate of recombination than in S. enterica.

Concluding comments
CA-like manB_{OAg} genes have previously been observed in both E. coli and S. enterica; however in this study, a distinct pattern has emerged. The prevalence of subspecies-specific CA-like manB_{OAg} genes in S. enterica indicates that the manB genes of individual strains are evolving in concert, with a strong preference for the acquisition and maintenance of an entire manB_{CA} gene within the O antigen gene cluster. The efficiency of individual O antigen genes may not be critical with respect to O antigen expression when these genes are first transferred to S. enterica. In contrast, if the O antigen occupies a particular niche, selection pressure may eventually drive replacement of specific genes with better adapted homologues (if available). In particular, if the manB_{CA} genes were better adapted to expression in S. enterica, this could explain the unidirectional gene conversion events (manB_{CA} genes replacing manB_{OAg} genes).

However, the replacement of existing CA-like manB_{OAg} genes with subspecies-specific sequence, derived from the manB_{CA} isogene, is unlikely to confer any selective advantage. This suggests that the direction of gene conversion is not random, and a possible explanation could relate to the chromosomal location of the individual manB isogenes. Abdulkarim & Hughes (1996) observed that for the tufA and tufB genes in S. enterica (which are evolving in concert), the rate of sequence transfer was different depending on which tuf gene was the donor. It was suggested that the distribution and frequency of chromosomal breakpoints and chi sites, in relation to the location of the individual tuf genes, contributed to the biased rates of sequence transfer in one direction (Abdulkarim & Hughes, 1996). At present we have no specific explanation for the unidirectional gene conversion of manB_{OAg} genes; however, there is a precedent in relation to the tufA and tufB genes.

We embarked on this study with the expectancy that the evolutionary relationships of manB and manC genes would provide information on the origin of O antigen gene clusters. Due to the unexpected finding that homogenization of S. enterica manB_{OAg} genes is widespread, limited information is provided in this regard. However, for both the ManB and ManC trees (Fig. 2a, b, respectively) there is a general clustering of sequences from E. coli, S. enterica and other related species (Klebsiella pneumoniae, Y. enterocolitica and Y. pseudotuberculosis), which encompasses a number of distinct relationships discussed above. This suggests that in relatively recent times, gene capture from a distant source (as suggested by atypical GC content) has occurred infrequently and that the manB and manC genes, and their associated gene assemblages, are maintained and continue to evolve within these species. It is not surprising that genetic exchange would occur more frequently between closely related species; however, a broad-range study of the GDP-mannose pathway genes (encompassing many bacterial species) is required.

ACKNOWLEDGEMENTS
This work was supported by a grant from the Australian Research Council. We especially thank Dr N. Kido (Nagoya University, Nagoya, Japan) for supplying the E. coli strain HU1190 and the plasmid pXX195. We also thank Dr Ruiting Lan and Qun Li for helpful discussions during preparation of this manuscript.

REFERENCES
Evolution of the manB and manC genes in *S. enterica*


Received 9 June 2000; revised 25 September 2000; accepted 23 October 2000.