ELSEVIER

Gene 270 (2001) 231–236

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Sequence of the *E. coli* O104 antigen gene cluster and identification of O104 specific genes

Lei Wang[a,1], Connie E. Briggs[b], Deborah Rothemund[a], Pina Fratamico[b], John B. Luchansky[b], Peter R. Reeves[a,*]

[a]*Department of Microbiology (GO8), The University of Sydney, Sydney 2006, NSW, Australia*
[b]*Microbial Food Safety Research Unit, Eastern Regional Research Center, Agricultural Research Service, U.S. Department of Agriculture, 600 East Mermaid Lane, Wyndmoor, PA 19038-8595, USA*

Received 19 December 2000; received in revised form 21 March 2001; accepted 30 March 2001
Received by A.M. Campbell

Abstract

The *Escherichia coli* O104 polysaccharide is an important antigen, which contains sialic acid and is often associated with EHEC clones. Sialic acid is a component of many animal tissues, and its presence in bacterial polysaccharides may contribute to bacterial pathogenicity. We sequenced the genes responsible for O104 antigen synthesis and have found genes which from their sequences are identified as an O antigen polymerase gene, an O antigen flippase gene, three CMP-sialic acid synthesis genes, and three potential glycosyl transferase genes. The *E. coli* K9 group IB capsular antigen has the same structure as the O104 O antigen, and we find using gene by gene PCR that the K9 gene cluster is essentially the same as that for O104. It appears that the distinction between presence as group IB capsule or O antigen for this structure does not involve any difference in genes present in the O antigen gene cluster. By PCR testing against representative strains for the 166 *E. coli* O antigens and some randomly selected Gram-negative bacteria, we identified three O antigen genes which are highly specific to O104/K9. This work provides the basis for a sensitive test for rapid detection of O104 *E. coli*. This is important both for decisions on patient care as early treatment may reduce the risk of life-threatening complications and for a faster response in control of food borne outbreaks. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Polymerase chain reaction (PCR); Pathogenic *E. coli*; Molecular typing; Serotyping

1. Introduction

Escherichia coli is a clonal species, with clones normally identified by their combination of O and H (and sometimes K) antigens. *E. coli* O104:H2 and O104:NM are important pathogens, causing bloody diarrhoea and haemolytic uraemic syndrome in humans. Since most laboratories do not screen stool samples for *E. coli* O104, the magnitude of the public health problem posed by these clones is probably underestimated.

The O antigen, which contains many repeats of an oligosaccharide unit (O unit), is part of the lipopolysaccharide present in the outer membrane of Gram-negtive bacteria. The O unit is synthesised by sequential transfer of a sugar phosphate and sugars from respective nucleotide sugars to the carrier lipid, undecaprenyl phosphate (UndP). O-units are then polymerized on UndPP into polysaccharide chains, which are transferred to the independently synthesised core-lipid A to form LPS (Reeves, 1994). Characteristically, all genes specific to O antigen synthesis in *E. coli* are clustered between the *galF* and *gnd* genes (Reeves et al., 1996).

The O antigen contributes major antigenic variability to the cell surface, and on the basis of this variation 166 O antigen forms have been recognised in *E. coli*. The surface O antigen is subject to intense selection by the host immune system, which may account for the maintenance of many different O antigen forms within species such as *E. coli*.

Sialic acid is an important component of glycoconjugates in animal tissues. The presence of this sugar in bacterial polysaccharides may contribute to pathogenicity by mimicking the host tissue component. *E. coli* O104 is one of the few *E. coli* O antigens containing sialic acid (Gamian et al., 1992). It is worth noting that O104 has the same

Abbreviations: CMP, cytidine monophosphate; EHEC, enterohemorrhagic *E. coli*; Gal, galactose; GalNAc, *N*-acetyl galactosamine; GlcNAc, *N*-acetyl glucoseamine; NeuNAc, sialic acid; PCR, polymerase chain reaction; UndP, Undecaprenol phosphate; μl, microlitre

* Corresponding author. Tel.: +612-9351-2536; fax: +612-9351-4571.
E-mail address: reeves@angis.org.au (P.R. Reeves).

[1] Collaborator, via a fellowship under the OECD Co-operative Research Program: Biological Resource Management for Sustainable Agriculture Systems.

0378-1119/01/$ - see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S0378-1119(01)00471-1

structure as the *E. coli* K9 antigen (Dutton and Parolis, 1989), a group IB K antigen which is usually associated with the O8 or O9 antigen (Whitfield and Roberts, 1999).

To better understand the genetics of the O104 antigen, we have sequenced the DNA between *galF* and *gnd* of the O104 type strain. Analysis of the sequence revealed three CMP-sialic acid biosynthetic pathway genes plus a putative acetyltransferase gene, three presumptive sugar transferase genes for synthesis of the O unit, the O antigen flippase gene and the O antigen polymerase gene. We showed that the *E. coli* K9 capsular antigen gene cluster is also located between *galF* and *gnd*, and has the same genes and same gene order as that of O104. By PCR testing against strains carrying the 166 known *E. coli* O antigens and a range of Gram-negative bacterial strains, we identified several genes specific for O104.

2. Materials and methods

2.1. Bacterial strains

E. coli O104:H12 (the type strain, isolate H519) was from the Institute of Medical and Veterinary Science, Adelaide, Australia. Other *E. coli* O antigen type strains (Lior, 1994) are listed in Table 1. Other strains used are also listed in Table 1 together with the names of those who supplied them. Plasmids were maintained in *E. coli* K-12 strain JM109.

2.2. Construction of random DNaseI bank for sequencing DNA fragments

Chromosomal DNA used as template for PCR was prepared using the Wizard DNA preparation kits from Promega. Long-PCR was carried out using the Expand Long Template PCR System from Roche, and products were subjected to DNaseI digestion and cloned into pGEM-T to make banks for sequencing using the method described previously by Wang and Reeves (1998).

2.3. Sequencing and analysis

DNA template for sequencing was prepared using the BioRobot 9600 of Qiagen, Inc. (Valencia, CA). Sequencing was performed with an Applied Biosystems 3700 automated

Table 1
Bacterial strains and PCR pools

Pool No.	Strains of which chromosomal DNA included in the pool	Source
1	*E. coli* type strains for O serotypes 1, 2, 3, 4, 8, 10, 16, 18, and 39	IMVS[a]
2	*E. coli* type strains for O serotypes 40, 41, 43, 49, 71, 73, 88, and 100	IMVS
3	*E. coli* type strains for O serotypes 102, 109, 119, 120, 121, 125, 126, and 137	IMVS
4	*E. coli* type strains for O serotypes 138, 139, 149, 7, 5, 6, 11, and 12	IMVS
5	*E. coli* type strains for O serotypes 13, 14, 15, 17, 19, 20, 21, and 22	IMVS
6	*E. coli* type strains for O serotypes 23, 24, 25, 26, 27, 28, 29, 30, and 32	IMVS
7	*E. coli* type strains for O serotypes 33, 34, 35, 36, 37, 38, 42, and 43	IMVS
8	*E. coli* type strains for O serotypes 44, 45, 46, 50, 51, 52, 53, and 54	IMVS
9	*E. coli* type strains for O serotypes 55, 56, 57, 58, 59, 60, 61, and 62	IMVS
10	*E. coli* type strains for O serotypes 63, 64, 65, 66, 68, 69, 70, and 74	IMVS
11	*E. coli* type strains for O serotypes 75, 76, 77, 78, 79, 80, 81, and 82	IMVS
12	*E. coli* type strains for O serotypes 83, 84, 85, 86, 87, 89, 90, and 91	IMVS
13	*E. coli* type strains for O serotypes 92, 95, 96, 97, 98, 99, 101, and 103	IMVS
14	*E. coli* type strains for O serotypes 104, 105, 106, 107, 108, 110, 111, and 112[a]	IMVS
15	*E. coli* type strains for O serotypes 112[b], 113, 114, 115, 116, 117, 118, and 123	See [b]
16	*E. coli* type strains for O serotypes 165, 166, 167, 168, 169, 170, 171, and 172	See [c]
17	*E. coli* type strains for O serotypes 173, 127, 128, 129, 130, 131, 132, and 133	IMVS
18	*E. coli* type strains for O serotypes 134, 135, 136, 140, 141, 142, 143, and 144	IMVS
19	*E. coli* type strains for O serotypes 145, 146, 147, 148, 150, 151, 152, and 153	IMVS
20	*E. coli* type strains for O serotypes 154, 155, 156, 157, 158, 159, 160, and 161	IMVS
21	*E. coli* type strains for O serotypes 163, 164, 9, and 124	IMVS
22	As pool #14, without *E. coli* 0104:H12 strain	
23	As pool #21, without *E. coli* O9 strain	
24	As pool #1, without *E. coli* O8 strain	
25	*Shigella* strains: *S. sonnei* 3354[d], *S. boydii* type17[e], *S. dysenteriae* type 2[e], 12[e] and 13[e]	See [d,e]
26	*Vibrio cholerae* strains with O-groups O1, O2, O5, O8, O11, O28 and O31	See [f]
27	*Streptococcus pneumoniae* strains type 4, 6A, 6B, 7, 12F, 18C and 19F	See [g]

[a] Institute of Medical and Veterinary Science, Adelaide, Australia.
[b] 123 from IMVS; the rest from Statens Serum Institut, Copenhagen, Denmark.
[c] 172 and 173 from Statens Serum Institut, Copenhagen, Denmark; the rest from IMVS.
[d] Swiss Serum and Vaccine Institute, Berne, Switzerland.
[e] Canadian National Laboratory for Enertic Pathogens (NLEP).
[f] Toshio Shimada, National Institute of Health, Tokyo, Japan.
[g] James Paton, Molecular Microbiology Unit Women's and Children's Hospital, North Adelaide, Australia.

DNA sequencer. Sequence data were assembled using the Phred/Phrap package of the University of Washington Genome Center, and the sequence annotation was done using the program Artemis from the Sanger Centre. We used the algorithm described by Eisenberg et al. (1984) to identify potential transmembrane segments from the amino acid sequence.

2.4. Specificity assay by PCR

Chromosomal DNA was isolated using the Promega Genomic isolation kit and checked by gel electrophoresis. The quality of chromosomal DNA from each *E. coli* strain was checked by PCR amplification of the *mdh* gene (coding for malate dehydrogenase and generally present in *E. coli*) using oligonucleotides as described in Wang and Reeves (Wang and Reeves, 1998). 28 pools were made, with 5–9 samples of DNA per pool (Table 1). PCR for O antigen genes was carried out in a total volume of 25 μl, and after PCR 10 μl was run on an agarose gel to check for amplified DNA.

2.5. Nucleotide sequence accession number

The *E. coli* O104 antigen gene cluster sequence has been deposited in GenBank under accession number AF361371.

3. Results and discussion

3.1. Sequencing

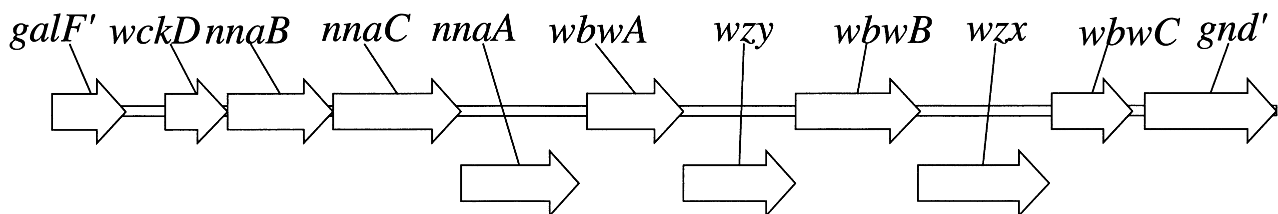
A total of 12105 bp, which covers DNA from *galF* to *gnd*, was sequenced by using a random DNaseI bank constructed from DNA amplified by L-PCR using primers #1523 (5′-ATTGTGGCTGCAGGGATCAAAGAAATC) and #1524 (5′-TAGTCRCGCTGNGCCTGRATYARGTTMGC (Code for degenerated bases: M = AC, N = ACTG, R = AG, Y = CT)) which bind to the 5′ end of *galF* and 3′ end of *gnd*, respectively. To limit the effect of PCR errors, ten individual PCR reaction products were pooled before making the bank.

Nine open reading frames between *galF* and *gnd* were predicted from the sequence (Fig. 1): all have the same transcriptional direction from *galF* to *gnd*. The nucleotide and amino acid sequences were used to search available databases for indication of possible function.

3.2. O104 O antigen genes

The structure of the O104 O unit is known (Fig. 1) and we expected to find transferase genes for two galactose residues, GalNAc and sialic acid (NeuNAc), genes for the synthesis of CMP-NeuNAc, an O antigen flippase gene (*wzx*), and an O antigen polymerase gene (*wzy*).

CMP-NeuNAc pathway genes: It is known that CMP-NeuNAc is synthesized from GlcNAc by a three step pathway in both *Neisseria meningitidis* and *E. coli* strains with NeuNAc-containing group II capsules (Annunziato et al., 1995; Petersen et al., 2000). Enzymes GlcNAc-2-epimerase, NeuNAc condensing enzyme and CMP-NeuNAc synthetase catalyse the three steps, and they are encoded by genes that have been named *neuC*, *neuB* and *neuA* in *E. coli* and *siaA*, *siaC* and *siaB* in *N. meningitidis*. We have found all three genes in the O104 gene cluster, identified because they encode proteins with high levels of identity (58, 68 and 49%, respectively) to those encoded by genes in the *E. coli* K1 gene cluster (GenBank entries: U05248, U05023, and M84026). The DNA identity levels between these three pairs of genes range from 61.7 to 68.6%, and indicating that these genes have evolved separately for a long time. As two different sets of names have been used for the CMP-NeuNAc biosynthetic pathway genes and as neither names the genes in pathway order, we have named these three genes *nnaA* (*neuC* or *siaA*), *nnaB* (*neuB* or *siaC*) and *nnaC* (*neuA* or *siaB*), respectively. Naming nucleotide sugar pathway genes in pathway order has been one of the



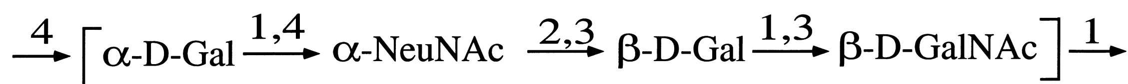


Fig. 1. *E. coli* O104 antigen structure (Gamian et al., 1992) and its gene cluster.

widely accepted benefits of the BPGD system (Reeves et al., 1996), and we prefer to use new names rather than choosing between existing names. We hope that these names can be used for many CMP-NeuNAc genes yet to be sequenced.

We have named the first open reading frame *wckD*. WckD shares 64% identity with NeuD of the *E. coli* K1 capsular gene cluster (Annunziato et al., 1995). *neuD* is present in addition to the three CMP-NeuNAc biosynthetic pathway genes in the *E. coli* K1 capsule gene cluster (Annunziato et al., 1995) and *Strepcoccus agalactiae* type III capsule gene cluster (GenBank entry. AF163833). NeuBAC are sufficient for CMP-NeuNAc synthesis when cloned into *E. coli* K-12 on a multicopy plasmid (Annunziato et al., 1995). However it has been shown recently that NeuD of the *E. coli* K1 capsule gene cluster has a role in the synthesis of NeuNAc by interacting with NeuB (Daines and Silver, 2000; Daines et al., 2000). WckD probably has the same or similar role in O104 synthesis.

Transferase genes: WbwA shares 30% identity with Lst of *Haemophilus ducreyi.* Lst is a NeuNAc transferase for making the linkage α-NeuNAc-(2-3)-β-Gal (Bozue et al., 1999), and *E. coli* O104 has the same linkage and we suggest that WbwA is the expected NeuNAc transferase. The C-terminal half of the WbwB shares 27% identity (53% similarity) with the C-terminal part of WaaB, a galactosyl transferase forming an α (1-6) linkage to α glucose in the synthesis of *E. coli* R2 core (GenBank entry. AF019375). Thus, it is likely that WbwB forms the α-Gal-(1-4)-α-NeuNAc linkage. WbwC shares similarity with many other putative bacterial polysaccharide transferases, which include WbbD of *E. coli* O7 (GenBank entry. U23775) with 38% identity. The *E. coli* O7 antigen contains a β-Gal-(1-3)-α-GlcNAc linkage and we suggest that WbwC is the galactosyl transferase forming the same linkage but to β-GalNAc to give a β-Gal-(1-3)-β-GalNAc in the O104 structure.

Typical *E. coli* O antigen gene clusters have a *wzx* and *wzy* gene. Both encode hydrophobic proteins and two such genes were found in the O104 gene cluster. One encodes a potential integral inner membrane protein with 12 predicted transmembrane segments and was treated as putative *wzx* gene. Wzx proteins can be difficult to identify with confidence by sequence searches as sequence identity levels are low, and motif searches are often more convincing. The putative Wzx protein of O104 was grouped by sequence comparisons with a range of Wzx proteins of O antigen gene clusters of *E. coli* O111 (GenBank entry: AF078736) and *E. coli* K-12 (GenBank entry. U09876): analysis using the BLOCKMAKER program (Henikoff et al., 1995) revealed two motifs which are conserved among this group of proteins and the consensus sequence was used to run the program PSI-Blast (Altschul et al., 1997) to search the Genpept database: the Wzx proteins of *E. coli* O111 and *E. coli* K-12 and many other distantly related Wzx proteins but no other proteins were retrieved (E-value $\leq 5e \times 10^{-4}$) after one iteration, confirming the designation.

The second of the two genes encodes a protein having ten predicted transmembrane segments with a periplasmic loop of 33 a.a. residues, a characteristic topology Wzy. Two motifs were found between this protein and Wzy proteins of *E. coli* O4 (GenBank entry: U39042) and *S. enterica* group B (GenBank entry: M84642) O clusters. PSI-Blast search using the consensus sequence retrieved the two input Wzy proteins and many other distantly related Wzy proteins but no other proteins (E-value $\leq 7e \times 10^{-8}$) after two iterations, confirming the designation. We conclude that it is the O antigen polymerase gene, *wzy*.

In summary, three putative transferase genes, three CMP-NeuNAc synthesis, a putative O antigen polymerase gene, and a putative flippase gene were identified. They account for all genes needed for the synthesis and processing of the O104 O unit, except for a gene for the transferase adding the first sugar, GalNAc, to the carrier lipid Undecaprenol phosphate (UndP) as a sugar phosphate. The WckD protein has no resemblance to known sugar transferase genes and is thought to be involved in facilitation of CMP-NeuNAc synthesis and not GalNAc transfer (see above). It is known that the first sugar, which links the O antigen to the core, is GalNAc in the O104 antigen (Gamian et al., 1992). Other than *wzx* and *wzy*, none of the genes in the cluster encode predicted integral membrane proteins, whereas the two proteins known to initiate O unit synthesis by transfer of a sugar phosphate to UndP, WecA (Rfe) and WbaP (RfbP), have several predicted transmembrane segments (Jiang et al., 1991; Meier-Dieter et al., 1992). We therefore suggest that as proposed on quite good grounds for *Yersinia enterocolitica* (Zhang et al., 1997) and *E. coli* O157 (Wang and Reeves, 1998), WecA transfers GalNAc phosphate to UndP to initiate O unit synthesis as well as initiating Enterobacterial common antigen synthesis by transfer of GlcNAc phosphate to UndP. The case is greatly strengthened by the observation of Amor and Whitfield (1997) that a *wecA* mutant blocks formation of the K9 antigen, known to have the same structure as the O104 (Dutton and Parolis, 1989) and shown by us (see below) to have the same genes in the same order.

3.3. The E. coli K9 and O104 gene clusters have the same genes and organization

The oligosaccharide unit of *the E. coli* K9 antigen has an identical structure to that of the O104 antigen (Dutton and Parolis, 1989). The *E. coli* K9 antigen belongs to the group IB K antigens, which are generally only present *in E. coli* strains with O antigens O8 and O9 (Whitfield and Roberts, 1999). Type IB K antigen genes are located in the normal O antigen region between *galF* and *gnd* (Whitfield and Roberts, 1999), while the *E. coli* O8 and O9 genes are located on the other side of the *gnd* gene between *gnd* and *his* (Jayaratne et al., 1994; Kido et al., 1995). This is the first sequence of an O antigen gene cluster where an antigen identified as a typical O antigen has the same structure as

a K antigen, so we carried out work to determine the level of similarity between the K9 and O104 gene clusters.

PCR was carried out on an *E. coli* O9:K9 strain (Bi316-42, the type strain for K9), using primers based on the O104 sequence. The primer pairs used each had primers from adjacent genes and covered each pair of adjacent genes from *galF* to *gnd*. Each PCR gave identical bands from the K9 and O104 strains (data not shown), and we conclude that the K9 gene cluster has the same genes in the same order as that of the O104 cluster.

3.4. Identification of O104 specific genes

We have shown that oligonucleotide primers based on the *E. coli* O157 and O111 O antigen transferase, *wzx* and *wzy* genes are serotype specific (Wang et al., 1998; Wang and Reeves, 1998). In this study, we tested primers based on O104 *wzy*, *wzx*, and *wbwB*.

Five pairs of oligonucleotide primers based on the O104 sequence (Table 2) were tested by PCR against each of 27 DNA pools. Each of the five pairs binding to *wbwB*, *wzx*, and *wzy* produces a band of predicted size with the pools (pools 14 and 21) containing DNA obtained from type strains of O104 and O9 (an O9:K9 strain) and none with any other pool (Table 1). The pools include DNA from strains representing the 166 known *E. coli* O antigens, five Shigella O antigens, seven *Vibrio cholerae* O antigens, and seven *Streptococcus pneumoniae* capsules. As pools 22 and 23 included DNA from all strains present in pools 14 and 21 other than DNA from the O104 and O9 strains, we conclude that the ten pairs of primers all give a positive PCR test only with strains carrying O104 or K9 genes. Thus, *wbwB*, *wzx* and *wzy* genes are highly specific for the O104 and K9 gene clusters.

As K9 is found only in association with O8 and O9 antigens, we also identified primers specific to O8 and O9 genes and these can be used to distinguish *E. coli* O104 strains from O8:K9 and O9:K9 strains. The *E. coli* O8, O9a and O9 antigens are similar to each other, with 3, 4 and 5 mannose residues, respectively (Jann and Jann, 1984; Sugiyama et al., 1998). Gene clusters for both O8 and O9a antigens have been sequenced (Kido et al., 1995; Sugiyama et al., 1998), and genes located at the ends of the two gene clusters are highly similar (Sugiyama et al., 1998). We tested primers based on three transferase genes of O9a (one pair per gene), and found that primers #3592 (GGCATCGGTCGG-TATTCC)/#3594 (TGCGCTAATCGCGTCTAC), based on the *wbdD* gene produce a band of predicted size with the pools 1 and 21 containing DNA obtained from type strains of O8 and O9 and none with any other pools (Table 1) including the pool containing O104 DNA. As pools 24 and 23 included DNA from all strains present in pools 1 and 21 other than DNA from the O8 and O9 strains, we conclude that this pair of primers is specific to O8, O9a and O9 strains (O9a on the basis of known sequence).

3.5. General conclusions

We now have the sequence of the entire *E. coli* O104 O antigen gene cluster and have identified with varying degrees of precision all genes required for synthesis of the O antigen. The *E. coli* O104 antigen has the same structure as a group IB K9 antigen, and we showed that the K9 antigen gene cluster has the same organization as that of O104. This is the first time that we can compare sequences of gene clusters encoding O and K antigens with the same structure: it is interesting that the two gene clusters are essentially the same. O antigen is linked to the Lipid A-core molecule to form a complete LPS molecule, which is then transferred to the bacterial outer membrane. K antigen is a capsule on the outside of the cell proper, and is a higher molecular weight molecule with many more repeat units than O antigen. The *E. coli* K40 (type IB) gene cluster has been cloned and sequenced, and was shown to be located at the typical O antigen locus (between *galF* and *gnd*) with all the typical O antigen gene cluster features (Amor and Whitfield, 1997). The K40 sequence had no characteristic to differentiate it from typical *E. coli* O antigen gene cluster, and our results both confirm that and show that for a structure reported for both a K and an O antigen, the gene clusters are identical at least in genes present and gene order as determined by PCR. Our results and the K40 sequence demonstrate that the factor(s) which determine the channelling of the repeat unit into either capsule or O antigen lie outside of the antigen gene cluster.

We have identified three genes specific to the gene clusters coding for both O104 and K9 antigens. The K9 capsule is only present in *E. coli* O8 or O9/9a strains (note O9 and O9a were recently distinguished and for most strains it is not known if O9 or O9a was present). We have identified primers specific to O8, O9 and O9a antigens. In routine use, separate PCR reactions using O104 and O8/O9/O9a specific primers can be carried out. Positive PCR results for both would indicate the presence of strains carrying the K9 antigen. Positive

Table 2
O104 specific primers

Gene	Base positions of the gene	Forward primer (base positions)	Reverse primer (base positions)
wzy	6226—7338	#3580 (6308–6325)	#3581 (7319–7302)
		#3629 (6493–6511)	#3630 (6952–6935)
wbwB	7344—8567	#3582 (7420–7437)	#3583 (8857–8540)
		#3631 (7663–7680)	#3632 (8283–8300)
wzx	8549—9844	#3633 (9099–9116)	#3634 (9630–9613)

PCR results for O104 primers only would indicate the presence of strains carrying the O104 antigen.

This work forms the basis for development of PCR based methods for identification of O104 strains to replace more time consuming plating and serotyping methods. The *E. coli* serotyping scheme is not yet fully comprehensive and there are known to be other as yet unidentified O antigens. For this reason field strains and conditions need to be tested to confirm the specificity, although we believe that all or most of these genes will be specific to O104 strains. Further specificity can be gained by use of a combination of these genes, perhaps by PCR using primers binding to adjacent genes.

Acknowledgements

We thank all the people and institutes listed in Table 1 for kindly supplying strains. This study was partly supported by the Australian Research Council. LW is a recipient of an OECD fellowship which covered travel and living allowances for work at the Microbial Food Safety Research Unit, Eastern Regional Research Center, USDA to carry out part of the work.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3398–3402.

Amor, P.A., Whitfield, C., 1997. Molecular and functional analysis of genes required for expression of group IB K antigens in *Escherichia coli*: characterization of the his-region containing gene clusters for multiple cell-surface polysaccharides. Mol. Microbiol. 26, 145–161.

Annunziato, P.W., Wright, L.F., Vann, W.F., Silver, R.P., 1995. Nucleotide sequence and genetic analysis of the *neuD* and *neuB* genes in region 2 of the polysialic acid gene cluster of *Escherichia coli* K1. J. Bacteriol. 177, 312–319.

Bozue, J.A., Tullius, M.V., Wang, J., Gibson, B.W., Munson, R.S., 1999. *Haemophilus ducreyi* produces a novel sialyltransferase. Identification of the sialyltransferase gene and construction of mutants deficient in the production of the sialic acid-containing glycoform of the lipooligosaccharide. J. Biol. Chem. 274, 4106–4114.

Daines, D.A., Silver, R.P., 2000. Evidence for multimerization of Neu proteins involved in polysialic acid synthesis in *Escherichia coli* K1 using improved LexA-based vectors. J. Bacteriol. 182, 5267–5270.

Daines, D.A., Wright, L.F., Chaffin, D.O., Rubens, C.E., Silver, R.P., 2000. NeuD plays a role in the synthesis of sialic acid in *Escherichia coli* K1. FEMS Microbiol. Lett. 189, 281–284.

Dutton, G.G.S., Parolis, L.A.S., 1989. Polysaccharide antigens of *Escherichia coli*. In: Dea, I.G.M.e.a. (Ed.). Biomedical and Biotechnological Advances in Industrial Polysaccharides, Proceedings of the 3rd International Workshop on Recent Developments in Industrial Polysaccharides, New York.

Eisenberg, D., Schwarz, E., Komaromy, M., Wall, R., 1984. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. J. Mol. Biol. 179, 125–142.

Gamian, A., Romanowska, E., Ulrich, J., Defaye, J., 1992. The structure of the sialic acid-containing *Escherichia coli* O104 O-specific polysaccharide and its linkage to the core region in lipopolysaccharide. Carbohydrate Res. 236, 195–208.

Henikoff, S., Henikoff, J.G., Alford, W.J., Pietrokovski, S., 1995. Automated construction and graphical presentation of protein blocks from unaligned sequences. Gene 163, GC17–GC26.

Jann, K., Jann, B., 1984. Structure and biosynthesis of O-antigen. In: Rietschel, E.T. (Ed.), Handbook of Endotoxin. Elsevier Scientific, Amsterdam, pp. 138–186.

Jayaratne, P., Bronner, D., MacLachlan, P.R., Dodgson, C., Kido, N., Whitfield, C., 1994. Cloning and analysis of duplicated *rfbM* and *rfbK* genes involved in the formation of GDP-mannose in *Escherichia coli* O9:K30 and participation of *rfb* genes in the synthesis of the group I K30 capsular polysaccharide. J. Bacteriol. 176, 3126–3139.

Jiang, X.M., Neal, B., Santiago, F., Lee, S.J., Romana, L.K., Reeves, P.R., 1991. Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). Mol. Microbiol. 5, 695–713.

Kido, N., Torgov, V.I., Sugiyama, T., Uchiya, K., Sugihara, H., Komatsu, T., Kato, N., Jann, K., 1995. Expression of the O9 polysaccharide of *Escherichia coli*: sequencing of the *E. coli* O9 *rfb* gene cluster, characterization of mannosyl transferases, and evidence for an ATP-binding cassette transport system. J. Bacteriol. 177, 2178–2187.

Lior, H., 1994. Classification of *Escherichia coli*. In: Gyles, C.L. (Ed.), *Escherichia coli* in domestic animals and humans. CAB International, Wallingford, UK, pp. 31–72.

Meier-Dieter, U., Barr, K., Starman, R., Hatch, L., Rick, P.D., 1992. Nucleotide sequence of the *Escherichia coli rfe* gene involved in the synthesis of enterobacterial common antigen. J. Biol. Chem. 267, 746–753.

Petersen, M., Fessner, W., Frosch, M., Luneberg, E., 2000. The *siaA* gene involved in capsule polysaccharide biosynthesis of *Neisseria meningitidis* B codes for N-acylglucosamine-6-phosphate 2-epimerase activity. FEMS Microbiol. Lett. 184, 161–164.

Reeves, P.R., 1994. Biosynthesis and assembly of lipopolysaccharide. In: Neuberger, A., van Deenen, L.L.M. (Eds.), Bacterial Cell Wall. Elsevier Science, Amsterdam, pp. 281–314.

Reeves, P.R., Hobbs, M., Valvano, M., Skurnik, M., Whitfield, C., Coplin, D., Kido, N., Klena, J., Maskell, D., Raetz, C., Rick, P., 1996. Bacterial polysaccharide synthesis and gene nomenclature. Trends Microbiol. 4, 495–503.

Sugiyama, T., Kido, N., Kato, Y., Koide, N., Yoshida, T., Yokochi, T., 1998. Generation of *Escherichia coli* O9a serotype, a subtype of *E. coli* O9, by transfer of the wb* gene cluster of *Klebsiella* O3 into *E. coli* via recombination. J. Bacteriol. 180, 2775–2778.

Wang, L., Curd, H., Qu, W., Reeves, P.R., 1998. Sequencing of *Escherichia coli* O111 O antigen gene cluster and identification of O111 specific genes. J. Clin. Microbiol. 36, 3182–3187.

Wang, L., Reeves, P.R., 1998. Organization of *Escherichia coli* O157 O antigen gene cluster and identification of its specific genes. Infect. Immun. 66, 3545–3551.

Whitfield, C., Roberts, I., 1999. Structure, assembly and regulation of expression of capsules in *Escherichia coli*. Mol. Microbiol. 31, 1307–1319.

Zhang, L., Radziejewska-Lebrecht, J., Krajewska-Pietrasik, D., Toivanen, P., Skurnik, M., 1997. Molecular and chemical characterization of the lipopolysaccharide O-antigen and its role in the virulence of *Yersinia enterocolitica* serotype O8. Mol. Microbiol. 23, 63–76.