JOURNAL OF BACTERIOLOGY, Aug. 1996, p. 4885–4893
0021-9193/96/$04.00+0

Vol. 178, No. 16

Organization of the *Escherichia coli* K-12 Gene Cluster Responsible for Production of the Extracellular Polysaccharide Colanic Acid

GORDON STEVENSON, KANELLA ANDRIANOPOULOS, MATTHEW HOBBS, AND PETER R. REEVES*

Department of Microbiology, University of Sydney, New South Wales 2006, Australia

Received 3 November 1995/Accepted 18 June 1996

Colanic acid (CA) is an extracellular polysaccharide produced by most *Escherichia coli* strains as well as by other species of the family *Enterobacteriaceae*. We have determined the sequence of a 23-kb segment of the *E. coli* K-12 chromosome which includes the cluster of genes necessary for production of CA. The CA cluster comprises 19 genes. Two other sequenced genes (*orf1.3* and *galF*), which are situated between the CA cluster and the O-antigen cluster, were shown to be unnecessary for CA production. The CA cluster includes genes for synthesis of GDP-L-fucose, one of the precursors of CA, and the gene for one of the enzymes in this pathway (GDP-D-mannose 4,6-dehydratase) was identified by biochemical assay. Six of the inferred proteins show sequence similarity to glycosyl transferases, and two others have sequence similarity to acetyl transferases. Another gene (*wzx*) is predicted to encode a protein with multiple transmembrane segments and may function in export of the CA repeat unit from the cytoplasm into the periplasm in a process analogous to O-unit export. The first three genes of the cluster are predicted to encode an outer membrane lipoprotein, a phosphatase, and an inner membrane protein with an ATP-binding domain. Since homologs of these genes are found in other extracellular polysaccharide gene clusters, they may have a common function, such as export of polysaccharide from the cell.

Polysaccharides are important constituents of the surface of the bacterial cell, and they play a critical role in its interaction with the environment. Many bacteria produce extracellular polysaccharides (EPSs), which can remain attached to the cell in a capsular form or alternatively be released as a slime. Gram-negative bacteria also produce lipopolysaccharide (LPS), which forms part of the outer membrane and which has a polysaccharide component called the O antigen. These polysaccharides are prominent antigens and are structurally and serologically diverse. In *Escherichia coli* serological studies have so far found 173 O antigens and 80 capsular antigens (K antigens) (63, 89). The *E. coli* K antigens have been divided into four groups (IA, IB, II, and III) based on a variety of criteria (33, 34, 64, 66). Colanic acid (CA) or M antigen (Fig. 1) is an EPS which is widely found within *E. coli*, as well as within other species of the family *Enterobacteriaceae* (28). This is in contrast to the K antigens, each of which is found in only a small proportion of *E. coli* strains. It has been suggested that CA is important for survival of *E. coli* outside the host and perhaps has a role in resistance to desiccation (62).

CA resembles the group IA capsular polysaccharides in several respects: high molecular weight, sugar composition, and the chromosomal location near *his* of the responsible biosynthetic genes (83). Furthermore, it has also been shown that synthesis of type I capsular polysaccharides is under control of *rcs* (regulator of capsule synthesis) genes, *rcsA*, *rcsB*, and *rcsC* (38), previously described as a control system for CA (for reviews, see references 26 and 27). Over the years there has been considerable interest in the transcriptional control of CA expression. Despite this interest, little is known about the nature of the genes responsible for CA production. In *E. coli* K-12 these genes map to a region (formerly *cps*) (83), which is upstream of the genes (formerly *rfb*) for production of O antigen. (In this paper we have renamed some genes and loci in accordance with a proposed new polysaccharide gene nomenclature (70) which it is hoped will do away with ambiguous and confusing gene names which have been accruing.) We have previously determined the DNA sequence of a ~6-kb segment of the K-12 CA gene cluster (4), and in this paper we report the completion of the sequencing of the entire gene cluster.

* Corresponding author. Mailing address: Department of Microbiology, Building G08, University of Sydney, NSW 2006, Australia. Phone: (612) 351 2536. Fax: (612) 351 4571. Electronic mail address: reeves@angis.su.oz.au.

MATERIALS AND METHODS

Strains and plasmids. *E. coli* K-12 strains and plasmids used in this study are listed in Table 1. Cultures for general use were grown at 37°C in nutrient broth (NB) (peptone, 10 g/liter; yeast extract [Gibco], 5 g/liter; sodium chloride, 5 g/liter) or on nutrient agar plates (1.5% bacteriological agar in NB). Antibiotics were added (when appropriate) to the following concentrations: ampicillin, 25 μg/ml; kanamycin, 50 μg/ml; tetracycline, 16 μg/ml.

General DNA manipulations. DNA manipulations, including PCR, restriction digestion, and ligations, were performed by standard methods (76). PCR fragments were cloned in the plasmid vector pGEM5Zf(+) by using the pGEM-T A-T cloning system (ProMega Corp.).

DNA sequencing. Random subclones of pPR1653 (Table 1 and Fig. 2) were created by partial digestion of the insert with DNaseI and cloning of the resulting fragments in pGEM7Zf(+). Preparations of double-stranded DNA of these plasmids were used as templates in sequencing reactions. The templates for all other sequencing reactions were PCR products produced with oligonucleotides tagged with the forward or reverse M13 sequencing primers. Thermocycle sequencing reactions based on the dideoxy termination method (77) were run in a Perkin-Elmer Cetus DNA Thermal Cycler using a procedure recommended by Perkin-Elmer Cetus. The reactions used primers labelled with the appropriate fluorescent dye and were run on an Applied Biosystems 373A automated DNA sequencer.

Sequence analysis. Computer programs for the editing and analysis of DNA were accessed through the Australian National Genomic Information Service (ANGIS) (71). Sequence data were assembled into contigs with the XDAP program (23). Sequence databases were searched with the National Center for Biotechnology Information BLAST network server (2). Analysis of open reading frames (ORFs) was carried out using the nucleotide interpretation program (80). Several Genetics Computer Group package (15) programs were used, including PILEUP for aligning multiple sequences and BESTFIT for pairwise sequence comparisons. The program RDF2 (67) was used to assess the statistical significance of pairwise alignments. The program ALOM (41) was used to identify potential transmembrane segments.

GMD assay. *gmd* was amplified with primers 471 (5′-AATCCCGCAGTGGT GATACGC-3′) and 472 (5′-CGCCTGATGGCGGAACCGACC-3′). The PCR

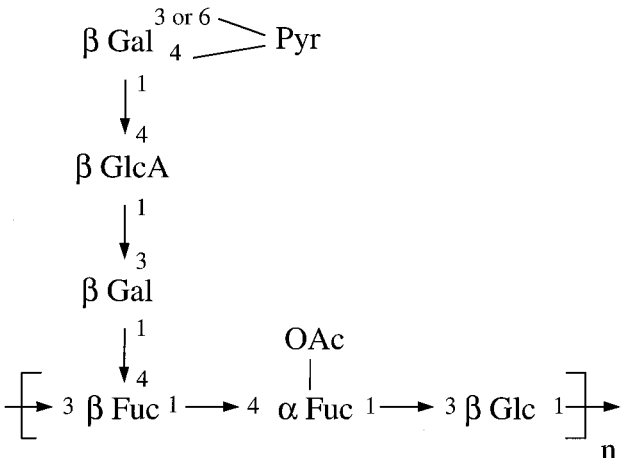

FIG. 1. Structure of CA of *E. coli* K-12 (3, 21, 28, 53). Fuc, L-fucose; Gal, D-galactose; GlcA, D-glucuronic acid; Glc, D-glucose; OAc, O-acetyl; Pyr, pyruvate linked acetalically to galactose.

product was cloned by using the pGEM-T system to give pPR1689 (Table 1 and Fig. 2), in which *gmd* is in the opposite transcriptional orientation to the *lacZ* gene. pPR1689 was introduced into *E. coli* SØ874, which has a large chromosomal deletion that includes the CA gene cluster. Crude protein extracts were prepared from the resulting strain by sonication of cell pellets suspended in 50 mM Tris (pH 7.5)–10 mM $MgCl_2$–1 mM EDTA–5 mM 2-mercaptoethanol. The extracts were tested for GDP-mannose 4,6-dehydratase (GMD) activity by using the spectrophotometric assay of Kornfeld and Ginsberg (45).

Generation of kanamycin resistance insertions. Recombinant plasmids pPR1742, pPR1743, and pPR1744 (Table 1 and Fig. 2) have insertional mutations in the cloned genes *wcaL*, *orf1.3*, and *galF*, respectively, and were created as follows. In each case two PCR reactions were performed to give amplification products which contained a portion of the target gene together with some flanking DNA. The PCR primers within the target gene incorporate an *Eco*RI site and this was used in joining the two PCR fragments together: each pair of segments was cloned together in the plasmid vector pGEM5Zf(+) in such a way as to give a construct with a novel and unique *Eco*RI site within the reconstituted target gene. The GenBlock *Eco*RI kanamycin resistance gene cassette (Pharmacia) was inserted into the *Eco*RI site. The primers used for PCR amplifications were as follows: 621 (5′-ATGCCATCTAAAAAACTCTC-3′) and 668 (5′-AAAGAATTCTCAGCATTGCATCAACGG-3′) for the left-hand end of the pPR1742 insert, 667 (5′-AAAGAATTCACTGGAACGTACTCAGCC-3′) and 679 (5′-GGTGCATCACTGCATCCG-3′) for the right-hand end of the pPR1742 insert, 622 (5′-ATGAAGGTCGGCTTCTTTTTAC-3′) and 694 (5′-AAAGAATTCTCCCGCAAGGCCTATGCC-3′) for the left-hand end of the pPR1743 insert, 693 (5′-AAAGAATTCTGACCATGTCATCGAGCG-3′) and 679 for the right-hand end of the pPR1743 insert, 687 (5′-′TGAAAGGGCAACTGGCGC-3′) and 720 (5′-AAAGAATTCTTCCGGCCCACAGATCGC-3′) for the left-hand end of the pPR1744 insert, and 721 (5′-AAAGAATTCTCAGCAACTGTTTCGCCG-3′) and 686 (5′-CCACGAACGTCTTTGCGC-3′) for the right-hand end of the pPR1744 insert.

The mutations were transferred into the *E. coli* chromosome by homologous recombination. The cloned DNA incorporating the kanamycin gene cassette was purified and the linear DNA fragment was then electroporated into the *E. coli recD* strain V355. Transformants were selected on kanamycin, and putative recombinants were screened for ampicillin sensitivity (i.e., original donor plasmid not present) and subsequently tested by PCR for the presence of the kanamycin resistance cassette at the expected site. In each of these confirmatory PCR reactions the primers used were on either side of the target site, and one was located outside of the segment of donor DNA. As expected, chromosomal DNA from mutants gave rise to amplification products of ~1.3 kb (i.e., the size of the inserted kanamycin resistance gene cassette) larger than was the case with a wild-type control. The ability of mutants to produce colanic acid was tested by introducing the plasmid pATC400 (which induces expression of colanic acid genes) and looking for the characteristic mucoid colonial morphology.

wcaL was amplified with primers 698 (5′-GGTTGCGGATCCCTTAGCCAGC-3′) and 699 (5′-AGATGGCATGCGTTCTCCTC-3′). The PCR product was cloned by using the pGEM-T system to give pPR1745 in which *wcaL* is in the opposite transcriptional orientation to *lacZ*. The pPR1745 insert was then subcloned in pACYC184 to give pPR1746 (Table 1), which was used to complement the mutation in P5340.

Nucleotide sequence accession number. The DNA sequence has been submitted to the GenBank and EMBL databases and assigned the accession number U38473.

RESULTS AND DISCUSSION

Size of the CA gene cluster. The promoter region of the CA gene cluster has recently been studied and sequenced (81a). We took this to be the 5′ end of the cluster and extended the sequence from the promoter region through our previously published sequence (4) to the beginning of the O-antigen gene

TABLE 1. Strains and plasmids used in this study

Strain or plasmid and laboratory stock no	Relevant characteristics	Source or reference
Strains		
JM109 (P3584)	*supE44 nalA96 recA1 gyrA96 relA1 endA1 thi hsdR17* Δ(*proAB-lac*), F′(*traD36 proAB lac*)	90
SØ874 (P4052)	*lacZ4503 try-355 upp-12 relA rpsL150* Δ(*sbcB-rfb*)	60
V355		
P5341	*lac-3350 galT22 galK2 recD1014*(Nuc$^-$), IN(*rrnD-rrnE*)*1*, *rpsL179*(*strR*)	11
P5336	V355 with kanamycin resistance gene insertion in *galF*	This study
P5338	V355 with kanamycin resistance gene insertion in *orf1.3*	This study
P5340	V355 with kanamycin resistance gene insertion in *wcaL*	This study
Phage K351-K355	Phage lambda clones of K-12 strain W3110	44
Plasmids		
pGEM5zf(+)	Cloning vector; pMB1 replicon; ampicillin resistance	ProMega Co.
pGEM7zf(+)		
pACYC184	Cloning vector; p15A replicon; tetracycline and chloramphenicol resistance	9
pATC400	*rcsA* gene cloned in pBR322; induces expression of CA genes	75
pPR1653	8-kb *Eco*RI-*Bam*HI fragment from K354 cloned in pGEM7Zf(+)	This study
pPR1689	PCR fragment containing *gmd* cloned in pGEM5Zf(+)	This study
pPR1742	Two PCR products together encompassing *galF* and flanking DNA cloned in pGEM5Zf(+), with kanamycin resistance gene cassette subsequently inserted into *galF*; used to generate P5336	This study
pPR1743	Two PCR products together encompassing *orf1.3* and flanking DNA cloned in pGEM5Zf(+), with kanamycin resistance gene cassette subsequently inserted into *orf1.3*; used to generate P5388	This study
pPR1744	Two PCR products together encompassing *wcaL* and flanking DNA cloned in pGEM5Zf(+), with kanamycin resistance gene cassette subsequently inserted into *wcaL*; used to generate P5340	This study
pPR1745	PCR fragment containing *wcaL* cloned in pGEM5Zf(+)	This study
pPR1746	1.3-kb *Eag*I-*Bam*HI fragment from pPR1745 cloned in pACYC184; compatible with pATC400	This study

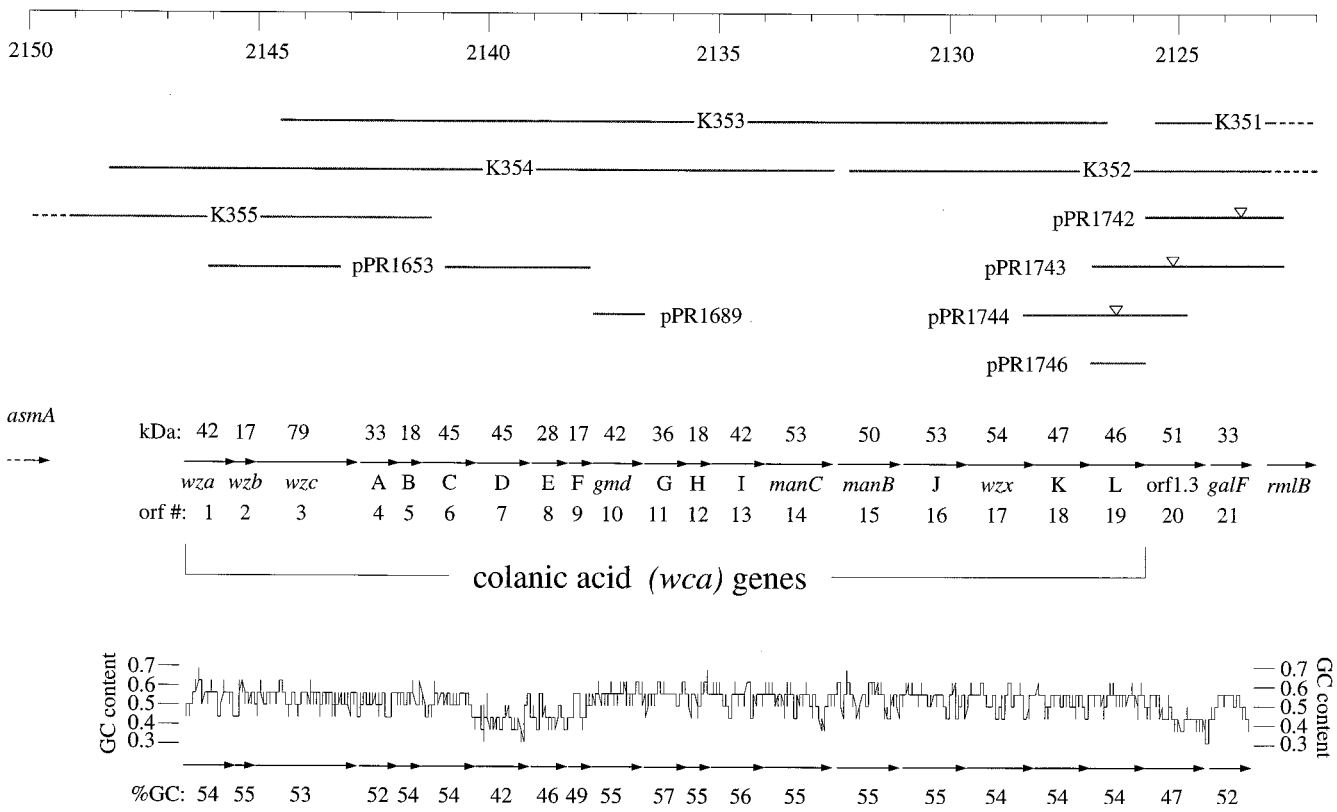

FIG. 2. Genetic organization of the *wca* (*cps*) region responsible for production of CA. The scale shows the location (in kilobases) on the Eco6 physical map of the *E. coli* K-12 chromosome (74). Horizontal lines represent the inserts of recombinant clones (Table 1). Triangles show the point of insertion of a kanamycin gene cassette used in mutagenesis. Arrows represent genes, which are transcribed in a rightward direction. *rmlB* (*rfbB*) is the first gene of the O-antigen gene cluster (81), and *asmA* is a gene involved in assembly of outer membrane proteins (54); neither is involved in CA production. The DNA sequence of the region between *asmA* and *wza* is not known. The sequenced ORFs (1 to 21) under consideration are shown, together with the sizes of predicted translation products. Gene names have been assigned to ORFs in accordance with the proposal of Reeves et al. (70): *manB*, *manC*, and *gmd* have established functions in the GDP-L-fucose biosynthetic pathway (Fig. 3); *wza*, *wzb*, *wzc*, and *wzx* appear to have a general role in EPS production; and other genes implicated in CA production have been given *wca* names (A to L). *orf1.3* and *galF* have no demonstrable role in CA production. Six of the CA genes (*gmd*, *wcaGHI*, and *manCB*) were previously sequenced (4) and were called (respectively) *orf0.0*, *orf0.9*, *orf1.9*, *orf2.4*, and *cpsBG*. *wcaL*, *orf1.3*, and *galF* are equivalent to the *S. enterica* LT2 genes called *orf0.0*, *orf1.3*, and *orf2.8* by Jiang et al. (35). The graph shows GC content of the *wca* cluster (mean value calculated for a window of 101 bases), and individual gene GC contents are also given.

cluster (81) (Fig. 2). This segment of DNA is 23 kb in length (Fig. 2). The promoter region is followed after about 60 bp by a JUMPstart sequence (81a), a highly conserved 39-bp element which is found upstream of many polysaccharide gene clusters (30). There are then 21 open reading frames all in the same transcriptional orientation (Fig. 2). The first of these begins 179 bp downstream of the JUMPstart sequence, and the last ends 373 bp upstream of *rmlB* (*rfbB*), the first gene of the O-antigen gene cluster. There is only one obvious transcriptional termination sequence in the region sequenced which is a 25-bp inverted repeat sequence 41 bp downstream of *orf21* (*galF*; see below), and 161 bp upstream of the JUMPstart sequence which preceeds the O antigen gene cluster. This suggests that the 21 genes are part of a single large transcription unit, but this has not been tested experimentally. Also, *galF* is preceeded by a potential promoter sequence (35) and so may also be independently transcribed.

orf19, *orf20*, and *orf21* of this study are equivalent to *orf0.0*, *orf1.3*, and *orf2.8* of *Salmonella enterica* LT2 (35). It has been suggested (35, 86) that *S. enterica orf2.8* could be the gene *galF*, identified by mutation and reported to have an effect on the chromatographic behavior of UDP-glucose pyrophosphorylase (57–59, 61). Recent work (83a) indicates that *orf2.8* does indeed encode functions similar to those described for *galF*, confirming that it is the *galF* gene described previously.

The 3′ end of the CA gene cluster was studied by making insertions in *orf19*, *orf20*, and *galF* in strain V355. The effect of these mutations on CA synthesis was determined by adding the *rcsA*-containing plasmid pATC400. RcsA is a positive control protein for the CA cluster, and *rcsA* on a multicopy plasmid confers mucoidy to a strain (75). After addition of pATC400 strains P5336 and P5338 (carrying insertions in *orf20* or *galF*, respectively) were mucoid. In contrast, P5340 carrying an insertion in *orf19* was nonmucoid. The defect in P5340 was corrected by introducing a second plasmid (pPR1746) carrying *orf19*, confirming that the only defect in P5340 was the insertion into *orf19*. We conclude that *orf19* is required for CA synthesis and is part of the CA gene cluster and name it *wcaL*, and we also give *wca* gene names to all ORFs upstream of *wcaL* where appropriate (Fig. 2).

galF is clearly not necessary for synthesis of CA. *galF* is a homolog of *galU* (58% sequence identity at the amino acid level) but apparently cannot compensate for loss of *galU* function since *galU* mutants are defective in production of CA (86, 87). In *S. enterica* the effect of deletion of *galF* is to change the kinetics of UDP-glucose synthase (57–59). While this may relate to regulation of synthesis of UDP-galactose for either LPS or CA synthesis, the loss of *galF* was not expected to lead to loss of ability to synthesize CA, and the data confirm this. *orf20* is also unnecessary for synthesis of CA, although it should be

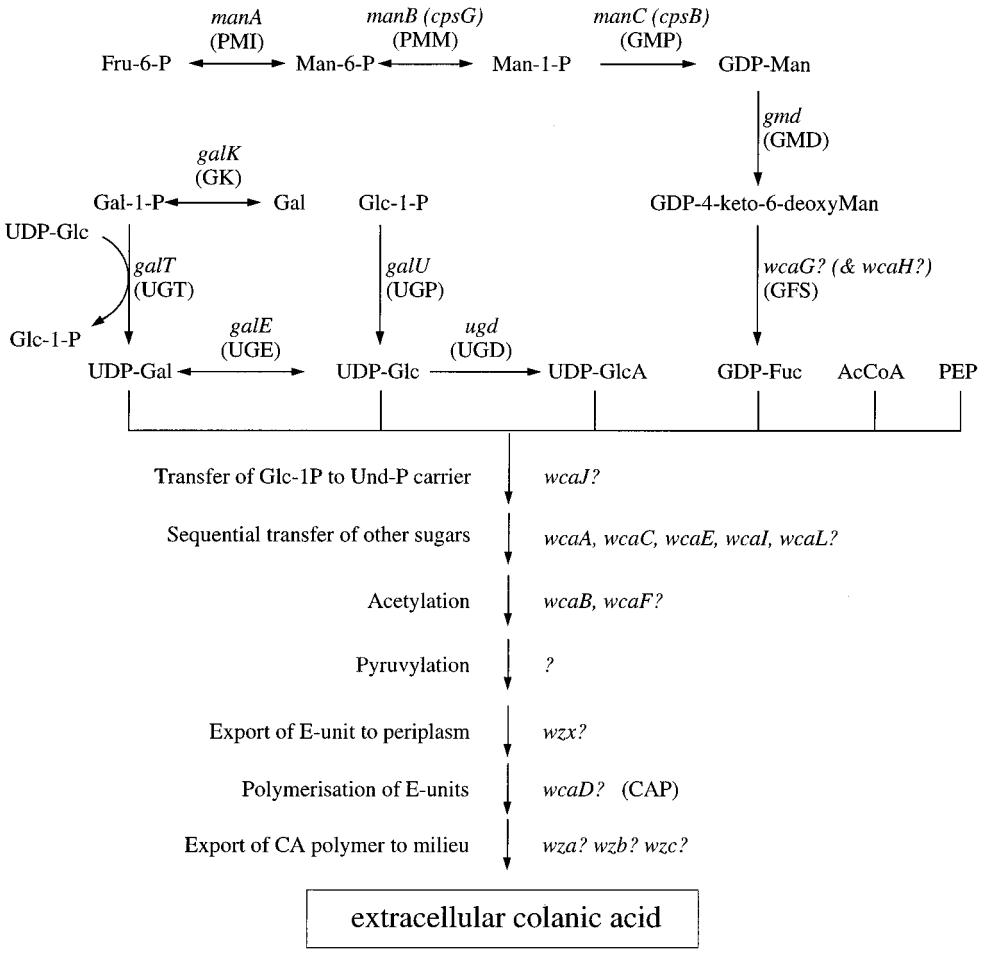

FIG. 3. Proposed pathway for biosynthesis of CA in *E. coli* K-12. Abbreviated enzyme names (in brackets) are shown, together with the encoding genes. Note that several gene assignments are speculative and that the stage at which acetylation of the E unit occurs is not known. AcCoA, acetyl coenzyme A; CAP, colanic acid polymerase; Fru, fructose; GFS, GDP-L-fucose synthetase; GK, galactokinase; GLK, glucokinase; GMP, GDP-D-mannose pyrophosphorylase; GMD, GDP-D-mannose dehydratase; Man, D-mannose; PEP, phosphoenolpyruvate; PGI, phosphoglucose isomerase; PGM, phosphoglucomutase; PMI, phosphomannose isomerase; PMM, phosphomannomutase; UGD, UDP-D-glucose dehydrogenase; UGE, UDP-D-galactose-4-epimerase; UGP, UDP-D-glucose pyrophosphorylase; UGT, UTP-D-galactose-1-phosphate uridylyl transferase. Other abbreviations are as in Fig. 1.

noted that the CA repeat unit has a three-sugar side branch and it is possible that a mutation in one of the side branch transferases would not affect polymerization of the incomplete repeat unit. Thus, although it is not proven, it is still quite possible that *orf20* is part of the CA gene cluster, but demonstration of a role in CA synthesis must await future research. We have therefore not given this orf a *wca* gene name but rename it *orf1.3* to maintain consistent nomenclature with *S. enterica*.

The CA gene cluster then comprises 19 or 20 genes. On the basis of sequence characteristics we have been able to assign possible functions to most of the CA genes (Fig. 3).

Genes for synthesis of nucleotide sugar precursors of CA. Genes responsible for production of the nucleotide sugars used as building blocks in polysaccharide synthesis are often found at the polysaccharide gene loci. Of the four different nucleotide sugars necessary for production of the CA repeat unit (the E unit), two (UDP-D-glucose and UDP-D-galactose) are used by the cell for other purposes and housekeeping genes responsible for their production (*galU*, *galT*, *galK*, and *galE*; Fig. 3) are located elsewhere on the K-12 chromosome.

Synthesis of a third CA sugar (UDP-D-glucuronic acid) from UDP-D-glucose is catalyzed by a single enzyme, UDP-glucose dehydrogenase (UGD). The *hasB* gene encoding UGD has been identified in *Streptococcus pneumoniae* (17). A very similar gene (*ugd*, with 55% amino acid identity to *hasB*) has been found in *E. coli* O111, *E. coli* Flexneri, and *S. enterica* LT2 in the region between *gnd* and *cld* (6, 56), and probably has the same function. Although this region has not yet been sequenced in K-12, it is very likely that *ugd* will also be found there. It is interesting to note that *ugd*, while not within *wca*, is nevertheless in the same vicinity of the chromosome (about 13 kb downstream of *wca*).

The fourth nucleotide sugar necessary for E-unit production is GDP-L-fucose, which is derived from mannose-6-phosphate (Fig. 2). Genes for production of this sugar reside within the CA gene cluster, and we have previously reported on likely candidates for some of the steps (4). The roles of the genes which act first in the pathway, *manB* and *manC* (*orf14* and *orf15*, formerly called *cpsG* and *cpsB*, respectively [4, 70]), are quite well established: they encode, respectively, phosphomannomutase and mannose-1-phosphate guanosyltransferase, which convert D-mannose-6-phosphate into GDP-D-mannose. The next enzyme in the pathway is GMD, which produces

GDP-4-keto-6-deoxy-D-mannose. Previously (4) we suggested that the gene most likely to encode GMD is *orf10* (which we partly sequenced and called then *orf0.0*) because it is very closely related to genes involved in other biochemical pathways in which we would also expect to find GMD (those for synthesis of colitose and perosamine [7, 82]). To test this assignment we looked for GMD activity in a *wca*-deleted strain (SØ874) harboring pPR1689, a plasmid in which *orf10* was cloned. When a crude extract of protein from this strain was assayed for GMD activity, it was found to produce 0.14 μmol of product per hour per milligram of protein, whereas control strains (P4052 alone and P4052 carrying the plasmid vector only) showed negligible amounts of GMD activity. These results show that *orf10* directs the expression of GMD, and so this gene was assigned the name *gmd*.

The final steps in the GDP-L-fucose pathway are epimerization of GDP-4-keto-6-deoxy-D-mannose at C3 and C5 and reduction at C4. In porcine liver only a single enzyme, an "epimerase-reductase," is necessary for the final steps to GDP-L-fucose (10), but it is not known whether one or two genes will be required in the case of *E. coli*. Although we have not yet studied this part of the pathway experimentally, both *wcaG* and *wcaH* (*orf11* and *orf12*) are perhaps implicated in these steps because of their location between other fucose pathway genes (i.e., *gmd* and *manC*) and because similar genes (*orf6.7* and *orf3.4*, with 37 and 46% amino acid identity, respectively) are present in part of the *wba* (*rfb*) region of *E. coli* O111 believed to be responsible for the colitose pathway in which there are probably analogous reduction and epimerization steps (7). Indeed, the predicted WcaG protein is homologous to a number of other sugar pathway enzymes and has a NAD-binding sequence motif consistent with the fact that the C4 reduction step utilizes NADPH (22).

Glycosyl transferase genes. The E unit is synthesized at the cytoplasmic face of the inner membrane on a lipid carrier, probably undecaprenolpyrophosphate (Und-PP) (36). It is reasonable to expect that six of the *wca* genes encode specific glycosyl transferases which utilize nucleotidylated sugars to add the sugars in turn to the growing E unit. The first sugar added is glucose (36) which is attached by adding glucose-1-phosphate to Und-P. We propose that the initiating glucose-1-phosphate transferase is encoded by *wcaJ* because of its similarity to the initiating galactose-1-phosphate transferase WbaP (RfbP) used in O-unit synthesis in a group of related *S. enterica* serogroups (35, 69). The sequence similarity between WcaJ and WbaP is in the C-terminal region, which we have shown to encode the transferase activity (85), but interestingly the N-terminal portions show the same distribution of four hydrophobic segments which may traverse the inner membrane. WcaJ is similar over its entire length to ORF14 (62% identity) from the *Klebsiella pneumoniae* K2 *cps* cluster (5) and to GumD (36% identity) from the *Xanthomonas campestris gum* gene cluster responsible for production of the EPS xanthan (84). The three proteins may perform the same function since like CA the *K. pneumoniae* K2 capsule and xanthan gum both have glucose in their repeat units, and in the latter it has been shown that glucose-1-phosphate is the first residue to be transferred to a lipid carrier during xanthan synthesis (32).

WcaA, WcaC, WcaE, WcaL, and WcaI all show some sequence similarity to other glycosyl transferases (Fig. 4) and are therefore good candidates for the other five glycosyl transferases proposed to be necessary for E-unit synthesis. Of the five genes, only *wcaL* has a closely related gene in the database: it is very similar (52% amino acid identity) to an *Erwinia amylovora* gene (*amsK*) involved in the synthesis of the EPS amylovoran (8), indicating that the two genes share a similar function. The resemblance of the E-unit side chain to part of the amylovoran structure suggests that possible roles for WcaL are either as the GlcA transferase or as the final Gal transferase. The location of *wcaI* within the set of fucose biosynthetic genes suggests the possibility that it may encode a fucosyl transferase.

Acetyltransferase genes. *wcaB* and *wcaF* (Fig. 2) both encode proteins which are clearly related to a large family of proteins that act as acetyltransferases. WcaB is most closely related to CysE, the serine acetyltransferase involved in cysteine biosynthesis (37% identity over 103 amino acids). WcaF is most closely related to NodL, an acetyltransferase involved in the O acetylation of *Rhizobium* Nod factors (30% sequence identity over 124 amino acids). Presumably either WcaB or WcaF is responsible for addition of the acetyl group that is attached through an O linkage to the first fucosyl residue of the E unit (Fig. 1). However, the precise role of the second O-acetyltransferase gene is not known.

Polymerization and export genes. The polysaccharide biosynthetic gene clusters studied thus far can be divided into two distinct classes depending upon whether an ABC (ATP-binding cassette) transporter gene pair is present. ABC transporters (16, 18, 72) are implicated in the export of group II capsules in *E. coli* (65, 79) and related capsules of *Neisseria meningitidis* (19) and *Haemophilus influenzae* (46), of the *S. enterica* Vi antigen (29), and of a class of O antigens which includes *E. coli* O9 (40), *Yersinia enterocolitica* O3 (91), and *Vibrio cholerae* O1 (52). A different mechanism of polysaccharide export is thought to operate in a second class of O antigens, of which the paradigm is the *S. enterica* group B O antigen (69, 88). Gene clusters for producing this second class of O antigen invariably have a gene (*wzx*, formerly *rfbX* [70]) encoding a predicted integral membrane protein with 12 transmembrane segments. *wzx* genes from different O-antigen gene clusters usually show little primary sequence similarity to one another (51, 69). It is thought that Wzx (RfbX) acts in O-unit export by flipping the Und-PP-O-unit from the cytoplasm into the periplasm (48), where polymerization occurs. The CA gene cluster does not have a pair of ABC transporter genes, and we suggest that CA may belong to a class of EPSs in which repeat unit assembly, export across the cytoplasmic membrane, and perhaps polymerization proceed in a manner analogous to that proposed for the Wzx-dependent class of O antigens. *orf17* (Fig. 2) is predicted to encode an integral inner membrane protein with 10 transmembrane segments and shows some sequence similarity (particularly in the amino-terminal region) to proteins encoded at other EPS gene clusters, as well as to Wzx (RfbX) proteins of O-antigen systems (Fig. 5). Although the degree of similarity between these proteins is very low (for example, Orf17 shows 22% identity and 51% similarity to RfbX of *Shigella flexneri* [51]), pairwise comparisons in which one of the sequences was randomly shuffled showed that this level of similarity is nevertheless higher than what is expected by chance alone. We propose that *orf17* fulfills a role analogous to that of the *wzx* genes of O-antigen systems; i.e., it is necessary for export of the E unit from the cytoplasm, and so we have called it *wzx*. By further analogy with Wzx-dependent O antigen biosynthetic systems, we expect the CA cluster to have a gene responsible for polymerization of the E units. The O antigen polymerases thus far identified are integral inner membrane proteins (14, 55), and it is reasonable to expect that the CA polymerase will have a similar form. We note that *wcaD*, one of the remaining unassigned *wca* ORFs, encodes a protein predicted to have nine transmembrane segments, and so a possible role for WcaD is CA polymerization. There is no sequence similarity between WcaD and any of the O antigen

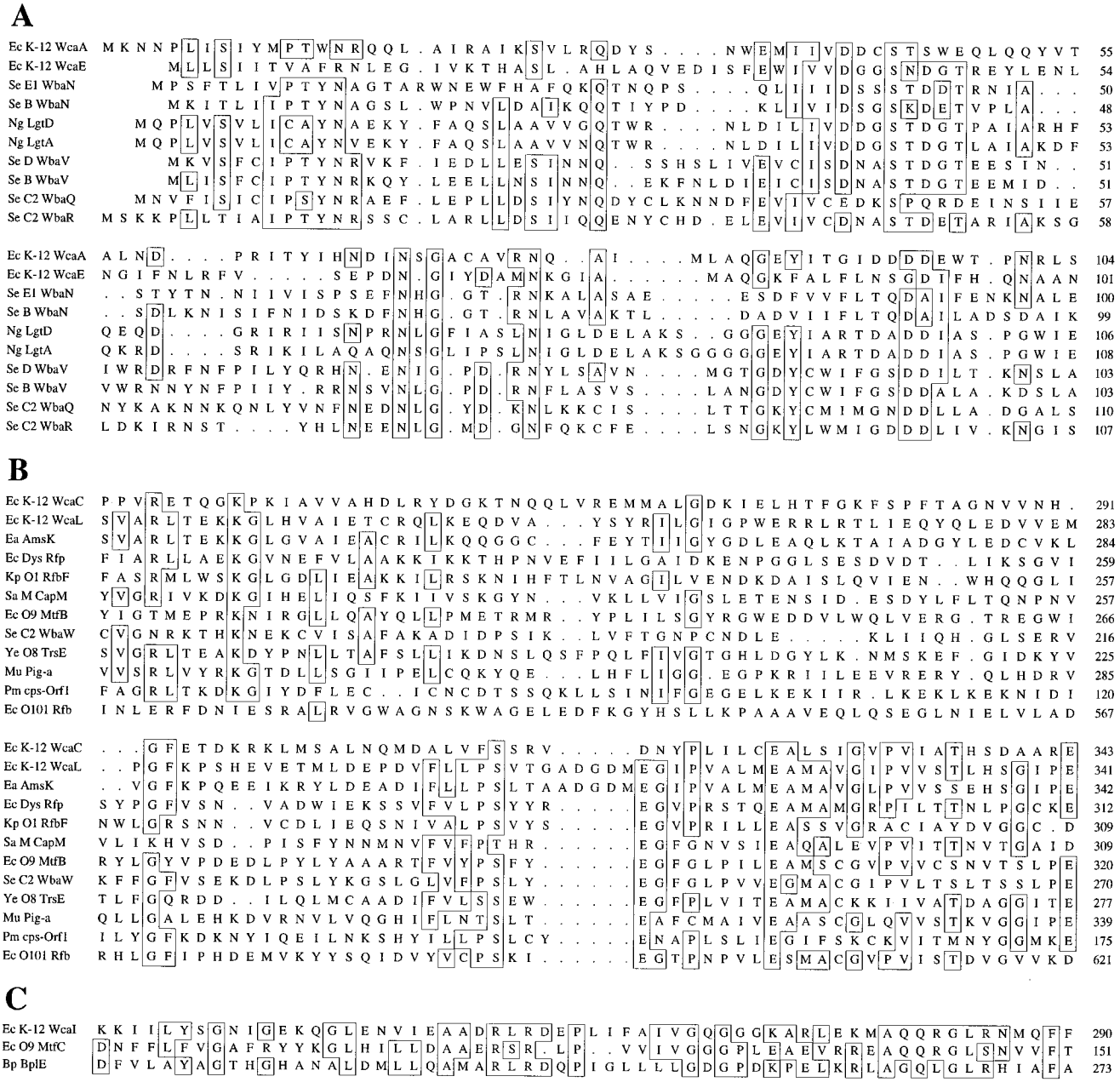

FIG. 4. Sequence alignments of sections of four K-12 Wca proteins with various glycosyl transferases. Gaps (.) have been introduced to improve the alignments. Positions where at least 50% of the sequences in the set have identical amino acids have been boxed. Ea, *E. amylovora*; Ec, *E. coli*; Dy, dysenteriae serogroup of *E. coli*; Kp, *K. pneumoniae*; Mu, mouse; Ng, *N. gonorrhoeae*; Pm, *Proteus mirabilis*; Sa, *Staphylococcus aureus*; Se, *S. enterica*; Ye, *Y. enterocolitica*. Numbers show the position within each protein sequence of the last amino acid on each line. (A) Amino-terminal sections of WcaA and WcaE. WbaN (RfbN) and WbaQ (RfbQ), rhamnosyl transferase (49); WbaV (RfbV), abequosyl transferase (50); LgtD, GalNac transferase; LgtA, GlcNac transferase (25). This region includes the conserved sequence previously observed by Liu et al. (50) in various glycosyltransferases. (B) Segments within the carboxy-terminal halves of WcaC and WcaL. Rfp, galactosyl transferase (42); RfbF, galactosyl transferase (13); MtfB, mannosyl transferase (40); WbaW (RfbW), mannosyl transferase (49); Pig-a, probable glucosaminyl transferase (37). The functions of the other proteins are not known but all are involved in polysaccharide production (8, 12, 47, 78); (see also GenBank entry PRMCPSFA) and so are probably also glycosyltransferases. (C) Segment within WcaI. MtfC, mannosyl transferase (40); BplE, possible glycosyl transferase from *Bordetella pertussis* lipopolysaccharide gene cluster (1).

polymerases which have been studied, or with any other protein in the sequence databases, but O antigen polymerases in general have little or no sequence similarity. The assignment of these functions to Wzx and WcaD is speculative and must be tested experimentally.

The first three genes of the CA cluster (*wza*, *wzb*, and *wzc*) have homologs in several other EPS gene clusters: *ams* of *E. amylovora* (8), *exo* of *Rhizobium meliloti* (73), *eps* of *Pseudomonas* spp. (31), and *cps* of *Klebsiella pneumoniae* (5). This suggests that their functions may not be specific to CA, but instead common to these sorts of EPSs. Since these genes are not present in O-antigen gene clusters, it is possible that they comprise an EPS processing system which acts in the later stages of EPS export.

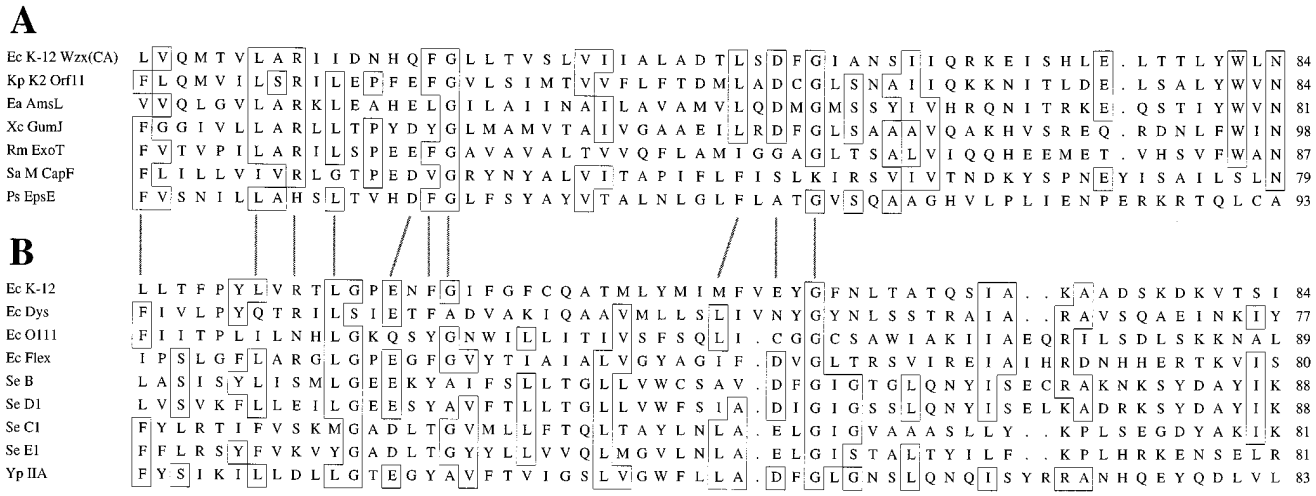

FIG. 5. Sequence alignments of Wzx (RfbX-like) proteins near their amino-terminal ends. Positions where at least four of the sequences in parts A or B have identical amino acids have been boxed. Lines show conserved residues common to parts A and B. (A) K-12 Wzx (colanic acid) and related proteins involved in production of other EPSs. Ps, *P. solanacearum* (31); Xc, *Xanthomonas campestris* (84) (see also GenBank entry U22511); other abbreviations and references are as in Fig. 4. (B) K-12 Wzx (O antigen) and related proteins encoded at *wba* (*rfb*) gene clusters (7, 39, 43, 51, 69, 81). Flex, flexneri serogroup of *E. coli*; other abbreviations and references are as in Fig. 4.

Wza is proposed to be an outer membrane lipoprotein and is homologous to *E. amylovora* AmsH, *K. pneumoniae* K2 *cps* ORF4, *P. solanacearum* EpsA, and *R. meliloti* ExoF. All of these proteins are predicted to have a signal peptidase II recognition site (68). This peptidase is specific for the processing of lipoproteins, cleaving the polypeptide upstream of a cysteine residue to which a glyceride-fatty acid lipid is subsequently attached. The residue at position 2 of the Wza predicted mature peptide, which is important in lipoprotein sorting, is Val rather than Asp, which suggests the protein is more likely to be directed to the outer membrane than the inner membrane. Consistent with this, Wza is similar to CtrA (29% amino acid identity over 379 amino acids), an outer membrane protein involved in the export of polysaccharide capsules of *N. meningitidis* (20).

Wzb is homologous to *E. amylovora* AmsI, *K. pneumoniae* K2 *cps* ORF5, and *Pseudomonas solanacearum* EpsP, and all these proteins are homologous to acid phosphatases.

Wzc is similar to *E. amylovora* AmsA, *K. pneumoniae* K2 *cps* ORF6, *P. solanacearum* EpsB, and *R. meliloti* ExoP. These proteins all have an ATP-binding sequence motif (but are otherwise unrelated to the ATP-binding components of the ABC transporter systems mentioned above), and also have three predicted transmembrane segments.

Other genes. Searches through sequence databases gave no indication of the functions of the remaining *wca* gene, *wcaK* (*orf18*). *wcaK* is closely related to the *E. amylovora* EPS gene *amsJ*, whose function is not known. The only function in the proposed CA biosynthetic pathway that has not yet been at least tentatively assigned is addition of the pyruvyl group to the terminal galactosyl residue of the E-unit side chain (Fig. 1); however, *wcaK* is not homologous to a previously sequenced pyruvylation gene of *R. meliloti* (24) and so there is no reason to suggest this assignment.

Evolution of *wca*. Figure 2 shows the GC content of the CA cluster. Most of the CA genes have a high GC content (52 to 57%) compared with that usual for *E. coli* genes (50%), indicating that they were acquired by lateral gene transfer from another species. Interestingly though, a block of three genes (*wcaD*, *wcaE*, and *wcaF*; Fig. 2), as well as *orf1.3*, have a lower GC content. This suggests that these genes have a different evolutionary history to the other genes of the cluster. In a previous comparison between *E. coli* K-12 and *S. enterica* LT2 of some of the *wca* genes, we pointed out an anomaly in the adjustment of P3 values (the corrected GC content of codon base 3), which we interpreted as meaning that *wca* was introduced into the two species at different times (4). It will be interesting to complete the sequence of the *S. enterica* LT2 *wca* cluster to extend this comparison.

ACKNOWLEDGMENTS

We thank Valerie Stout for making results available to us prior to publication and David Coplin and David Gutnick for helpful discussion.

The work was supported by a grant from the Australian Research Council.

REFERENCES

1. **Allen, A. G., and D. J. Maskell.** 1996. The identification, cloning and mutagenesis of a genetic locus required for lipopolysaccharide biosynthesis in *Bordetella pertussis*. Mol. Microbiol. **19:**37–52.
2. **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.
3. **Anderson, E. S., and A. H. Rogers.** 1963. Slime polysaccharide of the *Enterobacteriaceae*. Nature (London) **198:**714–715.
4. **Aoyama, K. M., A. M. Haase, and P. R. Reeves.** 1994. Evidence for effect of random genetic drift on G+C content after lateral transfer of fucose pathway genes to *Escherichia coli* K-12. Mol. Biol. Evol. **11:**829–838.
5. **Arakawa, Y., R. Wacharotayankun, T. Nagatsuka, H. Ito, N. Kato, and M. Ohta.** 1995. Genomic organization of the *Klebsiella pneumoniae cps* region responsible for serotype K2 capsular polysaccharide synthesis in the virulent strain chedid. J. Bacteriol. **177:**1788–1796.
6. **Bastin, D. A., P. K. Brown, A. Haase, G. Stevenson, and P. R. Reeves.** 1993. Repeat unit polysaccharides of bacteria: a model for polymerisation resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length. Mol. Microbiol. **7:**725–734.
7. **Bastin, D. A., and P. R. Reeves.** 1995. Sequence and analysis of the O antigen gene (*rfb*) cluster of *Escherichia coli* O111. Gene **164:**17–23.
8. **Bugert, P., and K. Geider.** 1995. Molecular analysis of the *ams* operon required for exopolysaccharide synthesis of *Erwinia amylovora*. Mol. Microbiol. **15:**917–933.
9. **Chang, A. C. Y., and S. N. Cohen.** 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. J. Bacteriol. **134:**1141–1156.
10. **Chang, S., B. Duerr, and G. Serif.** 1988. An epimerase-reductase in L-fucose

synthesis. J. Biol. Chem. **263:**1693–1697.
11. **Chaudhury, A. M., and G. R. Smith.** 1984. A new class of *Escherichia coli recBC* mutants: implications for the role of RecBC enzyme in homologous recombination. Proc. Natl. Acad. Sci. USA **81:**7850–7854.
12. **Cheah, K. C., and P. A. Manning.** 1993. Inactivation of the *Escherichia coli* B41 (O101:K99/F41) *rfb* gene encoding an 80-kDa polypeptide results in the synthesis of an antigenically altered lipopolysaccharide in *E. coli* K-12. Gene 9–15.
13. **Clarke, B. R., D. Bronner, W. J. Keenleyside, W. B. Severn, J. C. Richards, and C. Whitfield.** 1995. Role of *rfe* and *rfbF* in the initiation of biosynthesis of D-galactan I, the lipopolysaccharide O antigen from *Klebsiella pneumoniae* serotype O1. J. Bacteriol. **177:**5411–5418.
14. **Collins, L. V., and J. Hackett.** 1991. Molecular cloning, characterization, and nucleotide sequence of the *rfc* gene, which encodes an O antigen polymerase of *Salmonella typhimurium*. J. Bacteriol. **173:**2521–2529.
15. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12:**387–395.
16. **Doige, C. A., and G. F.-L. Ames.** 1993. ATP-dependent transport systems in bacteria and humans: relevance to cystic fibrosis and multidrug resistance. Annu. Rev. Microbiol. **47:**291–319.
17. **Dougherty, B. A., and I. van de Rijn.** 1993. Molecular characterization of *has*β from an operon required for hyaluronic acid synthesis in group A streptococci. J. Biol. Chem. **268:**7118–7124.
18. **Fath, M. J., and R. Kolter.** 1993. ABC transporters: bacterial exporters. Microbiol. Rev. **57:**995–1017.
19. **Frosch, M., U. Edwards, K. Bousset, B. Krausse, and C. Weisgerber.** 1991. Evidence for a common molecular origin of the capsule gene loci in gram-negative bacteria expressing group II capsular polysaccharides. Mol. Microbiol. **5:**1251–1263.
20. **Frosch, M., D. Müller, K. Bousset, and A. Müller.** 1992. Conserved outer membrane protein of *Neisseria meningitidis* involved in capsule expression. Infect. Immun. **60:**798–803.
21. **Garegg, P. J., B. Lindberg, T. Onn, and I. W. Sutherland.** 1971. Comparative structural studies on the M-antigen from *Salmonella typhimurium*, *Escherichia coli*, and *Aerobacter cloacae*. Acta Chem. Scand. **25:**2103–2108.
22. **Ginsburg, V.** 1966. Formation of GDP-L-fucose from GDP-D-mannose. Methods Enzymol. **8:**293–295.
23. **Gleeson, T. J., and R. Staden.** 1991. An X windows and UNIX implementation of our sequence analysis package. Comput. Appl. Biosci. **7:**398.
24. **Glucksmann, M. A., T. L. Reuber, and G. C. Walker.** 1993. Genes needed for the modification, polymerization, export, and processing of succinoglycan by *Rhizobium meliloti*: a model for succinoglycan biosynthesis. J. Bacteriol. **175:**7045–7055.
25. **Gotschlich, E. C.** 1994. Genetic locus for the biosynthesis of the variable portion of *Neisseria gonorrhoeae* lipooligosaccharide. J. Exp. Med. **180:**2181–2190.
26. **Gottesman, S.** 1995. Regulation of capsule synthesis: modification of the two-component paradigm by an accessory unstable regulator, p. 253–262. *In* J. A. Hoch and T. J. Silhavy (ed.), Two-component signal transduction. American Society for Microbiology, Washington, D.C.
27. **Gottesman, S., and V. Stout.** 1991. Regulation of capsular polysaccharide synthesis in *Escherichia coli* K12. Mol. Microbiol. **5:**1599–1606.
28. **Grant, W. D., I. W. Sutherland, and J. F. Wilkinson.** 1969. Exopolysaccharide colanic acid and its occurrence in the *Enterobacteriaceae*. J. Bacteriol. **100:**1187–1193.
29. **Hashimoto, Y., N. Li, H. Yokoyama, and T. Ezaki.** 1993. Complete nucleotide sequence and molecular characterization of ViaB region encoding Vi antigen in *Salmonella typhi*. J. Bacteriol. **175:**4456–4465.
30. **Hobbs, M., and P. R. Reeves.** 1994. The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters. Mol. Microbiol. **12:**855–856.
31. **Huang, J., and M. Schell.** 1995. Molecular characterization of the *eps* gene cluster of *Pseudomonas solanacearum* and its transcriptional regulation at a single promoter. Mol. Microbiol. **16:**977–989.
32. **Ielpi, L., R. O. Couso, and M. A. Dankert.** 1993. Sequential assembly and polymerization of the polyprenol-linked pentasaccharide repeat unit of the xanthan polysaccharide in *Xanthomonas campestris*. J. Bacteriol. **175:**2490–2500.
33. **Jann, B., and K. Jann.** 1990. Structure and biosynthesis of the capsular antigens of *Escherichia coli*. Curr. Top. Microbiol. **150:**19–42.
34. **Jann, K., and B. Jann.** 1992. Capsules of *Escherichia coli*, expression and biological significance. Can. J. Microbiol. **38:**705–710.
35. **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.
36. **Johnson, J. G., and D. B. Wilson.** 1977. Role of a sugar-lipid intermediate in colanic acid synthesis by *Escherichia coli*. J. Bacteriol. **129:**225–236.
37. **Kawagoe, K., J. Takeda, Y. Endo, and T. Kinoshita.** 1994. Molecular cloning of murine pig-a, a gene for GPI-anchor biosynthesis, and demonstration of interspecies conservation of its structure, function, and genetic locus. Genomics **23:**566–574.
38. **Keenleyside, W. J., P. Jayaratne, P. R. MacLachlan, and C. Whitfield.** 1992. The *rcsA* gene of *Escherichia coli* O9:K30:H12 is involved in the expression of the serotype-specific group I K (capsular) antigen. J. Bacteriol. **174:**8–16.
39. **Kessler, A., A. Haase, and P. R. Reeves.** 1993. Molecular analysis of the 3,6-dideoxyhexose pathway genes of *Yersinia pseudotuberculosis* serogroup IIa. J. Bacteriol. **175:**1412–1422.
40. **Kido, N., V. I. Torgov, T. Sugiyama, K. Uchiya, H. Sugihara, T. Komatsu, N. Kato, and K. Jann.** 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.
41. **Klein, P., M. Kanehisa, and C. Delisi.** 1985. The detection and classification of membrane spanning proteins. Biochem. Biophys. Acta **815:**468–476.
42. **Klena, J. D., R. S. Asford II, and C. A. Schnaitman.** 1992. Role of *Escherichia coli* K12 *rfa* genes and the *rfp* gene of *Shigella dysenteriae* 1 in generation of lipopolysaccharide core heterogeneity and attachment of O antigen. J. Bacteriol. **174:**7297–7307.
43. **Klena, J. D., and C. A. Schnaitman.** 1993. Function of the *rfb* gene cluster and the *rfe* gene in the synthesis of O antigen by *Shigella dysenteriae* 1. Mol. Microbiol. **9:**393–402.
44. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50:**495–508.
45. **Kornfeld, R. H., and V. Ginsburg.** 1966. Control of synthesis of guanosine 5′-diphosphate D-mannose and guanosine 5′-diphosphate L-fucose in bacteria. Biochim. Biophys. Acta **117:**79–87.
46. **Kroll, J. S., I. Hopkins, and E. R. Moxon.** 1988. Capsule loss in *H. influenzae* type b occurs by recombination-mediated disruption of a gene essential for polysaccharide export. Cell **53:**347–356.
47. **Lin, W., T. Cunneen, and C. Lee.** 1994. Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*. J. Bacteriol. **176:**7005–7016.
48. **Liu, D., R. Cole, and P. R. Reeves.** 1996. An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. J. Bacteriol. **178:**2102–2107.
49. **Liu, D., A. M. Haase, L. Lindqvist, A. A. Lindberg, and P. R. Reeves.** 1993. Glycosyl transferases of O-antigen biosynthesis in *Salmonella enterica*: identification and characterization of transferase genes of groups B, C2, and E1. J. Bacteriol. **175:**3408–3413.
50. **Liu, D., L. Lindquist, and P. R. Reeves.** 1995. Transferases of O-antigen biosynthesis in *Salmonella enterica*: dideoxhexosyl transferases of groups B and C2 and acetyltransferase of group C2. J. Bacteriol. **177:**4084–4088.
51. **Macpherson, D. F., P. A. Manning, and R. Morona.** 1995. Genetic analysis of the *rfbX* gene of *Shigella flexneri*. Gene **155:**9–17.
52. **Manning, P. A., U. H. Stroeher, L. E. Karageorgos, and R. Morona.** 1995. Putative O-antigen transport genes within the *rfb* region of *Vibrio cholerae* O1 are homologous to those for capsule transport. Gene **158:**1–7.
53. **Markovitz, A.** 1977. Genetics and regulation of bacterial capsular polysaccharide synthesis and radiation sensitivity, p. 415–462. *In* I. W. Sutherland (ed.), Surface carbohydrates of the prokaryotic cell, vol. I. Academic Press, New York.
54. **Misra, R., and Y. Miao.** 1995. Molecular analysis of *asmA*, a locus identified as the suppressor of OmpF assembly mutants of *Escherichia coli* K-12. Mol. Microbiol. **16:**779–788.
55. **Morona, R., M. Mavris, A. Fallarino, and P. A. Manning.** 1994. Characterization of the *rfc* region of *Shigella flexneri*. J. Bacteriol. **176:**733–747.
56. **Morona, R., L. Van Den Bosch, and P. Manning.** 1995. Molecular, genetic, and topological characterization of O antigen chain regulation in *Shigella flexneri*. J. Bacteriol. **177:**1059–1068.
57. **Nakae, T.** 1971. Multiple molecular forms of uridine diphosphate glucose pyrophosphorylase from Salmonella typhimurium: interconversion between various forms. J. Biol. Chem. **246:**4404–4411.
58. **Nakae, T., and H. Nikaido.** 1971. Multiple molecular forms of uridine diphosphate glucose pyrophosphorylase from Salmonella typhimurium: genetic determination of multiple forms. J. Biol. Chem. **246:**4397–4403.
59. **Nakae, T., and H. Nikaido.** 1971. Multiple molecular forms of uridine diphosphate glucose pyrophosphorylase from Salmonella typhimurium: catalytic properties of various forms. J. Biol. Chem. **246:**4386–4396.
60. **Neuhard, J., and E. Thomassen.** 1976. Altered deoxyribonucleotide pools in P2 eductants of *Escherichia coli* K-12 due to deletion of the *dcd* gene. J. Bacteriol. **126:**999–1001.
61. **Nikaido, H., M. Levinthal, K. Nikaido, and K. Nakane.** 1967. Extended deletions in the histidine-rough-B region of the *Salmonella* chromosome. Proc. Natl. Acad. Sci. USA **57:**1825–1832.
62. **Ophir, T., and D. L. Gutnick.** 1994. A role for exopolysaccharides in the protection of microorganisms from desiccation. Appl. Environ. Microbiol. **60:**740–745.
63. **Ørskov, F., and I. Ørskov.** 1992. *Escherichia coli* serotyping and disease in man and animals. Can. J. Microbiol. **38:**699–704.
64. **Ørskov, I., F. Ørskov, B. Jann, and K. Jann.** 1977. Serology, chemistry, and genetics of O and K antigens of *Escherichia coli*. Bacteriol. Rev. **41:**667–710.
65. **Pavelka, M. S., Jr., L. F. Wright, and R. P. Silver.** 1991. Identification of two genes, *kpsM* and *kpsT*, in region 3 of the polysialic gene cluster of *Escherichia*

coli K1. J. Bacteriol. **173:**4603–4610.
66. **Pearce, R., and I. S. Roberts.** 1995. Cloning and analysis of gene clusters for the production of the *Escherichia coli* K10 and K54 antigens: identification of a new group of *serA*-linked capsules gene clusters. J. Bacteriol. **177:**3992–3997.
67. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85:**2444–2448.
68. **Pugsley, A. P.** 1993. The complete general secretory pathway in gram-negative bacteria. Microbiol. Rev. **57:**50–108.
69. **Reeves, P. R.** 1993. Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. Trends Genet. **9:**17–22.
70. **Reeves, P. R., M. Hobbs, M. Valvano, M. Skurnik, C. Whitfield, D. Coplin, N. Kido, J. Klena, D. Maskell, C. Raetz, and P. Rick.** Proposal for a new nomenclature for bacterial surface polysaccharide genes. Submitted for publication.
71. **Reisner, A. H., C. A. Bucholtz, J. Smelt, and S. McNeil.** 1993. Australia's National Genomic Information Service. Proc. Twenty-Sixth Annu. Hawaii Int. Conf. Systems Sci. **1:**595–602.
72. **Reizer, J., A. Reizer, and M. H. Saier, Jr.** 1992. A new subfamily of bacterial ABC-type transport systems catalysing export of drugs and carbohydrates. Protein Sci. **1:**1326–1332.
73. **Reuber, T. L., and G. C. Walker.** 1993. Biosynthesis of succinoglycan, a symbiotically important exopolysaccharide of *Rhizobium meliloti*. Cell **74:**269–280.
74. **Rudd, K. E.** 1992. Alignment of *E. coli* DNA sequences to a revised, integrated genomic restriction map, p. 2.3–2.43. *In* J. Miller (ed.), A short course in bacterial genetics: a laboratory manual and handbook for *Escherichia coli* and related bacteria. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
75. **Russo, T. A., and G. Singh.** 1993. An extraintestinal, pathogenic isolate of *Escherichia coli* (O4/K54/H5) can produce a group 1 capsule which is divergently regulated from its constitutively produced group 2, K54 capsular polysaccharide. J. Bacteriol. **175:**7617–7623.
76. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
77. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
78. **Skurnik, M., R. Venho, P. Toivanen, and A. Alhendy.** 1995. A novel locus of *Yersinia enterocolitica* serotype O-3 involved in lipopolysaccharide outer core biosynthesis. Mol. Microbiol. **17:**575–594.
79. **Smith, A. N., G. J. Boulnois, and I. S. Roberts.** 1990. Molecular analysis of the *Escherichia coli* K5 *kps* locus: identification and characterization of an inner-membrane capsular polysaccharide transport system. Mol. Microbiol. **4:**1863–1869.
80. **Staden, R.** 1984. Graphic methods to determine the function of nucleic acid sequences. A summary of ANALYSEQ options. Nucleic Acids Res. **12:**521–538.
81. **Stevenson, G., B. Neal, D. Liu, M. Hobbs, N. H. Packer, M. Batley, J. W. Redmond, L. Lindquist, and P. R. Reeves.** 1994. Structure of the O antigen of *E. coli* K-12 and the sequence of its *rfb* gene cluster. J. Bacteriol. **176:**4144–4156.
81a. **Stout, V.** 1996. Identification of the promoter region for the colanic acid polysaccharide biosynthetic genes in *Escherichia coli* K-12. J. Bacteriol. **178:**4273–4280.
82. **Stroeher, U. H., L. E. Karageorgos, R. Morona, and P. A. Manning.** 1992. Serotype conversion in *Vibrio cholerae* O1. Proc. Natl. Acad. Sci. USA **89:**2566–2570.
83. **Trisler, P., and S. Gottesman.** 1984. *lon* transcriptional regulation of genes necessary for capsular polysaccharide synthesis in *Escherichia coli* K-12. J. Bacteriol. **160:**184–191.
83a. **Valvano, M.** Personal communication.
84. **Vanderslice, R. W., D. H. Doherty, M. A. Capage, M. R. Betlach, R. A. Hassler, N. M. Henderson, J. Ryan-Graniero, and M. Tecklenberg.** 1989. Genetic engineering of polysaccharide structure in *Xanthomonas campestris*, p. 145–156. *In* V. Crescenzi, I. C. M. Dea, S. Paoletti, S. S. Stivala, and I. Sutherland (ed.), Biomedical and biotechnological advances in industrial polysaccharides, Godon and Breach Science Publishers, New York.
85. **Wang, L., D. Liu, and P. R. Reeves.** 1996. C-terminal half of *Salmonella enterica* WbaP (RfbP) is the galactosyl-1-phosphate transferase domain catalysing the first step of O antigen synthesis. J. Bacteriol. **178:**2598–2604.
86. **Weissborn, A. C., Q. Y. Liu, M. K. Rumley, and E. P. Kennedy.** 1994. UTP-alpha-D-glucose-1-phosphate uridylyltransferase of *Escherichia coli*: isolation and DNA sequence of the *galU* gene and purification of the enzyme. J. Bacteriol. **176:**2611–2618.
87. **Weissborn, A. C., M. K. Rumley, and E. P. Kennedy.** 1992. Isolation and characterization of *Escherichia coli* mutants blocked in production of membrane-derived oligosaccharides. J. Bacteriol. **174:**4856–4859.
88. **Whitfield, C.** 1995. Biosynthesis of lipopolysaccharide O antigens. Trends Microbiol. **3:**178–185.
89. **Whitfield, C., W. J. Keenleyside, and B. R. Clarke.** 1994. Structure, function and synthesis of surface polysaccharides in *Escherichia coli*, p. 437–494. *In* C. L. Gyles (ed.), *Escherichia coli* in domestic animals and humans. CAB International, Wallingford, U.K.
90. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33:**103–119.
91. **Zhang, L., A. Al-Hendy, P. Toivanen, and M. Skurnik.** 1993. Genetic organization and sequence of the *rfb* gene cluster of *Yersinia enterocolitica* serotype O:3: similarities to the dTDP-L-rhamnose biosynthesis pathway of *Salmonella* and to the bacterial polysaccharide transport systems. Mol. Microbiol. **9:**309–321.