Application of a new nomenclature for bacterial surface polysaccharide genes.

Peter R. Reeves, Matthew Hobbs, Miguel A. Valvano, Mikael Skurnik, Chris Whitfield, David Coplin, Nobuo Kido, John Kleen, Duncan Maskell, Christian R. H. Raetz and Paul D. Rick

Affiliations and addresses of Authors at end of paper.

This communication expands on a proposal put forward in Trends in Microbiology by the same authors (Reeves et al. (1996) to meet two perceived needs: firstly, the need to resolve the problem created by the use of all 26 available rfb gene symbols for O antigen genes in some species with the same problem looming for cps; and secondly, the need to maintain a similar nomenclature for polysaccharide genes within a species, and where possible between species. The philosophy was to develop a naming system applicable to all current and foreseen surface polysaccharide genes. The proposed scheme (the Bacterial Polysaccharide Gene Nomenclature scheme (BPGN)) avoids re-use of names such as rfbA etc for genes of quite different function and makes it much easier to compare pathway and assembly genes from different species, which is very difficult at present because of the plethora of gene names currently used for some functions. This paper gives details of the 30 gene clusters to which it has been applied as yet and gives information on procedures for renaming other gene clusters using the BPGN system.

This paper is designed to be read in conjunction with a paper published in Trends in Microbiology (Reeves et al., 1996), which is also available from the same web site as this paper (1). An introduction with general references on bacterial polysaccharides was given in that paper. We take this opportunity to thank Elsevier for allowing us to make the Trends in Microbiology paper available on the web. This additional paper is not published other than on the world wide web but is nonetheless treated by us as a typical paper in the sense that has a date of release and will not be updated, other than for typographical errors. For updated information look at the Bacterial polysaccharide Gene Database (BPGD) and associated information.

The Bacterial polysaccharide Gene Nomenclature scheme (BPGN) system has been designed to overcome serious problems due to congestion - too many genes for the number of names. The case for a new scheme is developed in the Trends paper. In this paper we briefly describe the new nomenclature without repeating the rationale and then give some detail of the gene clusters to which the nomenclature has been applied in the first instance.

The Bacterial Polysaccharide Gene Nomenclature (BPGN) scheme in outline

The scheme is designed primarily for genes involved in bacterial polysaccharide synthesis, although there will be overlap with genes involved in other saccharide transformations including catabolism.

Three classes of genes are distinguished:

Firstly those involved in synthesis of sugar intermediates such as dTDP-rhamnose or GDP-mannose. For these genes the same principles are followed as for other biosynthetic pathways and for example the dTDP-L-rhamnose pathway genes are named rmlA, rmlB, rmlC and rmlD. This may not seem revolutionary to those outside the field but for historical reasons these genes were previously named after the type of polysaccharide, be it capsule, lipopolysaccharide (LPS) etc. and in some but not all cases had different names in different gene clusters. A diagram of pathways described to date is shown in the Trends paper.

The second class is for genes for steps common to synthesis of many polysaccharides, although they may display cluster specificity. The best known codes for the polymerase for O antigens. Typically the O antigen subunit, an oligosaccharide usually comprising 3 to 6 sugars known as the O-unit, is synthesised on a lipid carrier, undecaprenol pyrophosphate. It is then polymerised to a polysaccharide chain with perhaps 10-15 repeats of the O-unit. This polymerase was first described for Salmonella enterica LT2 and named rfc. Similar genes have been described for several O antigens and also named rfc. We are renaming them wzy for reasons given below. An important point is that each is specific to the particular O-antigen (or sometimes a group of related O antigens). There are several other genes in this class named wza, wzb, wzc for genes though to be involved in export of one group of polysaccharide, wxx and wzz, which with wzy are involved in processing classical O antigens, and wzm and wzt involved in export of Type II E. coli capsules and some O antigens. In each case genes of a given name perform the same general function and are homologous, but may have specificity such that one cannot complement another. This is in contrast to the pathway genes where genes of the same name have the same specific function and can complement each other. These genes are not in general well understood and are usually identified by similarity sequence and/or predicted secondary structure. There will probably be
more genes of this type recognised as we learn more of polysaccharide synthesis.

The third class is for those which do not fit into any other class. They are mostly genes for transferases which assemble the polysaccharide or its repeat unit.

The colanic acid gene cluster of *Erwinia* was first given to the locus for synthesis of the capsular polysaccharide, which is almost identical in sequence but different names (Lee et al., 1992). *gmd*, the GDP-mannose 4.6-dehydratase gene in the *E*. *coli* K-12 sequence. Putative processing genes were named *wza*, *wzb*, *wzc* and *wzx*. The other 13 genes present were named *wcaA-M*.

The gene cluster of *Erwinia stewartii* has been sequenced (Coplin et al., 1995) and it, and to a lesser extent the K-12 *cps* cluster show homology to that of *Erwinia amylovora* (Bugert & Geider, 1995).

Note that while the 'b' and 'c' in 'wbb' and 'wca' represent the historical description of the respective gene clusters as 'rfb' and 'cps' for O antigen and capsule clusters; the gene designation should not change should our description of the saccharide change, for example should the 'capsule' of *E. coli* K-12 be re-evaluated as an exopolysaccharide as is quite possible. Likewise should, for example, a gene first described in an O antigen cluster be found in a capsule cluster it will be given the same name in both.

### Enterobacterial common antigen genes

The genes of the Enterobacterial common antigen (ECA) cluster are present in both *E. coli* and *S. enterica* serovar Typhimurium (Daniels et al., 1992; Meier-Dieter et al., 1992), and they are presumably present in other members of the Enterobacteriaceae. The designation *wec* has been used for the genes unique to this cluster. ECA contains N-acetyl-glucosamine (GlcNAc), N-acetyl-mannosaminuronic acid (ManNAcA), and N-acetyl-fucosamine (Fuc4NAc). The donor of Fuc4NAc is in the GDP-L-fucose biosynthetic pathway (Fig. 1 of Trends paper), and the genes *rfpA* have been identified (Barr & Rick, 1993; Marolda & Valvano, 1995; Meier-Dieter et al., 1990). Although o389 and o379 have been tentatively identified as the *mnaA* and *mnaB* genes, respectively of the ManNAcA pathway, the functions of these open reading frames remains to be unequivocally established. Similarly, the *wzx* and *wzz* genes have been identified by similarity to the genes of O-antigen clusters.

### Capsule genes of *Salmonella* and *E. coli*

The name *cps* was first given to the locus for synthesis of colanic acid, found in many Enterobacteriaceae and studied in *E. coli* K-12 and *Salmonella* LT2. This colanic acid gene cluster is linked to the O antigen cluster. The gene clusters for production of group I capsules of *E. coli* appear to be allelic to the colanic acid *cps* cluster (Keenleyside et al., 1992), as do those of *Erwinia*, *Klebsiella pneumoniae* and probably other genera. As noted in the Trends paper, some have been called *cps* clusters but others given different names.

The colanic acid gene cluster of *E. coli* K-12, has been sequenced and the genes named in accordance with the current proposal (Stevenson et al., 1996). Part of both the K-12 and *Salmonella enterica* LT2 clusters had been sequenced and given *cps* names. They were in the GDP-L-fucose biosynthetic pathway and were renamed *manB* and *manC* respectively, removing the anomalous situation for *cpsG* and *rfbK* of group C1, which are almost identical in sequence but had different names (Lee et al., 1992). *gmd*, the GDP-mannose 4.6-dehydratase gene in the *E*. *coli* K-12 sequence. Putative processing genes were named *wza*, *wzb*, *wzc* and *wzx*. The other 13 genes present were named *wcaA-M*.

The gene cluster of *Erwinia stewartii* has been sequenced (Coplin et al., 1995) and it, and to a lesser extent the K-12 *cps* cluster show homology to that of *Erwinia amylovora* (Bugert & Geider, 1995).

Note that while the 'b' and 'c' in 'wbb' and 'wca' represent the historical description of the respective gene clusters as 'rfb' and 'cps' for O antigen and capsule clusters; the gene designation should not change should our description of the saccharide change, for example should the 'capsule' of *E. coli* K-12 be re-evaluated as an exopolysaccharide as is quite possible. Likewise should, for example, a gene first described in an O antigen cluster be found in a capsule cluster it will be given the same name in both.

### Enterobacterial common antigen genes

The genes of the Enterobacterial common antigen (ECA) cluster are present in both *E. coli* and *S. enterica* serovar Typhimurium (Daniels et al., 1992; Meier-Dieter et al., 1992), and they are presumably present in other members of the Enterobacteriaceae. The designation *wec* has been used for the genes unique to this cluster. ECA contains N-acetyl-glucosamine (GlcNAc), N-acetyl-mannosaminuronic acid (ManNAcA), and N-acetyl-fucosamine (Fuc4NAc). The donor of Fuc4NAc is in the GDP-L-fucose biosynthetic pathway (Fig. 1 of Trends paper), and the genes *rfpA* have been identified (Barr & Rick, 1993; Marolda & Valvano, 1995; Meier-Dieter et al., 1990). Although o389 and o379 have been tentatively identified as the *mnaA* and *mnaB* genes, respectively of the ManNAcA pathway, the functions of these open reading frames remains to be unequivocally established. Similarly, the *wzx* and *wzz* genes have been identified by similarity to the genes of O-antigen clusters.

### LPS lipid A and core genes

wa* has been be reserved for lipid A and LPS core genes in which the a in the second position indicates that it is from the lipopolysaccharide core "rfb" group. Thus for example *rfAG*, encoding a glucosyl transferase becomes waAG, the G representing the Gin its current name. *kdaA*, the KDO transferase gene, is in the same gene cluster and has been renamed *waaA* as it is part of
the same pathway, and htrB and msbB, also in the same pathway but mapping elsewhere, have been renamed waaM and waaN. Pathway genes and assembly genes have appropriate names as discussed above. The ADP-glycero-mannoheptose pathway genes have been allocated the names gmhA-D (Fig 1) of which gmhA and gmhB (were ipcA and ipcB respectively) map away from the main LPS cluster but rfaD becomes gmhD, while gmhC is not yet identified, although waaE (was rfaE) has been considered a candidate. For details of the genes see Raetz (1996).

The LPS genes of Bordetella pertussis

This gene cluster (Allen & Maskell, 1996) is particularly interesting as it has genes resembling genes from LPS core and O antigen clusters. The genes were named bpl and are now renamed wib. The structure of Bordetella pertussis LPS resembles a core with only one O unit. This is similar to the LPS structures sometimes called lipoooligosaccharides (LOS) and the name wib includes 'l' for LOS and 'b' for Bordetella. The cluster also includes waaA and waaC genes and a gene, wibG, which encodes a protein resembling wbaP.

O antigen genes of Pseudomonas aeruginosa

Since the Trends paper was submitted the O5 B-band O-antigen gene cluster of Pseudomonas aeruginosa has been sequenced and genes named using the BPNG scheme (Burrows et al., 1996). Most are given wbp names as only wzq and wzz genes were definitively identified. Several genes were tentatively identified as being in the putative ManNAc pathway and similarities were seen in some genes to genes of Bordetella pertussis LPS. The pathway genes will be renamed when better characterized and the steps common to the two gene clusters will be readily apparent.

Other gene clusters

The scheme could be applied to other gene clusters such as those for the type 2 capsules of E. coli and the related clusters for capsules of Neisseria meningitidis and Haemophilus influenzae but as yet no agreement has been reached on this.

Implementation of the BPNG scheme

The principles for allocation of names are:

1/ Pathway names only be used where there is a high degree of confidence that the gene carries out the particular function. However pathway gene names can be allocated before all the genes are known. For example the names gmhA,B,D have been reserved for the 4 step ADP-glycero-mannoheptose pathway (Fig 1 of the Trends paper) but the gmhC gene has not yet been identified. New pathway gene names should if possible avoid names already in use in E. coli or Salmonella or other species where it might cause confusion.

2/ Saccharide processing (wz*) gene names are only used where there is a high degree of confidence that the gene has the appropriate homology. In this case function is not necessarily well understood, as for example for wza, wzb or wzc, but there are good reasons to believe that the genes are homologous as they occur in the same order in addition to having good homology.

3/ w*** names are allocated to allow all genes within a cluster to share the first two letters and the fourth letter will normally be allocated in map order. However for gene clusters which require renaming the fourth letters from old names may be retained, eg Erwinia stewartii capsule genes. However where a w*** gene has already been named in another cluster it gets the same name. For example the rhamnose transferase genes in Salmonella group B and group E O antigen clusters are both named wbaN as they are clearly homologous: in this case the other w*** genes in both clusters are wba* but the gene would still be called wbaN if found in a cluster with different w** name for its cluster specific genes.

We suggest that gene names should be allocated after discussion with whoever is running the database, currently Peter Reeves in Sydney. This could precede publication if a name is needed for discussion within a lab or at meetings. In general a block of names would be reserved sufficient to cover the number of genes predicted for that cluster. The database would show that these names had been reserved but would not need to give any further information prior to publication. Names reserved but in the event not used would again be available for use. Many new genes will be named when only the sequence is known and before function is determined. If the gene is within an LPS or EPS cluster we suggest that it be named in the w*** series. If it is subsequently shown to have a function in a sugar biosynthetic pathway the gene would be given a new mnemonic name pertaining to the pathway. If it later turns out that the gene encodes a function which already has a symbol then precedence would normally apply and the name would have to change. Likewise if a gene is given an existing name, on the basis of sequence homology, but is later shown to encode a different function than expected, a new name would be given. To avoid this homology should be used as a basis for names only when the level of similarity is adequate to give a high level of confidence in the function.

The suggestion that researchers contact Peter Reeves is made to avoid the possibility that two research groups find the same gene at about the same time and give it different names, or that the same name be used by two different groups for different genes. However there can be no compulsion about this and there will inevitably be problems of overlap in names. This should become evident as soon as the data is published and hopefully one name will be changed.
It is preferable that discussions on the priority for the assignment of a gene symbol be resolved before publication and we suggest that the date for priority be the date that the sequence was released by GenBank or reserved by BPGD. Both are easily checked and it should be habit to check to see if a name has been used immediately prior to releasing it, and before it is too late to change the symbol in any publication. In this way the confusion will be confined to one lab and not enter the literature. It is of course advantageous in this sense at least to release or reserve names early as it removes the risk of having to change after the paper is published. It is also important to note that there are names which have been reserved but will not be seen in the database.

References

Examples of the application of the BPGN system


Affiliations and addresses of Authors.

P.R. Reeves is in the Department of Microbiology, University of Sydney, NSW 2006, Australia. tel: +61 2 9351 2536, fax: +61 2 9351 4571, email: reeves@angis.su.oz.au;

M. Hobbs is in the Centre for Molecular and Cellular Biology, University of Queensland, Qld 4072 Australia. tel: +61-7-3365 1819 Fax: +61-7-3365 4388, email: M.Hobbs@cmcb.uq.edu.au;

M. A. Valvano is in the Department of Microbiology and Immunology, University of Western Ontario, London, Ontario, N6A 5C1, Canada. tel: +1 519 6613996, fax: +1 519 661 3499, email: mvalvano@mni.uwo.ca;

M. Skurnik is in the Turku Centre for Biotechnology, PO Box 123, 20521 Turku, Finland. tel: +358-2-3338035, fax: +358-2-3338000, email: mskurnik@btu.utu.fi;

C. Whitfield is in the Department of Microbiology, University of Guelph, Guelph, Ontario, Canada N1G 2W1. tel: +1 519-824-4120 ext 3478, fax: +1 519-837-1802, email: CWHITFIE@micro.uoguelph.ca;

D. Coplin is in the Department of Plant Pathology, Ohio State University, Columbus, OH 43210-1087, USA. tel: +1 614-292-8503, fax: +1+ 614-292-4455, email: dave+@osu.edu;

N. Kido is in Biosystems, School of Informatics & Sciences, Nagoya University, Nagoya 464-01, Japan. tel: +81 52 789 4816, fax: +81 52 789 4818, email: j45811a@nucc.cc.nagoya-u.ac.jp;

J. Klena is a Lecturer in the Department of Plant and Microbial Sciences, University of Canterbury. Christchurch 4, New Zealand. tel +64 3 364 7001, fax +64 3 364 2083, email j.klena@botn.canterbury.ac.nz;

D. Maskell is in the Department of Clinical Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge CB3 0ES, UK. tel: +44 1223 339868, fax: +44 1223 337610, email: djm47@cam.ac.uk;

C. R. H. Raetz is in the Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, USA. tel: +1 919 684 5326, fax: +1 919 684 8885, email: raetz@bchm.biochem.duke.edu;

P. D. Rick is in the Department of Microbiology and Immunology, The Uniformed Services University of the Health Sciences, Bethesda, Maryland 20814-4799, USA. tel: +1 301 295-3418, fax: +1 301 295-1545, email: Rick@usuhsb.usuhs.mil.

Examples of the application of the BPGN system

There were 31 gene clusters referred to in table 1 of the Trends paper (Reeves et al., 1996), and listed below. Details including cluster map and structure of oligosaccharide synthesised have been extracted from BPGD and can be viewed. Note that only clusters those clusters in the original compilation are included. BPGD includes additional data and is updated regularly. A presentation similar to that provided here for 31 clusters can be obtained for any cluster in the database using "Cluster Report" in the BPGD Browser.

We thank Slade Jensen who prepared the information pages on the following gene clusters using BPGD.

<table>
<thead>
<tr>
<th>Gene Cluster</th>
<th>Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Bordetella pertussis</em></td>
<td>LPS</td>
</tr>
<tr>
<td><em>Erwinia stewartii</em></td>
<td>stewartan (Extracellular polysaccharide)</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>K-12 colanic acid</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>ECA</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>K-12 lipid A/core</td>
</tr>
<tr>
<td><em>Escherichia coli</em> (Shigella)</td>
<td>Dysenteriae type I O antigen</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>O9 O antigen</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>O16 (K-12) O antigen</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>O111 O antigen</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>O7 O antigen</td>
</tr>
<tr>
<td><em>Klebsiella pneumoniae</em></td>
<td>O1 O antigen</td>
</tr>
<tr>
<td><em>Klebsiella pneumoniae</em></td>
<td>O8 O antigen</td>
</tr>
<tr>
<td><em>Salmonella enterica</em></td>
<td>LT2 colanic acid</td>
</tr>
</tbody>
</table>
Salmonella enterica  ECA
Salmonella enterica  LT2 lipid A/core
Salmonella enterica  group B O antigen
Salmonella enterica  group C1 O antigen
Salmonella enterica  group C2 O antigen
Salmonella enterica  group D1 and A O antigens
Salmonella enterica  group D2 O antigen
Salmonella enterica  group E O antigen
Salmonella enterica  group O:54 sv Borreze O antigen
Serratia marcescens  O16 O antigen
Pseudomonas aeruginosa  O5 O antigen
Yersinia enterocolitica  O3 O antigen.
Yersinia enterocolitica  O8 O antigen.
Yersinia enterocolitica  O3 outer core.
Yersinia pseudotuberculosis  IA O antigen
Yersinia pseudotuberculosis  IIA O antigen
Yersinia pseudotuberculosis  IVA O antigen
Yersinia pseudotuberculosis  VA O antigen

**Bordetella pertussis** LPS

\[
\text{GlcNAc} \rightarrow \text{DAA} \rightarrow \text{FucNMeAc} \rightarrow \text{GlcN} \rightarrow \text{core} \rightarrow \text{Hep}
\]

DAA is 2,3-di-N-acetyl-2,3-dideoxyhexuronic acid or 2,3-di-N-acetyl-2,3-dideoxymannosaminuronic acid.
FucNMeAc is N-methyl-N-acetylfucosamine.

**B. pertussis**

**wlb** genes

\[
\text{Gene} \quad \text{Product Name} \quad \text{Old Names}
\begin{align*}
\text{waaA} & \quad \text{KDO transferase (inner core)} & \text{kdtA} \\
\text{waaC} & \quad \text{heptosyltransferase I (inner core)} & \text{rfaC} \\
\text{wlbA} & \quad \text{O-antigen biosynthetic protein} & \text{bplA} \\
\text{wlbB} & \quad \text{O-antigen biosynthetic protein} & \text{bplB} \\
\text{wlbC} & \quad \text{O-antigen biosynthetic protein} & \text{bplC} \\
\text{wlbD} & \quad \text{O-antigen biosynthetic protein} & \text{bplD} \\
\text{wlbE} & \quad \text{O-antigen biosynthetic protein} & \text{bplE} \\
\text{wlbF} & \quad \text{O-antigen biosynthetic protein} & \text{bplF}
\end{align*}
\]
wlbG  O-antigen biosynthetic protein  bplG
wlbH  O-antigen biosynthetic protein  bplH
wlbI  O-antigen biosynthetic protein  bplI
wlbJ  O-antigen biosynthetic protein  bplJ
wlbK  O-antigen biosynthetic protein  bplK
wlbL  O-antigen biosynthetic protein  bplL

References

1. Allen, A.G and Maskell, D.J.
The identification, cloning and mutagenesis of a genetic locus required for lipopolysaccharide biosynthesis in Bordatella pertussis.

*Erwinia stewartii* stewartan (Extracellular polysaccharide)

\[
\begin{align*}
\text{D-Glc} & \xrightarrow{\beta, 1,6} \text{D-Gal} \\
\text{D-Gal} & \xrightarrow{\alpha, 1,4} \text{D-GlcA} \\
\text{D-GlcA} & \xrightarrow{\beta, 1,4} 3 \left[ \text{D-Gal} \xrightarrow{\beta, 1,6} \text{D-Glc} \xrightarrow{\beta, 1,3} \text{D-Gal} \xrightarrow{\beta, 1} \right] \\
\text{D-Glc} & \xrightarrow{\alpha, 1,6} \\
\end{align*}
\]

*E. stewartii* wce genes

<table>
<thead>
<tr>
<th>G</th>
<th>wza</th>
<th>wzf</th>
<th>wzb</th>
<th>L</th>
<th>B</th>
<th>M</th>
<th>N</th>
<th>F</th>
<th>J</th>
<th>K</th>
<th>wzz</th>
<th>galF</th>
<th>galE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>5</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>galE</td>
<td>UDP-glucose 4-epimerase</td>
<td></td>
</tr>
<tr>
<td>galF</td>
<td>galactose-1-phosphate uridylyltransferase (regulatory subunit?)</td>
<td>cpsM</td>
</tr>
<tr>
<td>wceB</td>
<td>putative EPS biosynthetic protein</td>
<td>cpsE</td>
</tr>
<tr>
<td>wceF</td>
<td>putative EPS biosynthetic protein</td>
<td>cpsH</td>
</tr>
<tr>
<td>wceG</td>
<td>putative EPS biosynthetic protein</td>
<td>cpsA</td>
</tr>
<tr>
<td>wceJ</td>
<td>putative EPS biosynthetic protein</td>
<td>cpsJ</td>
</tr>
<tr>
<td>wceK</td>
<td>putative EPS biosynthetic protein</td>
<td>cpsK</td>
</tr>
<tr>
<td>wceL</td>
<td>putative EPS biosynthetic protein</td>
<td>cpsD</td>
</tr>
<tr>
<td>wceM</td>
<td>putative EPS biosynthetic protein</td>
<td>cpsF</td>
</tr>
<tr>
<td>wceN</td>
<td>putative EPS biosynthetic protein</td>
<td>cpsG</td>
</tr>
<tr>
<td>wza</td>
<td>outer membrane protein</td>
<td>cpsB</td>
</tr>
<tr>
<td>wzb</td>
<td>acid phosphatase</td>
<td>cpsI</td>
</tr>
<tr>
<td>wzc</td>
<td>ATP-binding protein with similarity to Wzz</td>
<td>cpsC</td>
</tr>
</tbody>
</table>
References

Nucleotide sequence analysis of the *Erwinia stewartii* *cps* gene cluster for synthesis of stewartan and comparison to the *Erwinia amylovora* *ams* cluster for synthesis of amylovoran.

Genetic transfer of amylovoran and stewartan synthesis between *Erwinia amylovora* and *Erwinia stewartii*.

3. Nimtz, M. Mort, A. Wray, V. Domke, T. Zhang, Y. Coplin, D.L. Geider, K.
Structural analysis of stewartan, the capsular exopolysaccharide from the corn pathogen *Erwinia stewartii*.

*Escherichia coli* K-12 colanic acid

\[
\begin{align*}
\text{D-Gal} & \quad \beta \quad \text{Pyr} \\
\end{align*}
\]

\[
\begin{align*}
\text{D-GlcA} & \quad \beta \quad \text{D-Gal} \\
\beta & \quad 1,3 \\
\end{align*}
\]

\[
\begin{align*}
\text{D-Gal} & \quad \beta \quad \text{OAc} \\
\beta & \quad 1,4 \\
\end{align*}
\]

\[
\begin{align*}
\text{L-Fuc} & \quad \beta \quad \text{L-Fuc} \quad \alpha \quad \text{D-Glc} \\
\beta & \quad 1,3 \\
\end{align*}
\]

\[
\begin{align*}
\text{Pyr is pyruvate linked acetalically to galactose}
\end{align*}
\]

\[
\begin{align*}
\text{E. coli} \quad \text{K-12 colanic acid} \\
\text{wca} \quad \text{genes}
\end{align*}
\]

\[
\begin{align*}
\text{galF} & \quad \text{galactase-1-phosphate uridylyltransferase} \quad (\text{regulatory subunit?}) \\
\text{gmd} & \quad \text{GDP-D-mannose 4,6-dehydratase} \\
\text{orf0.0} & \quad \text{adjacent}
\end{align*}
\]
manB phosphomannomutase
cpsG
manB phosphomannomutase
cpsB
manC D-mannose-1-phosphate guanylyltransferase
cpm
g
manC D-mannose-1-phosphate guanylyltransferase
pmi #
wcaA putative colanic acid glycosyltransferase
cpsB
wcaB colanic acid O-acetyltransferase II
cpsG
wcaC putative colanic acid glycosyltransferase
wcaD protein with multiple transmembrane segments
wcaE putative colanic acid glycosyltransferase
wcaF colanic acid O-acetyltransferase I
wcaG putative GDP-L-fucose pathway enzyme
wcaH putative colanic acid biosynthetic enzyme
wcaI putative colanic acid biosynthetic enzyme
wcaJ putative undecaprenylphosphate glucosephosphotransferase
wcaK putative colanic acid biosynthetic enzyme
wcaL putative colanic acid glycosyltransferase
wcaM putative colanic acid biosynthetic enzyme
wza outer membrane protein
wzb acid phosphatase
wzc ATP-binding protein with similarity to Wzz

# In M77127 manB and manC were incorrectly named pgm and pmi respectively.

References

1. Stevenson, G., Andrianopoulos, K., Hobbs, M. and Reeves, P.R.
   Organisation of the Escherichia coli K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid.

Escherichia coli ECA

\[
\begin{align*}
3 \overset{\text{D-Fuc4NAc} \overset{\alpha}{1,4} \text{D-ManNAc} \overset{\beta}{1,4} \text{D-GlcNAc}}{\rightarrow} \\
\end{align*}
\]

ECA has also been reported to be a cyclical structure

E. coli K-12 ECA
wec genes

\[
\begin{array}{|c|c|c|c|c|c|c|c|}
\hline
\text{rml} & A & \text{wzz} & B & C & \text{wzx} & D & E \\
\hline
\end{array}
\]

\[
\begin{array}{|c|}
\hline
\text{rml} \\
\hline
\end{array}
\]

<table>
<thead>
<tr>
<th>0</th>
<th>5</th>
<th>10</th>
<th>kb</th>
</tr>
</thead>
</table>

9
<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>rmlA</td>
<td>glucose-1-phosphate thymidylyltransferase</td>
<td>o292</td>
</tr>
<tr>
<td>rmlB</td>
<td>dtDP-D-glucose 4,6-dehydratase</td>
<td>o355</td>
</tr>
<tr>
<td>wecA</td>
<td>UDP-GlcNAc:undecaprenylphosphate</td>
<td>rfe</td>
</tr>
<tr>
<td></td>
<td>GlcNAc-1-phosphate transferase</td>
<td></td>
</tr>
<tr>
<td>wecB</td>
<td>UDP-N-acetylglucosamine-2-epimerase</td>
<td>nfrC</td>
</tr>
<tr>
<td>wecC</td>
<td>UDP-N-acetylglucosamine-2-epimerase</td>
<td>o389</td>
</tr>
<tr>
<td>wecD</td>
<td>ECA biosynthetic protein</td>
<td>o379</td>
</tr>
<tr>
<td>wecE</td>
<td>ECA biosynthetic protein</td>
<td>o181</td>
</tr>
<tr>
<td>wecF</td>
<td>ECA biosynthetic protein</td>
<td>o299</td>
</tr>
<tr>
<td>wecG</td>
<td>4-amino-4,6-dideoxy-D-galactosyltransferase</td>
<td>rffT</td>
</tr>
<tr>
<td></td>
<td>Lipopolysaccharide</td>
<td>o246</td>
</tr>
<tr>
<td>wzx</td>
<td>N-acetylmannosaminuronosyltransferase</td>
<td></td>
</tr>
<tr>
<td>wzz</td>
<td>Flippase</td>
<td>o416</td>
</tr>
<tr>
<td></td>
<td>Chain length determinant</td>
<td>o349</td>
</tr>
</tbody>
</table>

References

1. Meier-Dieter, U., Barr, K., Starman, R., Hatch, L. and Rick, P.D.
Nucleotide sequence of the *Escherichia coli* rfe gene involved in the synthesis of Enterobacterial common antigen. Moleculat cloning of the rfe-rff gene cluster.

Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes.

A cytoplasmic protein, NfrC, is required for bacteriophage N4 adsorption.

4. Marolda, C.L. and Valvano, M.A.

*Escherichia coli* K-12 lipid A/core
KDO is 3-deoxy-D-manno-octulosonic acid

PPEA is phosphoethanolamine

**References**

Cloning, expression and characterisation of the *Escherichia coli* K-12 rfaD gene.

2. Clementz, T. and Raetz, C.R.
A gene coding for 3-deoxy-D-manno-octulosonic-acid transferase in *Escherichia coli*. Identification, mapping, cloning and sequencing.

3. Parker, C.T., Pradel, E. and Schnaitman, C.A.
Identification and sequences of the lipopolysaccharide core biosynthetic genes rfaQ, rfaP and rfaG of *Escherichia coli* K-12.
4. Pradel, E., Parker, C.T. and Schnaitman, C.A.
Structures of the \textit{rfaB, rfaI, rfaJ, and rfaS} genes of \textit{Escherichia coli} K-12 and their roles in assembly of the lipopolysaccharide core.

5. Klena, J.D., Pradel, E. and Schnaitman, C.A.
Comparison of lipopolysaccharide biosynthesis genes \textit{rfaK, rfaL, rfaY, and rfaZ} of \textit{Escherichia coli} K-12 and \textit{Salmonella typhimurium}.

6. Clementz, T.
The gene coding for 3-deoxy-manno-octulosonic acid transferase and the \textit{rfaQ} gene are transcribed from divergently arranged promoters in \textit{Escherichia coli}.

7. Schnaitman, C.A. Parker, C.T. Klena, J.D. Pradel, E.L. Pearson, N.B. Sanderson, K.E. Maclachlan, P.R.
Physical maps of the \textit{rfa} Loci of \textit{Escherichia coli} K-12 and \textit{Salmonella typhimurium}

\textit{Escherichia coli} (Shigella) Dysenteriae type I O antigen

\[
\begin{align*}
\quad & \quad L-Rha \quad a_{1,2} \quad L-Rha \quad a_{1,3} \quad D-Gal \quad a_{1,2} \quad D-GlcNac \quad a_{1,3} \\
\quad & \quad 3 \\
\end{align*}
\]

Note that the Dysenteriae O-antigen genes are divided between chromosome and plasmid

\textit{Sh. dysenteriae}
\textit{wbb} genes (chromosome)

\[
\begin{center}
\begin{array}{cccccccccccc}
\text{rml} & \text{wzx} & \text{wzy} & \text{ORI} \\
\text{B} & \text{D} & \text{A} & \text{C} & \text{R} & \text{Q} & \text{O} \\
\end{array}
\end{center}
\]

\begin{center}
\begin{tikzpicture}
\end{tikzpicture}
\end{center}

non-functional gene

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{rmlA}</td>
<td>glucose-1-phosphate thymidylyltransferase</td>
<td>\textit{rfbA}</td>
</tr>
<tr>
<td>\textit{rmlB}</td>
<td>dTDP-D-glucose 4,6-dehydratase</td>
<td>\textit{rfbB}</td>
</tr>
<tr>
<td>\textit{rmlC}</td>
<td>dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase</td>
<td>\textit{rfbC}</td>
</tr>
</tbody>
</table>
**References**

1. Sturm S. Jann B. Jann K. Fortnagel P. Timmis K.N.  
Genetic and biochemical analysis of Shigella dysenteriae 1 O antigen polysaccharide biosynthesis in *Escherichia coli* K-12: stucture and functions of the *rfb* cluster.  

2. Klena JD. Schnaitman CA.  
Function of the *rfb* gene cluster and the *rfe* gene in the synthesis of O antigen by *Shigella dysenteriae* 1.  

\[\text{**Sh. dysenteriae pHW400**} \]
\[\text{**wbb** genes (plasmid)} \]

\[\text{Gene} \quad \text{Product Name} \quad \text{Old Name} \]
\[wbbP \quad \text{galactosyltransferase} \quad \text{rfp} \]
\[wbbP \quad \text{galactosyltransferase} \quad \text{rfpB} \]

**References**

Role of Escherichia coli K-12 *rfa* genes and the *rfp* gene of *Shigella dysenteriae* 1 in generation of lipopolysaccharide core heterogeneity and attatchment of O antigen.  

2. Gohmann S. Manning P.A. Alpert C.A. Walker M.J. Timmis K.N.  
Lipopolysaccharide O antigen biosynthesis in *Shigella dysenteriae* serotype 1: analysis of the plasmid-carried *rfp* determinant.  
Microbial Pathogenesis 16 53-64, 1994.
**Escherichia coli** O9 O antigen

Note that *E.coli* O9 O-antigen differs in the 2 cases sequenced

\[
\begin{align*}
\text{E. coli O9} \\
\text{wbd genes}
\end{align*}
\]

<table>
<thead>
<tr>
<th>Gene</th>
<th><strong>Product Name</strong></th>
<th><strong>Old Name</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>manB</td>
<td>phosphomannomutase</td>
<td>rfbK</td>
</tr>
<tr>
<td>manC</td>
<td>D-mannose-1-phosphate guanylyltransferase</td>
<td>rfbM</td>
</tr>
<tr>
<td>wbdA</td>
<td>mannosyltransferase I</td>
<td>mtfA</td>
</tr>
<tr>
<td>wbdB</td>
<td>mannosyltransferase II</td>
<td>mtfB</td>
</tr>
<tr>
<td>wbdC</td>
<td>mannosyltransferase III</td>
<td>mtfC</td>
</tr>
<tr>
<td>wbdD</td>
<td>O-antigen biosynthetic protein</td>
<td>orf708</td>
</tr>
<tr>
<td>wzm</td>
<td>ABC-2 type transport system integral membrane protein</td>
<td>orf261</td>
</tr>
<tr>
<td>wzt</td>
<td>ABC-2 type transport system ATP-binding protein</td>
<td>orf431</td>
</tr>
</tbody>
</table>

**References**


**E. coli O9**

*wbd* genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Name</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>manB</em></td>
<td>phosphomannomutase</td>
<td><em>rfbK1</em></td>
</tr>
<tr>
<td><em>manB</em></td>
<td>phosphomannomutase</td>
<td><em>rfbK2</em></td>
</tr>
<tr>
<td><em>manC</em></td>
<td>D-mannose-1-phosphate guanylyltransferase</td>
<td><em>rfbM1</em></td>
</tr>
<tr>
<td><em>manC</em></td>
<td>D-mannose-1-phosphate guanylyltransferase</td>
<td><em>rfbM2</em></td>
</tr>
</tbody>
</table>

References

1. Jayaratne P., Bronner D., MacLachlan P.R., Dodgson C., Kido N., Whitfield C.
   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.

**Escherichia coli** O16 (K-12) O antigen

**E. coli** K-12 O16 O Ag

*wbb* genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>glf</td>
<td>UDP-galactopyranose mutase</td>
<td><em>orf6</em></td>
</tr>
<tr>
<td><em>rmlA</em></td>
<td>glucose-1-phosphate thymidylyltransferase</td>
<td><em>rfbA</em></td>
</tr>
</tbody>
</table>
\begin{align*}
\text{rmIB} & \quad \text{dTDP-D-glucose 4,6-dehydratase} & \quad \text{rfbB} \\
\text{rmIC} & \quad \text{dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase} & \quad \text{rfbC} \\
\text{rmID} & \quad \text{dTDP-4-keto-L-rhamnose reductase} & \quad \text{rfbD} \\
\text{wbbH} & \quad \text{putative O-unit polymerase} & \quad \text{orf4} (\text{Yao}) \\
\text{wbbH} & \quad \text{putative P-unit polymerase} & \quad \text{rfc} (\text{Stevenson}) \\
\text{wbbI} & \quad \text{galactofuranosyltransferase} & \quad \text{orf3} (\text{Yao}) \\
\text{wbbI} & \quad \text{galactofuranosyltransferase} & \quad \text{orf8} (\text{Stevenson}) \\
\text{wbbJ} & \quad \text{acetyltransferase} & \quad \text{orf2} (\text{Yao}) \\
\text{wbbJ} & \quad \text{acetyltransferase} & \quad \text{orf8} (\text{Stevenson}) \\
\text{wbbK} & \quad \text{glucosyltransferase} & \quad \text{ORF224} (\text{Liu}) \\
\text{wbbK} & \quad \text{glucosyltransferase} & \quad \text{orf1} (\text{Yao}) \\
\text{wbbK} & \quad \text{glucosyltransferase} & \quad \text{orf10} (\text{Stevenson}) \\
\text{wbbL} & \quad \text{rhamnosyltransferase} & \quad \text{ORF264} (\text{Liu}) \\
\text{wbbL} & \quad \text{rhamnosyltransferase} & \quad \text{orf5} (\text{Yao}) \\
\text{wzx} & \quad \text{flippase} & \quad \text{rfbx}
\end{align*}

\textbf{References}

1. Liu, D. and Reeves, P.R.
\textit{Escherichia coli} K12 regains its O antigen.

2. Yao, Z. and Valvano, M.A.

Structure of the O antigen of \textit{Escherichia coli} K-12 and the sequence of its \textit{rfb} gene cluster.

\textit{Escherichia coli} O111 O antigen

\[
\begin{array}{c}
\text{Col} \\
\alpha \downarrow 1,3
\end{array}
\begin{array}{c}
\text{Col} \\
\alpha \downarrow 1,6
\end{array}
\begin{array}{c}
\alpha \downarrow 1,4
\end{array}
\begin{array}{c}
\alpha \downarrow 1,3
\end{array}
\begin{array}{c}
\beta \downarrow 1
\end{array}
\frac{4}{4} \left[ \text{D-Glc} \begin{array}{c}
\alpha \downarrow 1,3
\end{array} \text{D-Gal} \begin{array}{c}
\alpha \downarrow 1,3
\end{array} \text{D-GlcNAc} \begin{array}{c}
\beta \downarrow 1
\end{array} \right]
\text{Col is 3,6-dideoxy-L-galactose}
\]
**E. coli O111**

*wb* genes

![Gene Map](image)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>gmd</td>
<td>GDP-D-mannose 4,6-dehydratase</td>
<td>orf0.0</td>
</tr>
<tr>
<td>manB</td>
<td>phosphomannomutase</td>
<td>rfbK</td>
</tr>
<tr>
<td>manC</td>
<td>D-mannose-1-phosphate guanylyltransferase</td>
<td>rfbM</td>
</tr>
<tr>
<td>ugd</td>
<td>UDP-glucose 6-dehydrogenase</td>
<td>orf1</td>
</tr>
<tr>
<td>wbdI</td>
<td>putative O-antigen biosynthetic protein</td>
<td>orf3.4</td>
</tr>
<tr>
<td>wbdJ</td>
<td>putative O-antigen biosynthetic protein</td>
<td>orf6.7</td>
</tr>
<tr>
<td>wbdK</td>
<td>putative O-antigen biosynthetic protein</td>
<td>orf7.7</td>
</tr>
<tr>
<td>wzx</td>
<td>flippase</td>
<td>orf8.9</td>
</tr>
<tr>
<td>wzz</td>
<td>chain length determinant</td>
<td>cld</td>
</tr>
</tbody>
</table>

**References**

1. Bastin, D.A. and Reeves, P.R.  
   Sequence and analysis of the O antigen gene (*rfb*) cluster of *Escherichia coli* O111.  

2. Bastin, D.A., Stevenson, G., Brown, P.K., Haase, A. and Reeves, P.R.  
   Repeat unit polysaccharides of bacteria: a model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length.  
   Molecular Microbiology. 7: 725-734, 1993.

*Escherichia coli* O7 O antigen
VsmNAc is 4-acetamido-4,6-dideoxy-D-glucose

E. coli O7

\[ VsmNAc \xrightarrow{\beta} D-Man \xrightarrow{\alpha} D-Gal \xrightarrow{\beta} D-GlcNAc \xrightarrow{\alpha} 1 \]

\[ 3 \rightarrow \]

\[ L-Rha \xrightarrow{\alpha} 1,3 \]

\[ \xrightarrow{1,2} \]

\[ \xrightarrow{1,4} \]

\[ \xrightarrow{1,3} \]

\[ \xrightarrow{1} \]

References

1. Marolda, C.L. and Valvano, M.A.
Identification, expression and DNA sequence of the GDP-mannose biosynthesis genes encoded by the 07 rfb gene cluster of strain VW187 (Escherichia coli 07:K1).

2. Marolda, C.L. and Valvano, M.A.

Klebsiella pneumoniae O1 O antigen
\[
3 \left[ D\text{-Gal} \xrightarrow{\beta} D\text{-Galp} \right] \xrightarrow{\alpha} 1
\]

\[K. \textit{pneumoniae} \text{ O1}\]
\[wbb \text{ genes}\]

\[\begin{array}{cccc}
\text{wzm} & \text{wzt} & M & \text{glf} & N & O
\end{array}\]

\[
\begin{array}{ccc}
0 & \text{kb} & 5
\end{array}
\]

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>glf</td>
<td>UDP-galactopyranose mutase</td>
<td>rfbD</td>
</tr>
<tr>
<td>wbbM</td>
<td>protein of unknown function</td>
<td>rfbC</td>
</tr>
<tr>
<td>wbbN</td>
<td>protein of unknown function</td>
<td>rfbE</td>
</tr>
<tr>
<td>wbbO</td>
<td>galactosyltransferase</td>
<td>rfbF</td>
</tr>
<tr>
<td>wzm</td>
<td>ABC-2 type transport system integral membrane</td>
<td>rfbA</td>
</tr>
<tr>
<td>wzt</td>
<td>ABC-2 type transport system ATP-binding protein</td>
<td>rfbB</td>
</tr>
</tbody>
</table>

References

1. Bronner, D., Clarke, B.R. and Whitfield, C.
Identification of an ATP-binding cassette transport system required for translocation of lipopolysaccharide O-antigen side-chains across the cytoplasmic membrane of \textit{Klebsiella pneumoniae} serotype 01.

\[K. \textit{pneumoniae} \text{ O8} \text{ O antigen}\]

\[
3 \left[ D\text{-Gal} \xrightarrow{\beta} D\text{-Galp} \right] \xrightarrow{\alpha} 1
\]

\[K. \textit{pneumoniae} \text{ O8}\]
\[wbb \text{ genes}\]

\[\begin{array}{cccc}
\text{wzm} & \text{wzt} & M & \text{glf} & N & O
\end{array}\]

\[
\begin{array}{ccc}
0 & \text{kb} & 5
\end{array}
\]
<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>glf</td>
<td>UDP-galactopyranose mutase</td>
<td>rfbD</td>
</tr>
<tr>
<td>wbbM</td>
<td>protein of unknown function</td>
<td>rfbC</td>
</tr>
<tr>
<td>wbbN</td>
<td>protein of unknown function</td>
<td>rfbE</td>
</tr>
<tr>
<td>wbbO</td>
<td>galactosyltransferase</td>
<td>rfbF</td>
</tr>
<tr>
<td>wzm</td>
<td>ABC-2 transport system integral membrane protein</td>
<td>rfbA</td>
</tr>
<tr>
<td>wzt</td>
<td>ABC-2 type transport system ATP-binding protein</td>
<td>rfbB</td>
</tr>
</tbody>
</table>

**References**


*Salmonella enterica* LT2 colanic acid

```
  D-Gal \beta^{1,4}  Pyr
     \downarrow \beta^{1,4}
  D-GlcA
      \downarrow \beta^{1,3}
  D-Gal
       \downarrow \beta^{1,4}
 \left[\begin{array}{c}
    L-Fuc \\
    \beta^{1,4} \\
  \end{array}\right] \xrightarrow{\beta^{1,3}} D-Glc
```

Pyr is pyruvate linked acetalically to galactose
**Gene** | **Product Name** | **Old Names**
--- | --- | ---
galF | galactose-1-phosphate uridylyltransferase | orf2.8
GDP-D-mannose 4,6-dehydratase | gmd
Phosphomannomutase | manB
D-mannose-1-phosphate guanylyltransferase | manC
cpsG
Putative colanic acid glycosyltransferase | wcaL
Putative colanic acid biosynthetic enzyme | wcaM
Orf0.0

**References**

   Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2).

2. Stevenson, G., Lee, S.J., Romana, L.K. and Reeves, P.R.
   The *cps* gene cluster of Salmonella strain LT2 includes a second mannose pathway: sequence of two genes and relationship to genes in the *rfb* gene cluster.

**Salmonella enterica** LT2 lipid A/core

KDO is 3-deoxy-D-manno-octulosonic acid
PPEA is phosphoethanolamine
<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>gmhD</td>
<td>ADP-heptose 6-epimerase</td>
<td>rfaD</td>
</tr>
<tr>
<td>waaC</td>
<td>heptosyltransferase I (inner core)</td>
<td>rfaC</td>
</tr>
<tr>
<td>waaF</td>
<td>heptosyltransferase II (inner core)</td>
<td>rfaF</td>
</tr>
<tr>
<td>waaI</td>
<td>glycosyltransferase</td>
<td>rfaI</td>
</tr>
<tr>
<td>waaJ</td>
<td>glucosyltransferase II (outer core)</td>
<td>rfaJ</td>
</tr>
<tr>
<td>waaK</td>
<td>protein of unknown function</td>
<td>rfaK</td>
</tr>
<tr>
<td>waaL</td>
<td>O-antigen ligase</td>
<td>rfaL</td>
</tr>
<tr>
<td>waaY</td>
<td>protein of unknown function</td>
<td>rfaY</td>
</tr>
<tr>
<td>waaZ</td>
<td>protein of unknown function</td>
<td>rfaZ</td>
</tr>
</tbody>
</table>

References

1. MacLachlan, P.R., Kadam, S.K. and Sanderson, K.E.
   Cloning, characterisation and DNA sequence of the rfaLK region for lipopolysaccharide synthesis in *Salmonella typhimurium* LT2.

2. Sirisena, D.M., Brozek, K.A., MacLachlan, P.R., Sanderson, K.E. and Raetz, C.R.
   the rfaC gene of Salmonella typhimurium. Cloning, sequencing and enzymatic function in heptose transfer to lipopolysaccharide.

3. Schnaitman, C.A. Parker, C.T. Klena, J.D. Pradel, E.L. Pearson, N.B. Sanderson, K.E. Maclachlan, P.R.
   Physical maps of the rfa loci of *Escherichia coli* K-12 and *Salmonella typhimurium*

4. Carstenius, P. Flock, J.I. Lindberg, A.
   Nucleotide sequence of rfaI and rfaJ genes encoding lipopolysaccharide glycosyl transferases from *Salmonella typhimurium*
   Nucleic Acids Research 18: 6128

*Salmonella enterica* group A O antigen
Gene cluster same as Se D1 O-Ag except tyv has a frame shift mutation

References

1. Liu, D., Verma, N.K., Romana, L.K. and Reeves, P.R.
Relationships among the rfb regions of Salmonella serovars A, B and D.

Salmonella enterica group B O antigen
Abe is 3,6-dideoxy-D-galactose

S. enterica sv. Typhimurium
wba genes

References

1. Wyk, P. and Reeves, P. R.
Identification and sequence of the gene for abequose synthase, which confers antigenic specificity on group B Salmonellae: homology with galactose epimerase.
2. Jiang, X.M., Neal, B., Santiago, F., Lee, S.J., Romano, L.K. and Reeves, P.R.
Structure and sequence of the \textit{rfb} (O antigen) gene cluster of \textit{Salmonella} serovar typhimurium (strain LT2).

3. Reeves, P.
Evolution of \textit{Salmonella} O antigen variation by interspecific gene transfer on a large scale. [Review].

Repeat unit polysaccharides of bacteria: a model for polymerization resembling that of ribosomes and fatty acid synthetase, with a novel mechanism for determining chain length.
Molecular Microbiology 7: 725-34, 1993.

5. Liu, D., Haase, A.M., Lindqvist, L., Lindberg, A.A. and Reeves, P.R.

6. Liu, D., Lindqvist, L. and Reeves, P.R.

\textbf{\textit{Salmonella enterica} group C1 O antigen}

\[
\text{Glc} \rightarrow 2 \left[ \text{D-Man}_1^\alpha \text{D-Man}_2^\alpha \text{D-Man}_2^\beta \text{D-Man}_2^\beta \text{D-GlcNAc}_1^\beta \right]
\]
**S. enterica sv. Montevideo**

**wba genes**

![Genome Map](image)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>galF</td>
<td>galactose-1-phosphate uridylyltransferase</td>
<td>orf2.8</td>
</tr>
<tr>
<td>manB</td>
<td>phosphomannomutase</td>
<td>orf9.79</td>
</tr>
<tr>
<td>manC</td>
<td>D-mannose-1-phosphate guanylyltransferase</td>
<td>orf8.36</td>
</tr>
<tr>
<td>wbaA</td>
<td>possible O-antigen polymerase</td>
<td>orf4.11</td>
</tr>
<tr>
<td>wbaB</td>
<td>possible mannosyltransferase I</td>
<td>orf5.19</td>
</tr>
<tr>
<td>wbaC</td>
<td>possible mannosyltransferase II</td>
<td>orf6.17</td>
</tr>
<tr>
<td>wbaD</td>
<td>possible mannosyltransferase III</td>
<td>orf7.17</td>
</tr>
<tr>
<td>wzx</td>
<td>flippase</td>
<td>rfbX</td>
</tr>
</tbody>
</table>

### References

1. Lee, S.J., Romano, L.K. and Reeves, P.R.  
Cloning and structure of group C1 O antigen (rfb gene cluster) from *Salmonella enterica* serovar montevideo.  

2. Lee, S.J., Romana, L.K. and Reeves, P.R.  
Sequence and structural analysis of the rfb (O antigen) gene cluster from a group C1 Salmonella enterica strain.  

3. Reeves, P.  
Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale. [Review].  

4. Hobbs, M. and Reeves, P.R.  
The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters.  
Salmonella enterica group C2 O antigen

\[
\text{Abe} \xrightarrow{\alpha_1,3} \text{L-Rha} \xrightarrow{\alpha_1,2} \text{D-Man} \xrightarrow{\alpha_1,2} \text{D-Man} \xrightarrow{\alpha_1,3} \text{D-Gal} \xrightarrow{\alpha_1}
\]

Abe is 3,6-dideoxy-D-galactose

S. enterica sv. Muenchen

\text{wba} \text{genes}

\text{rml}  \quad \text{ddh}  \quad \text{man}

\begin{array}{cccccccccc}
\text{B} & \text{D} & \text{A} & \text{C} & \text{D} & \text{A} & \text{B} & \text{C} & \text{abe} & \text{wzx} & \text{R} & \text{L} & \text{Q} & \text{wzy} & \text{W} & \text{Z} & \text{C} & \text{B} & \text{P} & \text{gnd}
\end{array}

0 \quad 5 \quad 10 \quad \text{kb}

<table>
<thead>
<tr>
<th>Gene</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>abe</td>
<td>CDP-abequose synthase</td>
</tr>
<tr>
<td>ddhA</td>
<td>D-glucose-1-phosphate cytidylyltransferase</td>
</tr>
<tr>
<td>ddhB</td>
<td>CDP-glucose 4,6-dehydratase</td>
</tr>
<tr>
<td>ddhC</td>
<td>CDP-4-keto-6-D-glucose 3-dehydrase</td>
</tr>
<tr>
<td>ddhD</td>
<td>CDP-6-deoxy-delta3,4-glucose reductase</td>
</tr>
<tr>
<td>manB</td>
<td>phosphomannomutase</td>
</tr>
<tr>
<td>manC</td>
<td>D-mannose-1-phosphate guanylyltransferase</td>
</tr>
<tr>
<td>rmlA</td>
<td>glucose-1-phosphate thymidylyltransferase</td>
</tr>
<tr>
<td>rmlB</td>
<td>dTDP-D-glucose 4,6-dehydratase</td>
</tr>
<tr>
<td>rmlC</td>
<td>dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase</td>
</tr>
<tr>
<td>rmlD</td>
<td>dTDP-4-keto-L-rhamnose reductase</td>
</tr>
<tr>
<td>wbaL</td>
<td>O-acetyltransferase</td>
</tr>
<tr>
<td>wbaL</td>
<td>O-acetyltransferase</td>
</tr>
<tr>
<td>wbaP</td>
<td>undecaprenylphosphate galactosephosphotransferase</td>
</tr>
<tr>
<td>wbaQ</td>
<td>rhamnosyltransferase</td>
</tr>
<tr>
<td>wbaR</td>
<td>abequosyltransferase</td>
</tr>
<tr>
<td>wbaR</td>
<td>abequosyltransferase</td>
</tr>
<tr>
<td>wbaW</td>
<td>mannosyltransferase II</td>
</tr>
<tr>
<td>wbaW</td>
<td>mannosyltransferase II</td>
</tr>
<tr>
<td>wbaZ</td>
<td>mannosyltransferase I</td>
</tr>
<tr>
<td>wbaZ</td>
<td>mannosyltransferase I</td>
</tr>
<tr>
<td>wzx</td>
<td>flippase</td>
</tr>
<tr>
<td>wzx</td>
<td>flippase</td>
</tr>
<tr>
<td>wzy</td>
<td>polymerase</td>
</tr>
</tbody>
</table>

References
1. Brown, P.K., Romana, L.K. and Reeves, P.R.
Molecular analysis of the rfb gene cluster of *Salmonella* serovar muenchen (strain M67): the genetic basis of the polymorphism between groups C2 and B.
Molecular Microbiology. 6: 1385-1394, 1992.

2. Liu, D., Haase, A.M., Lindqvist, L., Lindberg, A.A. and Reeves, P.R.
Glycosyl transferases of O-antigen biosynthesis in *Salmonella enterica*: identification and characterisation of transferase genes of groups B, C2 and E1.

**Salmonella enterica** group D1O antigen

Tyv
\[
\begin{array}{c}
\text{Glc-2OAc} \\
\alpha \downarrow 1,3 \\
\end{array}
\]

\[
\begin{array}{c}
\text{D-Man} \\
\alpha \downarrow 1,4 \\
\end{array}
\begin{array}{c}
\text{L-Rha} \\
\alpha \downarrow 1,3 \\
\end{array}
\begin{array}{c}
\text{D-Gal} \\
\beta \downarrow 1 \\
\end{array}
\]

Tyv is 3,6-dideoxy-D-mannose

*S. enterica* sv. Typhi

**wba** genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>ddhA</td>
<td>D-glucose-1-phosphate cytidylyltransferase</td>
<td>rfbF</td>
</tr>
<tr>
<td>ddhB</td>
<td>CDP-glucose 4,6-dehydratase</td>
<td>rfbG</td>
</tr>
<tr>
<td>ddhC</td>
<td>CDP-4-keto-6-deoxy-D-glucose 3-hydrase</td>
<td>rfbF</td>
</tr>
<tr>
<td>ddhD</td>
<td>CDP-6-deoxy-delta3,4-glucose reductase</td>
<td>rfbI</td>
</tr>
<tr>
<td>manB</td>
<td>phosphomannomutase</td>
<td>rfbK</td>
</tr>
<tr>
<td>manC</td>
<td>D-mannose-1-phosphate guanylyltransferase</td>
<td>rfbM</td>
</tr>
<tr>
<td>prt</td>
<td>paratose synthetase</td>
<td>rfbS</td>
</tr>
<tr>
<td>rmlA</td>
<td>glucose-1-phosphate guanylyltransferase</td>
<td>rfbA</td>
</tr>
<tr>
<td>rmlB</td>
<td>dTDP-D-glucose 4,6-dehydratase</td>
<td>rfbB</td>
</tr>
<tr>
<td>rmlC</td>
<td>dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase</td>
<td>rfbC</td>
</tr>
<tr>
<td>rmlD</td>
<td>dTDP-4-keto-L-rhamnose reductase</td>
<td>rfbD</td>
</tr>
<tr>
<td>tyv</td>
<td>CDP-paratose epimerase</td>
<td>rfbE</td>
</tr>
<tr>
<td>wbaN</td>
<td>rhamnosyltransferase</td>
<td>rfbN</td>
</tr>
<tr>
<td>wbaP</td>
<td>undecaprenylphosphate galactosephosphotransfer</td>
<td>rfbB</td>
</tr>
<tr>
<td>wbaU</td>
<td>mannosyltransferase</td>
<td>orf14.1</td>
</tr>
<tr>
<td>wbaV</td>
<td>dideoxyhexosyltransferase</td>
<td>rfbV</td>
</tr>
<tr>
<td>wxz</td>
<td>flippase</td>
<td>orf12.8</td>
</tr>
<tr>
<td>wxz</td>
<td>flippase</td>
<td>rfbX</td>
</tr>
</tbody>
</table>
References

1. Liu, D., Verma, N.K., Romana, L.K. and Reeves, P.R.

Relationships among the *rfb* regions of *Salmonella* serovars A, B and D.

*Salmonella enterica* group D2 O antigen

\[
\text{Tvyv} \\
\alpha \begin{array}{c}
\beta \\
1,3 \\
\alpha \\
1,4 \\
\alpha \\
1,3 \\
\end{array} \\
\begin{array}{c}
\rightarrow \\
\rightarrow \\
\rightarrow \\
\rightarrow \\
\rightarrow \\
\rightarrow \\
\end{array}
\]

\[
\begin{array}{c}
\text{D-Man} \\
\text{L-Rha} \\
\text{D-Gal} \\
\end{array}
\]

\[
\text{Tvyv is 3,6-dideoxy-D-mannose}
\]

*S. enterica* sv. Strasbourg

*wb* genes

<table>
<thead>
<tr>
<th>rml</th>
<th>ddh</th>
<th>man</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>D</td>
<td>A</td>
</tr>
<tr>
<td>D</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>C</td>
<td>par</td>
<td>tyv</td>
</tr>
<tr>
<td>wzy</td>
<td>V</td>
<td>Hept</td>
</tr>
<tr>
<td>wzy</td>
<td>G</td>
<td>N</td>
</tr>
<tr>
<td>C</td>
<td>B</td>
<td>P</td>
</tr>
</tbody>
</table>

Gene Product Name Old Names

ddhA D-glucose-1-phosphate cytidylyltransferase rfbF

ddhB CDP-glucose 4,6-dehydratase rfbG

ddhC CDP-4-keto-6-deoxy-D-glucose 3-dehydratase rfbH

ddhD CDP-6-deoxy-delt3,4-glucoseon reductase rfbI

manB phosphomannomutase rfbK

manC D-mannose-1-phosphate guanylyltransferase rfbM

prt paratose synthetase rfbS

rmlA glucose-1-phosphate thymidylyltransferase rfbA

rmlB dTDP-D-glucose 4,6-dehydratase rfbB

rmlC dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase rfbC

rmlD dTDP-4-keto-L-rhamnose reductase rfbD

tyv CDP-paratose epimerase rfbE

wbaN rhamnosyltransferase rfbN

wbaO mannosyltransferase rfbO

wbaP undecaprenylphosphate galactosephosphotransferase rfbP

wbaV dideoxyhexosyltransferase rfbV

wzx flippase rfbX

wzy polymerase rfc

References
1. Reeves, P.
Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale. [Review].

2. Xiang, S.H, Hobbs, M. and Reeves, P.R.
Molecular analysis of the rfb gene cluster of a group D2 Salmonella enterica strain: evidence for its origin from an insertion sequence-mediated recombination event between group E and D1 strains.

Salmonella enterica group E O antigen

\[
\begin{align*}
\text{OAc} & \\
\rightarrow & \\
\text{D-Man} & \quad \beta \\
\rightarrow & \quad 1,4 \\
\text{L-Rha} & \quad \alpha \\
\rightarrow & \quad 1,3 \\
\text{D-Gal} & \quad \alpha \\
\end{align*}
\]

S. enterica sv. Anatum
wba genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>manB</td>
<td>phosphomannomutase</td>
<td>rfbK</td>
</tr>
<tr>
<td>manC</td>
<td>D-mannose-1-phosphate guanylyltransferase</td>
<td>rfbM</td>
</tr>
<tr>
<td>rmlA</td>
<td>glucose-1-phosphate thymidylyltransferase</td>
<td>orf6.1</td>
</tr>
<tr>
<td>rmlA</td>
<td>glucose-1-phosphate thymidylyltransferase</td>
<td>rfbA</td>
</tr>
<tr>
<td>rmlB</td>
<td>dTDP-D-glucose 4,6-dehydratase</td>
<td>rfbB</td>
</tr>
<tr>
<td>rmlC</td>
<td>dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase</td>
<td>orf7.0</td>
</tr>
<tr>
<td>rmlD</td>
<td>dTDP-4-keto-L-rhamnose reductase</td>
<td>orf5.2</td>
</tr>
<tr>
<td>wbaE</td>
<td>possible additional O-antigen polymerase</td>
<td>orf17.4</td>
</tr>
<tr>
<td>wbaN</td>
<td>rhamnosyltransferase</td>
<td>orf11.9</td>
</tr>
<tr>
<td>wbaN</td>
<td>rhamnosyltransferase</td>
<td>rfbN</td>
</tr>
<tr>
<td>wbaO</td>
<td>mannosyltransferase</td>
<td>orf10.9</td>
</tr>
<tr>
<td>wbaO</td>
<td>mannosyltransferase</td>
<td>rfbO</td>
</tr>
<tr>
<td>wbaP</td>
<td>undecaprenylphosphate galactosephosphotransferase</td>
<td>orf11.9</td>
</tr>
<tr>
<td>wzx</td>
<td>flippase</td>
<td>orf7.9</td>
</tr>
<tr>
<td>wzy</td>
<td>polymerase</td>
<td>orf9.6</td>
</tr>
</tbody>
</table>
References

1. Wang, L., Romana, L.K. and Reeves, P.R.
Molecular analysis of a Salmonella enterica group E1 rfb gene cluster: O antigen and the genetic basis of the major polymorphism.

2. Liu, D., Haase, A.M., Lindqvist, L., Lindberg, A.A. and Reeves, P.R.

Salmonella enterica group O:54 sv Borreze O antigen

\[
\begin{align*}
4 \left[ D-\text{ManN}Ac \xrightarrow{\beta} D-\text{ManN}Ac \right] \xrightarrow{\beta} 1
\end{align*}
\]

S. enterica sv. Borreze pWQ799
wbb genes (plasmid)

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>wbbE</td>
<td>N-acetylmannosaminotransferase</td>
<td>rfbA</td>
</tr>
<tr>
<td>wbbF</td>
<td>protein of unknown function</td>
<td>rfbB</td>
</tr>
<tr>
<td>wecB</td>
<td>UDP-N-acetylglucosamine-2-epimerase</td>
<td>rfbC</td>
</tr>
</tbody>
</table>

References

1. Keenleyside, W.J. and Whitfield, C.
Lateral transfer of rfb genes: a mobilizable ColE1-type plasmid carries the rfb O:54 (O:54 antigen biosynthesis) gene cluster from Salmonella enterica serovar Borreze.

Serratia marcescens O16 O antigen
\[
3 \rightarrow [\text{D-Galf} \xrightarrow{\beta}{1,3} \text{D-Galp}] \rightarrow [\alpha 1]
\]

\textit{S. marcescens} O16

\textit{wbb} genes

\[
\begin{array}{ccc}
\text{wzm} & \text{wzt} & \text{M} \\
\end{array}
\]

0 \hspace{1cm} \text{kb} \hspace{1cm} 5

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>wbbM</td>
<td>protein of unknown function</td>
<td>rfbF</td>
</tr>
<tr>
<td>wzm</td>
<td>ABC-2 type transport system integral membrane protein</td>
<td>rfbA</td>
</tr>
<tr>
<td>wzt</td>
<td>ABC-2 type transport system ATP-binding protein</td>
<td>rfbB</td>
</tr>
</tbody>
</table>

References

1. Szabo, M., Bronner, D. and Whitfield, C.
Relationships between \textit{rfb} gene clusters required for biosynthesis of identical D-galactose-containing \textit{O} antigens in \textit{Klebsiella pneumoniae} serotype O1 and \textit{Serratia marcescens} serotype O16.

\textit{Pseudomonas aeruginosa} O5 \textit{O} antigen

\[
\begin{array}{c}
\frac{\beta}{4} \left[D-\text{Man}(2\text{NAc3N})A \xrightarrow{\beta}{1,4} D-\text{Man}(2\text{NAc3N})A \xrightarrow{\alpha}{1,3} D-\text{FucNAc}\right] \rightarrow
\end{array}
\]

\textit{P.aeruginosa} O5

\textit{wbp} genes

\[
\begin{array}{cccccccccccccc}
wzz \hspace{1cm} A \hspace{1cm} B \hspace{1cm} C \hspace{1cm} D \hspace{1cm} E \hspace{1cm} \text{wzy} \hspace{1cm} F \hspace{1cm} G \hspace{1cm} H \hspace{1cm} I \hspace{1cm} J \hspace{1cm} K \hspace{1cm} L \hspace{1cm} \text{M} \hspace{1cm} \text{N}
\end{array}
\]

0 \hspace{1cm} \text{kb} \hspace{1cm} 5 \hspace{1cm} 10
<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>wbpA</td>
<td>putative O-antigen biosynthetic protein</td>
<td></td>
</tr>
<tr>
<td>wbpB</td>
<td>putative O-antigen biosynthetic protein</td>
<td></td>
</tr>
<tr>
<td>wbpC</td>
<td>putative O-antigen biosynthetic protein</td>
<td></td>
</tr>
<tr>
<td>wbpD</td>
<td>putative O-antigen biosynthetic protein</td>
<td></td>
</tr>
<tr>
<td>wbpE</td>
<td>putative O-antigen biosynthetic protein</td>
<td></td>
</tr>
<tr>
<td>wbpF</td>
<td>putative O-antigen biosynthetic protein</td>
<td></td>
</tr>
<tr>
<td>wbpG</td>
<td>putative O-antigen biosynthetic protein</td>
<td></td>
</tr>
<tr>
<td>wbpH</td>
<td>putative O-antigen biosynthetic protein</td>
<td></td>
</tr>
<tr>
<td>wbpI</td>
<td>putative O-antigen biosynthetic protein</td>
<td></td>
</tr>
<tr>
<td>wbpJ</td>
<td>putative O-antigen biosynthetic protein</td>
<td></td>
</tr>
<tr>
<td>wbpK</td>
<td>putative O-antigen biosynthetic protein</td>
<td></td>
</tr>
<tr>
<td>wbpL</td>
<td>UDP-GlcNAc transferase</td>
<td>rfbA</td>
</tr>
<tr>
<td>wbpM</td>
<td>putative O-antigen biosynthetic protein</td>
<td></td>
</tr>
<tr>
<td>wbpN</td>
<td>putative O-antigen biosynthetic protein</td>
<td></td>
</tr>
<tr>
<td>wzy</td>
<td>polymerase</td>
<td>rfc</td>
</tr>
<tr>
<td>wzz</td>
<td>flippase</td>
<td></td>
</tr>
</tbody>
</table>

**References**

1. Dasgupta, T. Lam, J.S.
   Identification of rfbA, involved in B-band lipopolysaccharide biosynthesis in *Pseudomonas aeruginosa* serotype O5

2. de Kievit, T.R. Dasgupta, T. Schweizer, H. Lam, J.S.
   Molecular cloning and characterization of the rfc gene of *Pseudomonas aeruginosa* (serotype O5)
   Molecular Microbiology 16: 565-574, 1995.

   Molecular characterization of the *Pseudomonas aeruginosa* serotype O5 (PAO1) B-band lipopolysaccharide gene cluster.
   Molecular Microbiology 22: 481-495.

**Yersinia enterocolitica** O3 O antigen.

\[
2 \left[ 6d-L-\text{Alt} \right] \xrightarrow{\beta} 1
\]
**Y. enterocolitica O3 OAg**

**wbb genes**

Gene | Product Name | Old Names
--- | --- | ---
wbbS | putative 6-dAlt biosynthetic enzyme 1 | rfbA
wbbT | putative glycosyltransferase | rfbB
wbbU | putative glycosyltransferase | rfbC
wbbV | putative 6-dAlt biosynthetic enzyme (2) | rfbF
wbbW | putative 6-dAlt biosynthetic enzyme (3) | rfbG
wbbX | O-antigen biosynthetic protein | rfbH
wzm | ABC-2 type transport system integral membrane protein | rfbD
wzt | ABC-2 type transport system ATP-binding protein | rfbE

**References**


**Yersinia enterocolitica O8 O antigen.**

\[
\begin{align*}
\text{D-Man} \quad & \text{D-Gal} \quad & \text{D-GalNAc} \\
\quad & \alpha \quad & 1 \\
\end{align*}
\]

6d-D-Gul is 6-deoxy-D-gulose
**Gene Cluster of Yersinia enterocolitica O:8**

The genetic characterization of a novel locus of Yersinia enterocolitica serotype O:8 responsible for temperature regulation of O-side chain biosynthesis. 

Genes involved:

- **ddhA** - D-glucose-1-phosphate cytidylyltransferase
- **ddhB** - CDP-glucose 4,6-dehydratase
- **galE** - UDP-glucose 4-epimerase
- **gmd** - GDP-D-mannose 4,6-dehydratase
- **manB** - phosphomannomutase
- **rosA** - putative regulator of O-antigen expression
- **rosB** - putative regulator of O-antigen expression
- **wbcA** - putative 6-deoxygulose synthetase (1)
- **wbcB** - putative 6-deoxygulose synthetase (2)
- **wbcC** - putative 6-deoxygulose transferase
- **wbcD** - putative glycosyltransferase
- **wbcE** - putative O-antigen biosynthetic protein
- **wbcF** - putative O-antigen biosynthetic protein
- **wbcG** - putative glycosyltransferase
- **wbcH** - putative galactoside 2-L-fucosyltransferase
- **wbcI** - putative galactosyltransferase
- **wbcJ** - putative O-antigen biosynthetic protein
- **wzx** - flippase
- **wzy** - polymerase

**References**

1. Zhang, L., Toivanen, P. and Skurnik, M. 
   Genetic characterization of a novel locus of *Yersinia enterocolitica* serotype O:8 responsible for temperature regulation of O-side chain biosynthesis.
   Unpublished.

2. Zhang, L., Toivanen, P. and Skurnik, M. 
   Molecular and chemical characterization of the lipopolysaccharide O-side chain biosynthesis of Yersinia enterocolitica serotype O:8.
   Unpublished.

3. Zhang, L., Toivanen, P. and Skurnik, M. 
   The gene cluster directing O-antigen biosynthesis in Yersinia enterocolitica serotype O:8: identification of the genes for mannose and galactose biosynthesis and the gene for the O-antigen polymerase.
4. Tomshich, S.V., Gorshkova, R.P. and Ovodov, Y.S.
Structural studies on lipopolysaccharide from Y. enterocolitica serovar O:8.

**Yersinia enterocolitica** O3 outer core.

\[
\text{D-GalNAc} \overset{\alpha}{\beta} \text{D-Gal} \overset{\alpha}{1,6} \text{D-GalNAc} \overset{\alpha}{1,4} \text{D-GalNAc} \overset{\beta}{1,3} \text{D-FucNAc} \overset{\beta}{1,3} \text{LD-Hep} \overset{\alpha}{1} \text{D-Glc} \overset{\beta}{1,6} \text{D-Glc} \]

**Y. enterocolitica O3**

*wb* _c_ genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>galE</td>
<td>UDP-glucose 4-epimerase</td>
<td>trsB</td>
</tr>
<tr>
<td>wbcK</td>
<td>putative glycosyltransferase</td>
<td>trsC</td>
</tr>
<tr>
<td>wbcL</td>
<td>putative glycosyltransferase</td>
<td>trsD</td>
</tr>
<tr>
<td>wbcM</td>
<td>putative glycosyltransferase</td>
<td>trsE</td>
</tr>
<tr>
<td>wbcN</td>
<td>putative glycosyltransferase</td>
<td>trsF</td>
</tr>
<tr>
<td>wbcO</td>
<td>putative glycosyltransferase</td>
<td>trsG</td>
</tr>
<tr>
<td>wbcP</td>
<td>putative FucNAc biosynthetic enzyme</td>
<td>trsH</td>
</tr>
<tr>
<td>wbcQ</td>
<td>putative glycosyltransferase</td>
<td>trsA</td>
</tr>
<tr>
<td>wzx</td>
<td>flippase</td>
<td></td>
</tr>
</tbody>
</table>

References

1. Skurnik, M., Venho, R., Toivanen, P. and Alhendy, A.
A novel locus of *Yersinia enterocolitica* serotype O-3 involved in lipopolysaccharide outer core biosynthesis.

2. Shaskov, A.S., Radziejewska-Lebrecht, J., Kochanowski, H. and Mayer, H.
The chemical structure of the outer core region of the Yersinia enterocolitica O:3 lipopolysaccharide.
**Yersinia pseudotuberculosis** IA O antigen

\[
\text{Par} \quad \alpha_\text{1,3} \quad 6d\text{-Hep} \quad \beta_\text{1,4} \\
\downarrow \\
\text{Gal} \quad \alpha_\text{1,3} \quad \text{GlcNAC} \quad \beta_\text{1} \\
\]

**Y. pseudotuberculosis** IA

\[\text{wby genes}\]

<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>ddhA</td>
<td>D-glucose-1-phosphate cytidylyltransferase</td>
<td>rfbF</td>
</tr>
<tr>
<td>ddhB</td>
<td>CDP-glucose 4,6-dehydratase</td>
<td>rfbG</td>
</tr>
<tr>
<td>ddhC</td>
<td>CDP-4-keto-6-deoxy-D-glucose 3-dehydrase</td>
<td>rfbH</td>
</tr>
<tr>
<td>ddhD</td>
<td>CDP-6-deoxy-delta3,4-glucose reductase</td>
<td>rfbI</td>
</tr>
<tr>
<td>prt</td>
<td>paratose synthetase</td>
<td>rfbS</td>
</tr>
<tr>
<td>wbyB</td>
<td>putative O-antigen biosynthetic protein</td>
<td>orf9.8</td>
</tr>
<tr>
<td>wbyC</td>
<td>putative O-antigen biosynthetic protein</td>
<td>orf1</td>
</tr>
<tr>
<td>wbyD</td>
<td>putative O-antigen biosynthetic protein</td>
<td>orf2</td>
</tr>
</tbody>
</table>

**References**

1. Hobbs, M. and Reeves, P.R.
   
   Genetic organisation and evolution of **Yersinia pseudotuberculosis** 3,6-dideoxyhexose biosynthetic genes.
   

**Yersinia pseudotuberculosis** IIA O antigen

\[
\text{Abe}_\text{p} \quad \alpha_\text{1,3} \quad 6d\text{-Hepp} \quad \beta_\text{1,4} \\
\downarrow \\
\text{Gal} \quad \alpha_\text{1,3} \quad \text{GlcNAC}_p \quad \beta_\text{1} \\
\]

37
<table>
<thead>
<tr>
<th>Gene</th>
<th>Product Name</th>
<th>Old Names</th>
</tr>
</thead>
<tbody>
<tr>
<td>abe</td>
<td>CDP-abequose synthase</td>
<td>rfbJ</td>
</tr>
<tr>
<td>ddhA</td>
<td>D-glucose-1-phosphate cytidylyltransferase</td>
<td>rfbF</td>
</tr>
<tr>
<td>ddhB</td>
<td>CDP-glucose 4,6-dehydratase</td>
<td>rfbG</td>
</tr>
<tr>
<td>ddhC</td>
<td>CDP-4-keto-6-deoxy-D-glucose 3-dehydrase</td>
<td>rfbH</td>
</tr>
<tr>
<td>ddhD</td>
<td>CDP-6-deoxy-delta3,4-glucose reductase</td>
<td>rfbI</td>
</tr>
<tr>
<td>wbyA</td>
<td>abequosyltransferase</td>
<td>orf8.7</td>
</tr>
<tr>
<td>wbyB</td>
<td>putative O-antigen biosynthetic protein</td>
<td>orf9.8</td>
</tr>
<tr>
<td>wzx</td>
<td>flippase</td>
<td>orf7.4</td>
</tr>
<tr>
<td>wzx</td>
<td>flippase</td>
<td>rfbX</td>
</tr>
<tr>
<td>wzz</td>
<td>chain length determinant</td>
<td>cld</td>
</tr>
</tbody>
</table>

**References**

1. Kessler, A.C., Haase, A. and Reeves, P.R.  
Molecular analysis of the 3,6-dideoxyhexose pathway genes of *Yersinia pseudotuberculosis* serogroup IIA.  

2. Stevenson, G., Kessler, A. and Reeves P.R.  
A plasmid-borne O-antigen chain length determinant and its relationship to other chain length determinants.  

**Yersinia pseudotuberculosis IVA O antigen**

$$3 \left[ Manp^{1,3} \alpha^1 \right] Manp^{1,2} \beta \left[ Manp^{1,3} \alpha^1 \right] GalNAcp \alpha^1$$

$$\alpha^{1,6}$$
Gene | Product Name | Old Names
---|---|---
ddhA | D-glucose-1-phosphate cytidylyltransferase | rfbF
ddhB | CDP-glucose 4,6-dehydratase | rfbG
ddhC | CDP-4-keto-6-deoxy-D-glucose 3-dehydrase | rfbH
ddhD | CDP-6-deoxy-delta3,4-glucose reductase | rfbI
prt | paratose synthetase | rfbS
tyv | CDP-paratose epimerase | rfbE
wzx | flippase | rfbX

References


Yersinia pseudotuberculosis VA O antigen

\[
\text{Ascp} \\
\text{Manp}_{1,4} \text{Fucp}_{1,3} \text{GalNAcp}_{1} \quad \text{Fucp}_{1,3} \text{Manp}_{1,4} \text{Fucp}_{1,3} \text{GalNAcp}_{1}
\]

Y. pseudotuberculosis VA wby genes
<table>
<thead>
<tr>
<th>Gene</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ascE</td>
<td>CDP-3,6-dideoxy-D-glycero-D-glycero-4-hexulose-5-epimerase</td>
</tr>
<tr>
<td>ascF</td>
<td>CDP-3,6-dideoxy-L-glycero-D-glycero-4-hexulose-4-reductase</td>
</tr>
<tr>
<td>ddhA</td>
<td>D-glucose-1-phosphate cytidylyltransferase</td>
</tr>
<tr>
<td>ddhB</td>
<td>CDP-glucose 4,6-dehydratase</td>
</tr>
<tr>
<td>ddhC</td>
<td>CDP-4-keto-6-deoxy-D-glucose 3-dehydrase</td>
</tr>
<tr>
<td>ddhD</td>
<td>CDP-6-deoxy-delta3,4-glucoseen reductase</td>
</tr>
<tr>
<td>wbyE</td>
<td>putative O-antigen biosynthetic protein</td>
</tr>
</tbody>
</table>

References

1. Thorson, J.S., Lo, S.F., Ploux, O., He, X. and Liu, H.W.
   Studies of the biosynthesis of 3,6-dideoxyhexoses: molecular cloning and characterization of the asc (ascarylose) region from Yersinia pseudotuberculosis serogroup VA.