

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 Klena, 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 (). 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, *wzx* 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. There are a very large number of sugars and potential linkages in bacterial polysaccharides, and it appears that there are many different genes involved. These have been named w^{***} in which * represent any letter. This gives 17,576 names from *waaA* to *wzzZ* of which 676 are committed for the wz^* series. w^{***} names are also given for genes in relevant gene clusters for which there is not information to confidently assign them to the pathway or $wz##$ classes.

There were 30 gene clusters referred to in Table 1 of the Trends paper. In the following sections we give a brief description of them to illustrate the application of the nomenclature in more detail. We also give as an appendix a map and other details of each of the gene clusters.

Application of the scheme to *rfb* genes

The O antigen genes of *Salmonella* and *E. coli* are the ones for which the nomenclature problem is most acute. The new names are of the form wb^* , with the "b" signifying a link to the old name *rfb*, and in many instances the fourth letter remaining unchanged, with for example *rfbV* becoming *wbaV*. Under our proposals *rfbQ* and *rfbR* from *Salmonella* group C2 will become *wbaQ* and *wbaR*, while *rfbQ* and *rfbR* from *E. coli* (*Shigella*) Dysenteriae will become *wbbQ* and *wbbR* (see Fig 3 of the Trends paper). There is also an anomalous name for the galactose transferase of *E. coli* Dysenteriae type 1 O antigen which maps by itself on a plasmid and has been named *rfp* or *rfpA*. This gene can be renamed in the wb^* series and we propose the name *wbbP* (Fig 3). In addition to the O antigen clusters shown in Figs 2 and 3 of the Trends paper, the database includes those of *E. coli* O7, O9 and O111 (incomplete), *Yersinia enterocolitica* O8 and the outer core of O3 (given $wb\#$ names as the structure resembles a single O unit and the genes map between *hemH* and *gsk* as do other O antigen genes of *Yersinia* (Skurnik et al., 1995; Stevenson et al., 1995)).

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 GDP fucose pathway was also identified in the 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 '*wba*' 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 residues for ECA synthesis is dTDP-Fuc4NAc. The pathway for the synthesis of dTDP-Fuc4NAc branches off from the pathway for the synthesis of dTDP-rhamnose (Fig. 1 of Trends paper), and the genes *rmlA*, *rmlB*, and *fcnA* 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 reserved for lipid A and LPS core genes in which the *a* in the second position indicates that it is from the lipopolysaccharide core "*rfa*" group. Thus for example *rfaG*, encoding a glucosyl transferase becomes *waaG*, the *G* representing the *G* in its current name. *kdtA*, 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-glyceromannoheptose pathway genes have been allocated the names *gmhA-D* (Fig 1) of which *gmhA* and *gmhB* (were *lpcA* and *lpcB* 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 *wlb*. The structure of *Bordetella pertussis* LPS resembles a core with only one O unit. This is similar to the LPS structures sometimes called lipooligosaccharides (LOS) and the name *wlb* includes 'l' for LOS and 'b' for *Bordetella*. The cluster also includes *waaA* and *waaC* genes and a gene, *wlbG*, 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 BPGN scheme (Burrows et al., 1996). Most are given *wbp* names as only *wzy* 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 BPGN 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-glyceromannoheptose 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

- Allen, A. G. & Maskell, D. J. (1996). The identification, cloning and mutagenesis of a genetic locus required for lipopolysaccharide biosynthesis in *Bordetella pertussis*. *Molecular Microbiology* **19**, 37-52.
- Barr, K. & Rick, P. D. (1993). Physical map location of the *rffC* and *rffA* genes of *Escherichia coli*. *Journal of Bacteriology* **175**, 5738-5739.
- Bugert, P. & Geider, K. (1995). Molecular analysis of the *ams* operon required for exopolysaccharide synthesis of *Erwinia amylovora*. *Molecular Microbiology* **15**, 917-933.
- Burrows, L. L., Charter, D. F. & Lam, J. S. (1996). Molecular characterization of the *Pseudomonas aeruginosa* serotype O5 (PA01) B-band lipopolysaccharide gene cluster. *Molecular Microbiology* **22**, in press.
- Coplin, D. L., Majerczak, D. R., Bugert, P. & Geider, K. (1995). 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. *Acta Hort* **411**, 251-255.
- Daniels, D. L., Plunkett III, G., Burland, V. & Blattner, F. R. (1992). Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes. *Science* **257**, 771-778.
- Keenleyside, W. J., Jayaratne, P., MacLachlan, P. R. & Whitfield, C. (1992). The *rcaA* gene of *Escherichia coli* O9:K30:H12 is involved in the expression of the serotype-specific group I K (capsular) antigen. *Journal of Bacteriology* **174**, 8-16.
- Lee, S. J., Romana, L. K. & Reeves, P. R. (1992). Sequence and structural analysis of the *rfb* (O antigen) gene cluster from a group C1 *Salmonella enterica* strain. *Journal of General Microbiology* **138**, 1843-1855.
- Marolda, C. L. & Valvano, M. A. (1995). Genetic analysis of the dTDP-rhamnose biosynthesis region of the *Escherichia coli* VW187 (O7/K1) *rfb* gene cluster - identification of functional homologs of *rfbB* and *rfbA* in the *rff* cluster and correct location of the *rffE* gene. *Journal of Bacteriology* **177**, 5539-5546.
- Meier-Dieter, U., Barr, K., Starman, R., Hatch, L. & Rick, P. D. (1992). Nucleotide sequence of the *Escherichia coli rfe* gene involved in the synthesis of enterobacterial common antigen. *Journal of Biological Chemistry* **267**, 746-753.
- Meier-Dieter, U., Starman, R., Barr, K., Mayer, H. & Rick, P. D. (1990). Biosynthesis of Enterobacterial common antigen in *Escherichia coli*. *Journal of Biological Chemistry* **265**, 13490-13497.
- Raetz, C. R. H. (1996). Bacterial lipopolysaccharides: a remarkable family of bioactive macroamphiphiles. In *Escherichia and Salmonella typhimurium: Cellular and Molecular Biology 2nd edition*, Vol. 1, pp. 1035 - 1063. Edited by F. D. Neidhardt. Washington, D. C.: American Society for Microbiology.
- Reeves, P. R., Hobbs, M., Valvano, M., Skurnik, M., Whitfield, C., Coplin, D., Kido, N., Klena, J., Maskell, D., Raetz, C. & Rick, P. (1996). Bacterial polysaccharide synthesis and gene nomenclature. *Trends in Microbiology* **4**, 495-503.
- Skurnik, M., Venho, R., Toivanen, P. & Alhendy, A. (1995). A novel locus of *Yersinia enterocolitica* serotype O-3 involved in lipopolysaccharide outer core biosynthesis. *Molecular Microbiology* **17**, 575-594.
- Stevenson, G., Andrianopoulos, K., Hobbs, H. & Reeves, P. R. (1996). Organization of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid. *Journal of Bacteriology* **178**, 4885-4893.
- Stevenson, G. S., Kessler, A. & Reeves, P. R. (1995). A plasmid-borne O-antigen chain length determinant and its relationship to other chain length determinants. *FEMS Microbiology Letters* **125**, 23-30.

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@btk.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.

<i>Bordetella pertussis</i>	LPS
<i>Erwinia stewartii</i>	stewartan (Extracellular polysaccharide)
<i>Escherichia coli</i>	K-12 colanic acid
<i>Escherichia coli</i>	ECA
<i>Escherichia coli</i>	K-12 lipid A/core
<i>Escherichia coli</i> (Shigella)	Dysenteriae type I O antigen
<i>Escherichia coli</i>	O9 O antigen
<i>Escherichia coli</i>	O16 (K-12) O antigen
<i>Escherichia coli</i>	O111 O antigen
<i>Escherichia coli</i>	O7 O antigen
<i>Klebsiella pneumoniae</i>	O1 O antigen
<i>Klebsiella pneumoniae</i>	O8 O antigen
<i>Salmonella enterica</i>	LT2 colanic acid

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

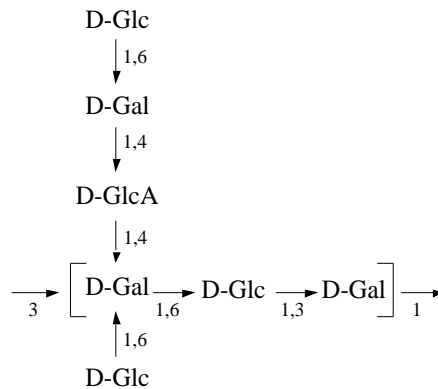
References

1. Allen, A.G and Maskell, D.J.

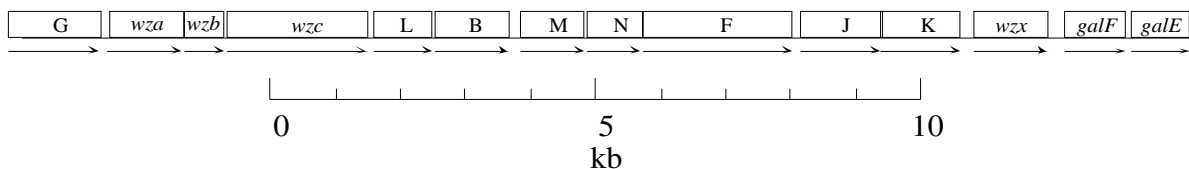
The identification, cloning and mutagenesis of a genetic locus required for lipopolysaccharide biosynthesis in *Bordetella pertussis*.

Molecular Microbiology 19:37-52, 1996.

Erwinia stewartii stewartan (Extracellular polysaccharide)



E. stewartii *wce* genes

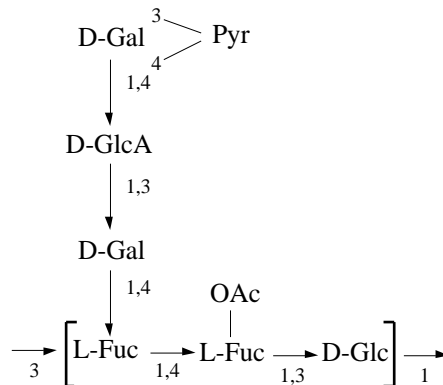


Gene	Product Name	Old Names
galE	UDP-glucose 4-epimerase	
galF	galactose-1-phosphate uridylyltransferase (regulatory subunit?)	cpsM
wceB	putative EPS biosynthetic protein	cpsE
wceF	putative EPS biosynthetic protein	cpsH
wceG	putative EPS biosynthetic protein	cpsA
wceJ	putative EPS biosynthetic protein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

References

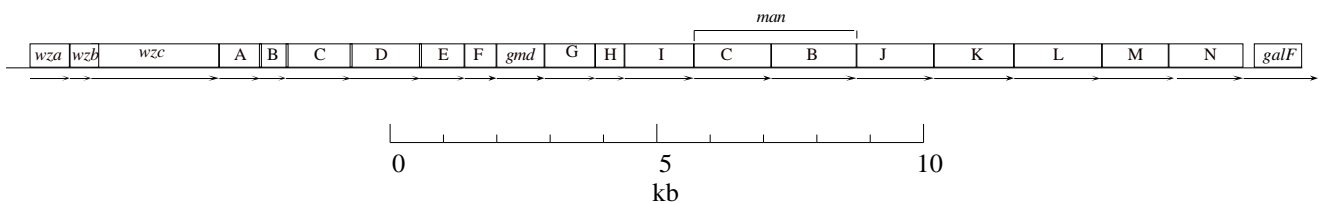
1. Coplin, D.L. Majerczak, D.R. Bugert, P. Geider, K.
Nucleotide sequence analysis of the *Erwinia stewartii* cps gene cluster for synthesis of stewartan and comparison to the *Erwinia amylovora* ams cluster for synthesis of amylovoran.
Acta Hort. 411: 251-255, 1995.
2. Bernhard, F. Schullerus, D. Belleman, P. Nimtz, M. Coplin, D.L. Geider, K.
Genetic transfer of amylovoran and stewartan synthesis between *Erwinia amylovora* and *Erwinia stewartii*.
Microbiology 42: 1087-1096, 1996.
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*.
Carbohydrate Research 288: 189-201, 1996.

Escherichia coli K-12 colanic acid



Pyr is pyruvate linked acetalically to galactose

E. coli K-12 colanic acid
wca genes



galF galactase-1-phosphate uridylyltransferase (regulatory ~~subunit?~~)
gmd GDP-D-mannose 4,6-dehydratase orfO.O

manB	phosphomannomutase	cpsG
manB	phosphomannomutase	pgm#
manC	D-mannose-1-phosphate guanylyltransferase	cpsB
manC	D-mannose-1-phosphate guanylyltransferase	pmi #
wcaA	putative colanic acid glycosyltransferase	
wcaB	colanic acid O-acetyltransferase II	
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	orf0.9
wcaH	putative colanic acid biosynthetic enzyme	orf1.9
wcaI	putative colanic acid fucosyltransferase	orf2.4
wcaJ	putative undecaprenylphosphate glucosephosphotransferase	
wcaK	putative colanic acid biosynthetic enzyme	
wcaL	putative colanic acid glycosyltransferase	orf0.0
wcaM	putative colanic acid biosynthetic enzyme	orf1.3
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

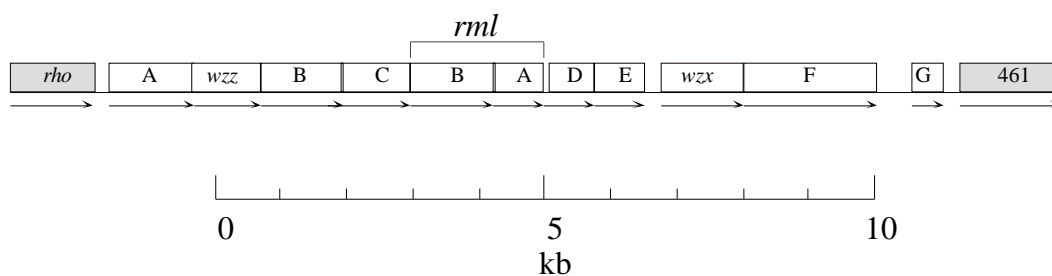
1. Stevenson, G., Andrianopoulos, K., Hobbs, M. and Reeves, P.R.
 Organisation of the *Escherichia coli* K-12 gene cluster responsible for production of the extracellular polysaccharide colanic acid.
 Journal of Bacteriology 178: 4885-4893, 1966.

Escherichia coli ECA



ECA has also been reported to be a cyclical structure

E. coli K-12 ECA wec genes

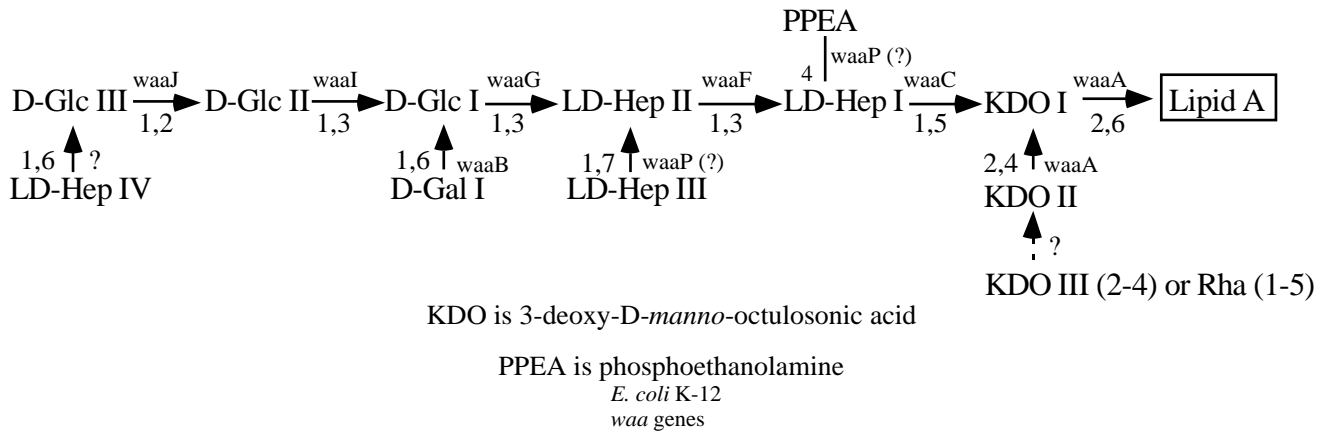


Gene	Product Name	Old Names
rmIA	glucose-1-phosphate thymidyltransferase	o292
rmIB	dtDP-D-glucose 4,6-dehydratase	o355
wecA	UDP-GlcNAc:undecaprenylphosphate GlcNAc-1-phosphate transferase	rfe
wecB	UDP-N-acetylglucosamine-2-epimerase	nfrC
wecB	UDp-N-acetylglucosamine-2-epimerase	o389
wecC	ECA biosynthetic protein	o379
wecD	ECA biosynthetic protein	o181
wecE	ECA biosynthetic protein	o299
wecF	4-amino-4,6-dideoxy-D-galactosyltransferase	rffT
wecG	lipopolysaccharide N-acetylmannosaminuronosyltransferase	o246
wzx	flippase	o416
wzz	chain length determinant	o349

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. Molecular cloning of the *rfe-rff* gene cluster.
Journal of Biological Chemistry 267: 746-753, 1992.
2. Daniels, D.L., Plunkett, G. 3d., Burland, V. and Blattner F.R.
Analysis of the *Escherichia coli* genome: DNA sequence of the region from 84.5 to 86.5 minutes.
Science 257: 771-778, 1992.
3. Kiino, D.R., Licudine, R., Wilt, K., Yang, D.H. and Rothman-Denes, L.B.
A cytoplasmic protein, NfrC, is required for bacteriophage N4 adsorption.
Journal of Bacteriology 175: 7074-7080, 1993 .
4. Marolda, C.L. and Valvano, M.A.
Genetic analysis of the dTDP-rhamnose biosynthesis region of the *Escherichia coli* VE187 (O7:K1) *rfb* gene cluster: identification of functional homologs of *rfbB* and *rfbA* in the *rff* cluster and correct location of the *rffE* gene.
Journal of Bacteriology 177: 5539-5546, 1995.

Escherichia coli K-12 lipid A/core



Gene	Product Name	Old Names
gmhD	ADP-heptose 6-epimerase	rfaD
waaA	KDO transferase (inner core)	kdtA
waaB	galactosyltransferase (outer core)	rfaB
waaC	heptosyltransferase I (inner core)	rfaC
waaF	heptosyltransferase II (inner core)	rfaF
waaG	glucosyltransferase I (outer core)	rfaG
waal	glycosyltransferase	rfaI
waaJ	glucosyltransferase II (outer core)	rfaJ
waaK	protein of unknown function	rfaK
waaL	O-antigen ligase	rfaL
waaP	protein of unknown function	rfaP
waaQ	protein of unknown function	rfaQ
waaS	protein of unknown function	rfaS
waaY	protein of unknown function	rfaY
waaZ	protein of unknown function	rfaZ

References

1. Pegues, J.C., Chen, L.S., Gordon, A.W., Ding, L. and Coleman, W.G. Jr. Cloning, expression and characterisation of the *Escherichia coli* K-12 *rfaD* gene. *Journal of Bacteriology* 172: 4652-4660, 1990.
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. *Journal of Biological Chemistry* 266: 9687-9696, 1991.
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.

Journal of Bacteriology 174: 930-934, 1992.

4. Pradel, E., Parker, C.T. and Schnaitman, C.A.

Structures of the *rfaB*, *rfaJ* and *rfaS* genes of *Escherichia coli* K-12 and their roles in assembly of the lipopolysaccharide core.

Journal of Bacteriology 174: 4736-4745, 1992.

5. Klena, J.D., Pradel, E. and Schnaitman, C.A.

Comparison of lipopolysaccharide biosynthesis genes *rfaK*, *rfaL*, *rfaY* and *rfaZ* of *Escherichia coli* K-12 and *Salmonella typhimurium*.

Journal of Bacteriology 174: 4746-4752, 1992.

6. Clementz, T.

The gene coding for 3-deoxy-manno-octulosonic acid transferase and the *rfaQ* gene are transcribed from divergently arranged promoters in *Escherichia coli*.

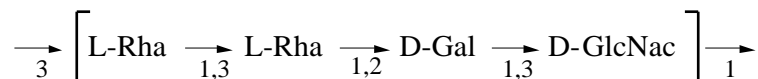
Journal of Bacteriology 174: 7750-7756, 1992.

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 *rfa* Loci of *Escherichia coli* K-12 and *Salmonella typhimurium*

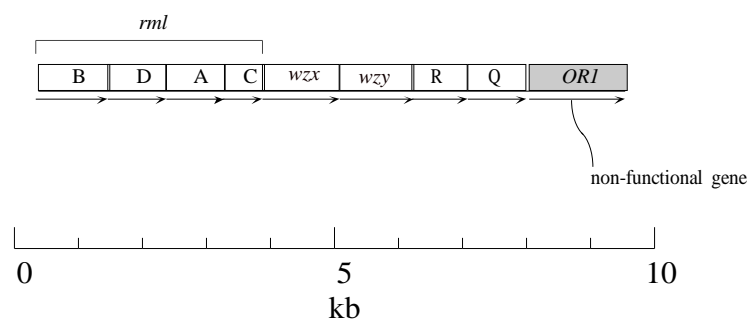
Journal of Bacteriology 173: 7410-7411, 1991.

Escherichia coli (Shigella) Dysenteriae type I O antigen



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

Sh. dysenteriae *wbb* genes (chromosome)



Gene

rmlA
rmlB
rmlC

Product Name

glucose-1-phosphate thymidyltransferase
dTDP-D-glucose 4,6-dehydratase
dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase

Old Name

rfbA
rfbB
rfbC

mlD
wbbQ
wbbR
wzx
wzy

dTDP-4-keto-L-rhamnose reductase
 rhamnosyltransferase I
 rhamnosyltransferase II
 flippase
 polymerase

rfbD
rfbQ
rfbR
rfbX
rfa

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: structure and functions of the *rfb* cluster.

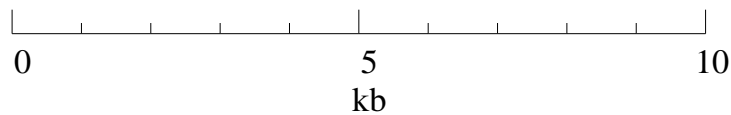
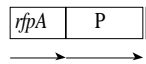
Microbial Pathogenesis 1: 307-324, 1986.

2. Klena JD. Schnaitman CA.

Function of the *rfb* gene cluster and the *rfa* gene in the synthesis of O antigen by *Shigella dysenteriae* 1.

Molecular Microbiology 9: 393-402, 1993.

Sh. dysenteriae pHW400 *wbb* genes (plasmid)



Gene
wbbP
wbbP

Product Name
 galactosyltransferase
 galactosyltransferase

Old Name
rfa
rfaB

References

1. Klena J.D. Ashford R.S. 2d. Schnaitman C.A.

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

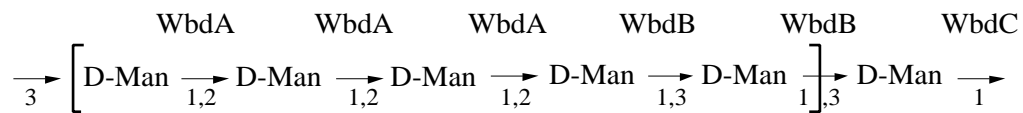
Journal of Bacteriology 174: 7297-7307, 1992.

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 *rfa* determinant.

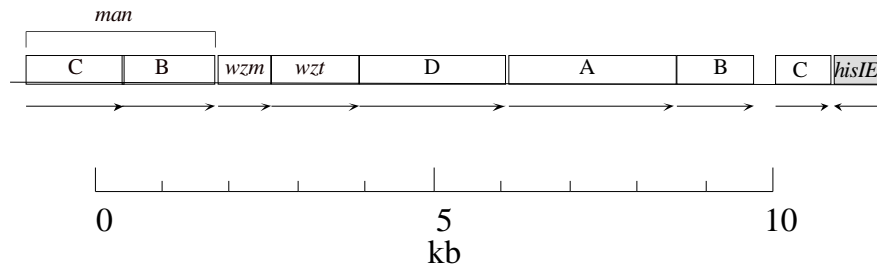
Microbial Pathogenesis 16 53-64, 1994.

Escherichia coli O9 O antigen



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

E. coli O9 *wbd* genes

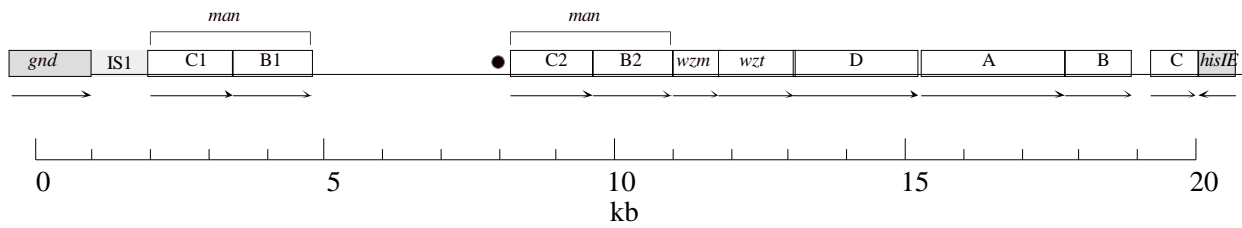


Gene	Product Name	Old Name
<i>manB</i>	phosphomannomutase	<i>rfbK</i>
<i>manC</i>	D-mannose-1-phosphate guanylyltransferase	<i>rfbM</i>
<i>wbdA</i>	mannosyltransferase I	<i>mtfA</i>
<i>wbdB</i>	mannosyltransferase II	<i>mtfB</i>
<i>wbdC</i>	mannosyltransferase III	<i>mtfC</i>
<i>wbdD</i>	O-antigen biosynthetic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

References

1. Sugiyama T. Kido N. Komatsu T. Ohta M. Jann K. Jann B. Saeki A. Kato N.
Genetic analysis of *Escherichia coli* O9 *rfb*: identification and DNA sequence of phosphomannomutase and GDP-mannose pyrophosphorylase genes.
Microbiology. 140: 59-71, 1994.
2. Kido N. Torgov V.I. Sugiyama T. Uchiya K. Sugihara H. Komatsu T. Kato N. Jann K.
Expression of the O9 polysaccharide of *Escherichia coli*: sequencing of the *E.coli* O9 *rfb* gene cluster, characterisation of mannosyl transferases, and evidence for an ATP-binding cassette transport system.
Journal of Bacteriology 177: 2178-2187, 1995.

E. coli O9
wbd genes

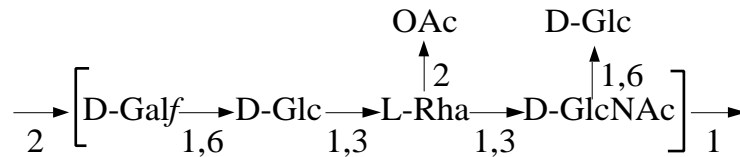


Gene	Product Name	Old Name
<i>manB</i>	phosphomannomutase	<i>rfbK1</i>
<i>manB</i>	phosphomannomutase	<i>rfbK2</i>
<i>manC</i>	D-mannose-1-phosphate guanylyltransferase	<i>rfbM1</i>
<i>manC</i>	D-mannose-1-phosphate guanylyltransferase	<i>rfbM2</i>

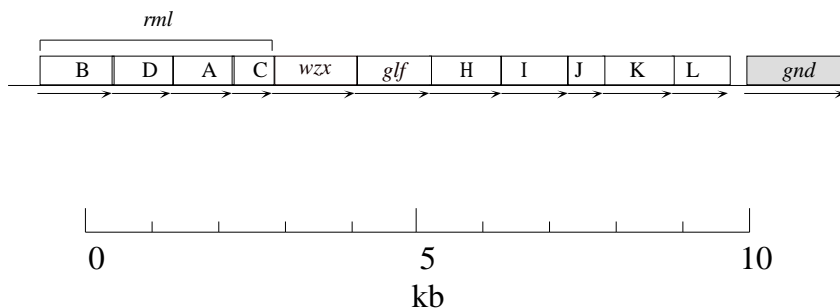
References

- 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. *Journal of Bacteriology*. 176: 3126-39, 1994.

Escherichia coli O16 (K-12) O antigen



E. coli K-12 O16 O Ag
wbb genes



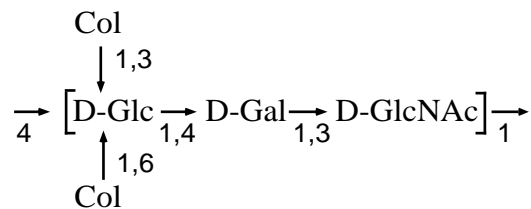
Gene	Product Name	Old Names
<i>glf</i>	UDP-galactopyranose mutase	<i>orf6</i>
<i>rmlA</i>	glucose-1-phosphate thymidyltransferase	<i>rfbA</i>

rmIB	dTDP-D-glucose 4,6-dehydratase	rfbB
rmIC	dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase	rfbC
rmID	dTDP-4-keto-L-rhamnose reductase	rfbD
wbbH	putative O-unit polymerase	orf4(Yao)
wbbH	putative P-unit polymerase	rfc (Stevenson)
wbbI	galactofuranosyltransferase	orf3(Yao)
wbbI	galactofuranosyltransferase	orf8(Stevenson)
wbbJ	acetyltransferase	orf2(Yao)
wbbJ	acetyltransferase	orf8(Stevenson)
wbbK	glucosyltransferase	ORF224 (Liu)
wbbK	glucosyltransferase	orf1(Yao)
wbbK	glucosyltransferase	orf10(Stevenson)
wbbL	rhamnosyltransferase	ORF264 (Liu)
wbbL	rhamnosyltransferase	orf5(Yao)
wzx	flippase	rfbx

References

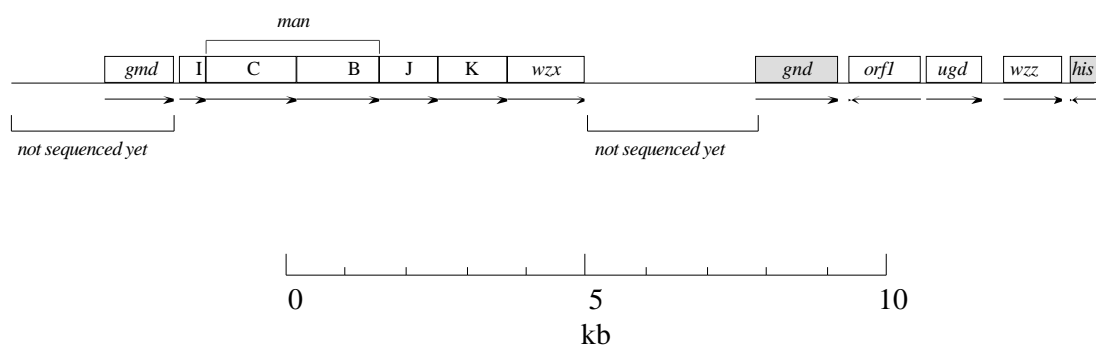
1. Liu, D. and Reeves, P.R.
Escherichia coli K12 regains its O antigen.
Microbiology 140: 49-57, 1994.
2. Yao, Z. and Valvano, M.A.
Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (*rfb*) of *Escherichia coli* K-12 W3110: identification of genes that confer group 6 specificity to *Shigella flexneri* serotypes Y and 4a.
Journal of Bacteriology. 176: 4133-43, 1994.
3. Stevenson, G., Neal, B., Hobbs, M., Packer, N.H., Batley, M., Redmond, J.W., Lindquist, L. and Reeves, P.R.
Structure of the O antigen of *Escherichia coli* K-12 and the sequence of its *rfb* gene cluster.
Journal of Bacteriology. 176: 4144-56, 1994.

Escherichia coli O111 O antigen



Col is 3,6-dideoxy-L-galactose

E. coli O111
wbd genes

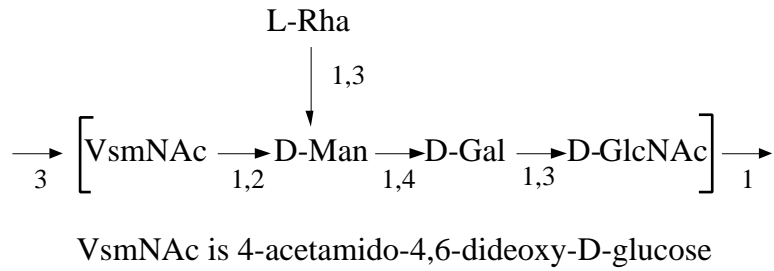


Gene	Product Name	Old Names
gmd	GDP-D-mannose 4,6-dehydratase	orf0.0
manB	phosphomannomutase	rfbK
manC	D-mannose-1-phosphate guanylyltransferase	rfbM
ugd	UDP-glucose 6-dehydrogenase	orf1
wbdI	putative O-antigen biosynthetic protein	orf3.4
wbdJ	putative O-antigen biosynthetic protein	orf6.7
wbdK	putative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

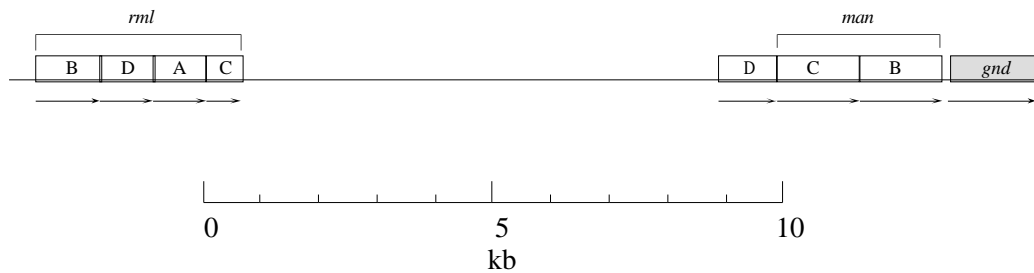
References

1. Bastin, D.A. and Reeves, P.R.
Sequence and analysis of the O antigen gene (*rfb*) cluster of *Escherichia coli* O111.
Gene 164: 17-23, 1995.
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



E. coli O7
wbb genes

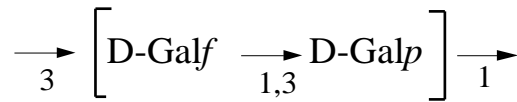


Gene	Product Name	Old Names
manB	phosphomannomutase	rfbK
manC	D-mannose-1-phosphate guanylyltransferase	rfbM
rmlA	glucose-1-phosphate thymidyltransferase	rfbP
rmlB	dTDP-D-glucose 4,6-dehydratase	rfbB
rmlC	dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase	rfbC
rmlD	dTDP-4-keto-L-rhamnose reductase	rfbD
wbbD	putative O-antigen biosynthetic protein	orf275

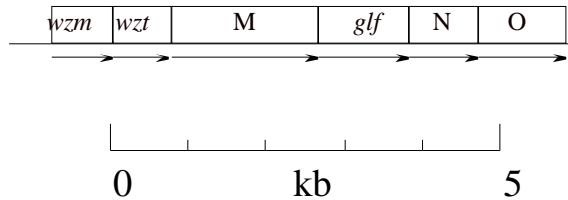
References

1. Marolda, C.L. and Valvano, M.A.
Identification, expression and DNA sequence of the GDP-mannose biosynthesis genes encoded by the O7 *rfb* gene cluster of strain VW187 (*Escherichia coli* O7:K1).
Journal of Bacteriology. 175: 148-158, 1993.
2. Marolda, C.L. and Valvano, M.A.
Genetic analysis of the dTDP-rhamnose biosynthesis region of the *Escherichia coli* VW187 (O7:K1) *rfb* gene cluster: identification of functional homologs of *rfbB* and *rfbA* in the *rff* cluster and correct location of the *rffE* gene.
Journal of Bacteriology. 177: 5539-5546, 1995.

Klebsiella pneumoniae O1 O antigen



K. pneumoniae O1
wbb genes

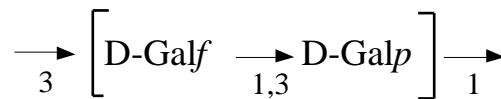


Gene	Product Name	Old Names
<i>glf</i>	UDP-galactopyranose mutase	rfbD
<i>wbbM</i>	protein of unknown function	rfbC
<i>wbbN</i>	protein of unknown function	rfbE
<i>wbbO</i>	galactosyltransferase	rfbF
<i>wzm</i>	ABC-2 type transport system integral membrane protein	rfbA
<i>wzt</i>	ABC-2 type transport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

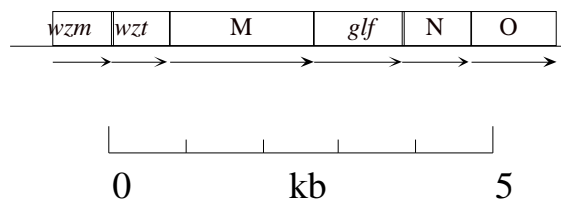
References

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 *Klebsiella pneumoniae* serotype O1.
 Molecular Microbiology 14: 505-19, 1994.

Klebsiella pneumoniae O8 O antigen



K. pneumoniae O8
wbb genes

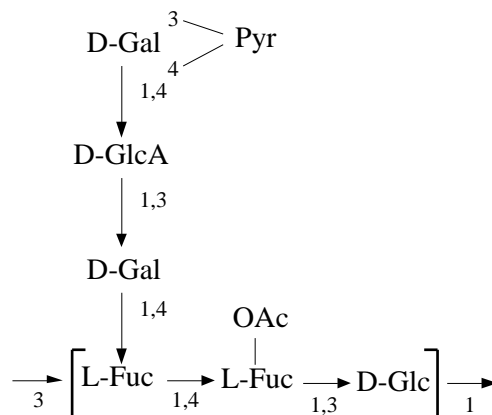


Gene	Product Name	Old Names
glf	UDP-galactopyranose mutase	rfbD
wbbM	protein of unknown function	rfbC
wbbN	protein of unknown function	rfbE
wbbO	galactosyltransferase	rfbF
wzm	ABC-2 transport system integral membrane protein	rfaA
wzt	ABC-2 type transport system ATP-binding protein	rfaB

References

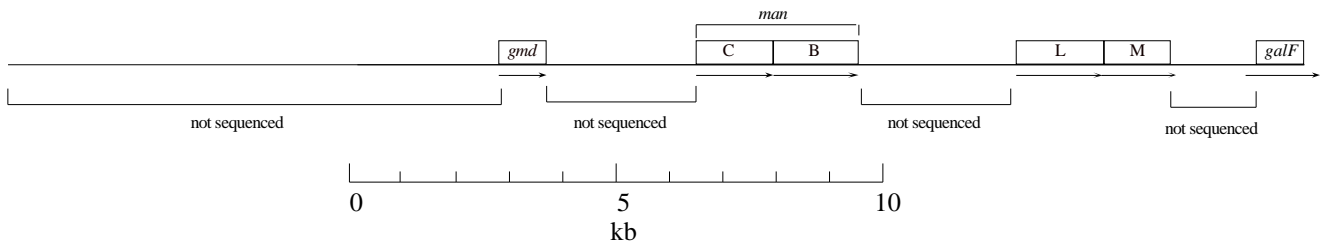
1. Kelly, R.F., Severn, W.B., Richards, J.C., Perry, M.B., Maclean, L.L., Tomas, J.M., Merino, S. and Whitfield, C. Structural variation in the O-specific polysaccharides of *Klebsiella pneumoniae* serotype 01 and 08 lipopolysaccharide: evidence for clonal diversity in *rfb* genes. *Molecular Microbiology* 10: 615-25, 1993.
2. Szabo, M., Bronner, D. and Whitfield, C. Relationships between *rfb* gene clusters required for biosynthesis of identical D-galactose-containing O antigens in *Klebsiella pneumoniae* serotype 01 and *Serratia marcescens* serotype 016. *Journal of Bacteriology*. 177: 1544-1553, 1995.

Salmonella enterica LT2 colanic acid



Pyr is pyruvate linked acetalically to galactose

S. enterica sv. Typhimuriam
colanic acid
wca genes

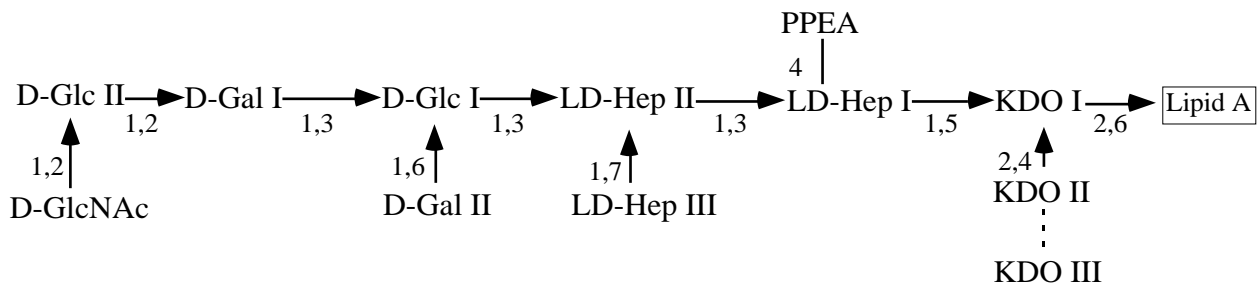


Gene	Product Name	Old Names
galF	galactose-1-phosphate uridylyltransferase (regulatory subunit?)	orf2.8
gmd	GDP-D-mannose 4,6-dehydratase	
manB	phosphomannomutase	cpsG
manC	D-mannose-1-phosphate guanylyltransferase	cpsB
wcaL	putative colanic acid glycosyltransferase	orfO.0
wcaM	putative colanic acid biosynthetic enzyme	orf1.3

References

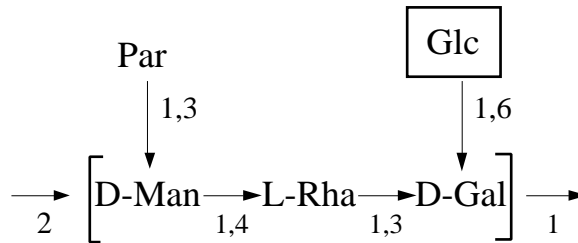
- Jiang, X.M., Neal, B., Santiago, R., Lee, S.J., Romana, L.K. and Reeves, P.R.
Structure and sequence of the *rfb* (O antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2).
Molecular Microbiology 5: 695-713, 1991.
- Stevenson, G., Lee, S.J., Romana, L.K. and Reeves, P.R.
The *cps* gene cluster of *Salmonella* strain LT2 includes a second mannose pathway: sequence of two genes and relationship to genes in the *rfb* gene cluster.
Molecular & General Genetics 227: 173-180, 1991.

Salmonella enterica LT2 lipid A/core

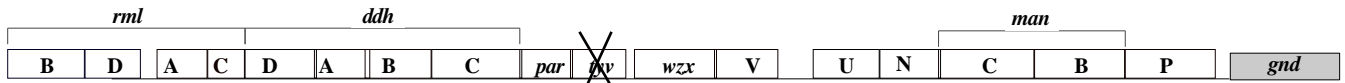


KDO is 3-deoxy-D-manno-octulosonic acid

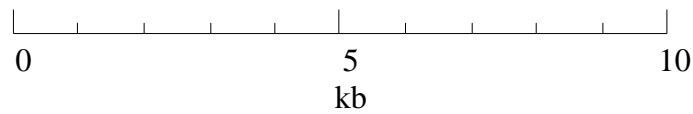
PPEA is phosphoethanolamine



S. enterica sv. Paratyphi
wba genes



Gene cluster same as Se D1 O-Ag except *tyv* has a frame shift mutation

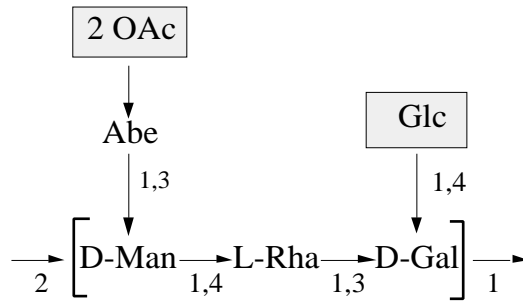


Gene	Product Name	Old Names
<i>ddhA</i>	D-glucose-1-phosphate cytidyltransferase	<i>rfbF</i>
<i>ddhB</i>	CDP-glucose 4,6-dehydratase	<i>rfbG</i>
<i>ddhC</i>	CDP-4-keto-6-deoxy-D-glucose 3-dehydrase	<i>rfbH</i>
<i>ddhD</i>	CDP-6-deoxy-delta3,4-glucoseen reductase	<i>rfbI</i>
<i>manB</i>	phosphomannomutase	<i>rfbK</i>
<i>manC</i>	D-mannose-1-phosphate guanylyltransferase	<i>rfbM</i>
<i>prt</i>	paratose synthetase	<i>rfbS</i>
<i>rmlA</i>	glucose-1-phosphate thymidyltransferase	<i>rfbA</i>
<i>rmlB</i>	dTDP-D-glucose 4,6-dehydratase	<i>rfbB</i>
<i>rmlC</i>	dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase	<i>rfbC</i>
<i>rmlD</i>	dTDP-4-keto-L-rhamnose reductase	<i>rfbD</i>
<i>wbaN</i>	thamnosyltransferase	<i>rfbN</i>
<i>wbaP</i>	undecaprenylphosphate galactosephosphotransferase	<i>rfbP</i>
<i>wbaU</i>	mannosyltransferase	<i>rfbU</i>
<i>wbaV</i>	dideoxyhexosyltransferase	<i>rfbV</i>
<i>wzx</i>	flippase	<i>rfbX</i>

References

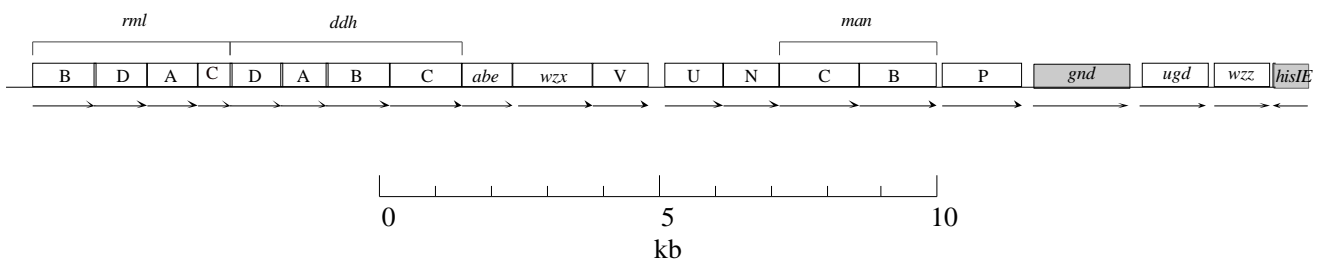
- Liu, D., Verma, N.K., Romana, L.K. and Reeves, P.R.
Relationships among the *rfb* regions of *Salmonella* serovars A, B and D.
Journal of Bacteriology 173: 4814-4819, 1991.

Salmonella enterica group B O antigen



Abe is 3,6-dideoxy-D-galactose

S. enterica sv. Typhimurium
wba genes



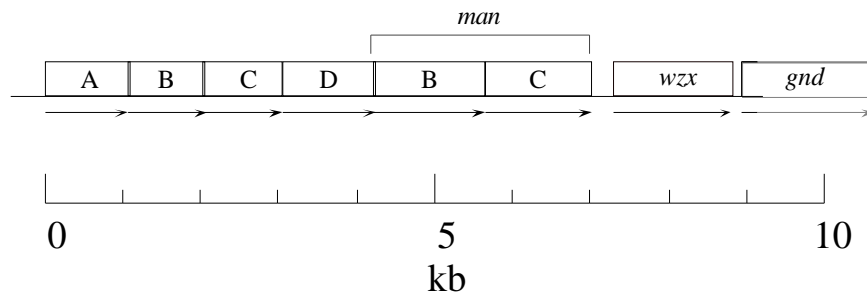
Gene	Product Name	Old Names
ddhA	D-glucose-1-phosphate cytidyltransferase	rfbF
ddhB	CDP-glucose 4,6-dehydratase	rfbG
ddhC	CDP-4-keto-6-deoxy-D-glucose 3-dehydrase	rfbH
ddhD	CDP-6-deoxy-delta3,4-glucoseen reductase	orf7.6
ddhD	CDP-6-deoxy-delta3,4-glucoseen reductase	rfbI
manB	phosphomannomutase	rfbK
manC	D-mannose-1-phosphate guanylyltransferase	rfbM
rm1A	glucose-1-phosphate thymidyltransferase	rfbA
rm1B	dTDP-D-glucose 4,6-dehydratase	rfbB
rm1C	dTDP-4-keto-6-deoxy-D-glucose 3,5-epimerase	rfbC
rm1D	dTDP-4-keto-L-rhamnose reductase	rfbD
ugd	UDP-glucose 6-dehydrogenase	orf1
wbaN	rhamnosyltransferase	rfbN
wbaP	undecaprenylphosphate galactosephosphotransferase	rfbP
wbaU	mannosyltransferase	rfbU
wbaV	dideoxyhexosyltransferase	rfbV
wzx	flippase	rfbX
wzz	chain length determinant	clb

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: homolgy with galactose epimerase.

S. enterica sv. Montevideo
wba genes

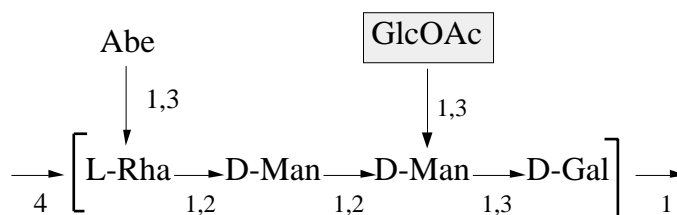


Gene	Product Name	Old Names
galF	galactose-1-phosphate uridylyltransferase regulatory subunit?)	orf2.8
manB	phosphomannomutase	orf9.79
manC	D-mannose-1-phosphate guanylyltransferase	orf8.36
wbaA	possible O-antigen polymerase	orf4.11
wbaB	possible mannosyltransferase I	orf5.19
wbaC	possible mannosyltransferase II	orf6.17
wbaD	possible mannosyltransferase III	orf7.17
wzx	flippase	rfbX

References

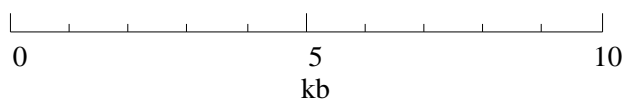
- Lee, S.J., Romano, L.K. and Reeves, P.R.
Cloning and structure of group C1 O antigen (*rfb* gene cluster) from *Salmonella enterica* serovar montevideo.
Journal of General Microbiology 138 : 305-312, 1992.
- Lee, S.J., Romana, L.K. and Reeves, P.R.
Sequence and structural analysis of the *rfb* (O antigen) gene cluster from a group C1 *Salmonella enterica* strain.
Journal of General Microbiology. 138: 1843-1455, 1992.
- Reeves, P.
Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. [Review].
Trends in Genetics. 9: 17-22, 1993 .
- Hobbs, M. and Reeves, P.R.
The JUMPstart sequence: a 39 bp element common to several polysaccharide gene clusters.
Molecular Microbiology 12: 855-6, 1994.

Salmonella enterica group C2 O antigen



Abe is 3,6-dideoxy-D-galactose

S. enterica sv. Muenchen wba genes



Gene	Product Name	Old Names
abe	CDP-abequose synthase	rfbJ
ddhA	D-glucose-1-phosphate cytidyltransferase	rfbF
ddhB	CDP-glucose 4,6-dehydratase	rfbG
ddhC	CDP-4-keto-6-D-glucose 3-dehydrase	rfbH
ddhD	CDP-6-deoxy-delta3,4-glucoseen reductase	rfbI
manB	phosphomannomutase	rfbK
manC	D-mannose-1-phosphate guanylyltransferase	rfbM
rmlA	glucose-1-phosphate thymidyltransferase	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
wbaL	O-acetyltransferase	orf14.9
wbaL	O-acetyltransferase	rfbL
wbaP	undecaprenylphosphate galactosephosphotransferase	rfbP
wbaQ	rhamnosyltransferase	orf15.6
wbaQ	rhamnosyltransferase	rfbQ
wbaR	abequosyltransferase	orf13.9
wbaR	abequosyltransferase	rfbR
wbaW	mannosyltransferase II	orf17.9
wbaW	mannosyltransferase II	rfbW
wbaZ	mannosyltransferase I	orf18.9
wbaZ	mannosyltransferase I	rfbZ
wzx	flippase	orf12.6
wzx	flippase	rfbX
wzy	polymerase	rfc

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.

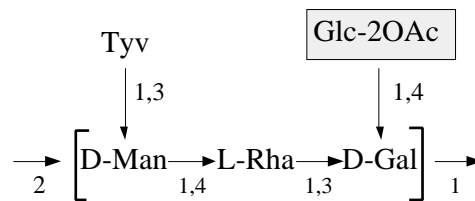
Molecular Microbiology. 6: 1385-1394, 1992.

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.

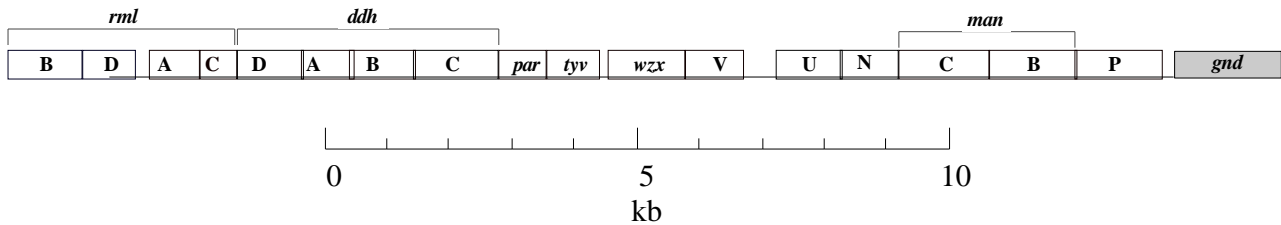
Journal of Bacteriology. 175: 3408-1343, 1993.

Salmonella enterica group D1O antigen



Tyv is 3,6-dideoxy-D-mannose

S. enterica sv. Typhi *wba* genes



Gene	Product Name	Old Names
ddhA	D-glucose-1-phosphate cytidyltransferase	rfbF
ddhB	CDP-glucose 4,6-dehydratase	rfbG
ddhC	CDP-4-keto-6-deoxy-D-glucose 3-hydrase	rfbG
ddhD	CDP-6-deoxy-delta3,4-glucoseen reductase	rfbI
manB	phosphomannomutase	rfbK
manC	D-mannose-1-phosphate guanylyltransferase	rfbM
prt	paratose synthetase	rfbS
rmlA	glucose-1-phosphate guanylyltransferase	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
wbaP	undecaprenylphosphate galactosephosphotransferase	rfbP
wbaU	mannosyltransferase	orf14.1
wbaV	dideoxyhexosyltransferase	rfbV
wzx	flippase	orf12.8
wzx	flippase	rfbX

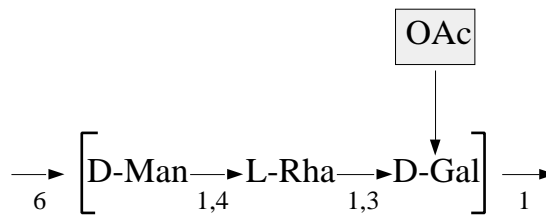
1. Reeves, P.

Evolution of Salmonella O antigen variation by interspecific gene transfer on a large scale. [Review].
Trends in Genetics 9 17-22, 1993.

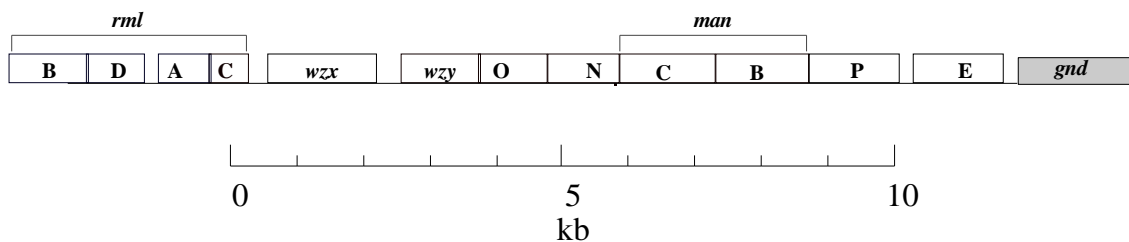
2. Xiang, S.H, Hobbs, M. and Reeves, P.R.

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.
Journal of Bacteriology 176: 4357-4365, 1994.

Salmonella enterica group E O antigen



S. enterica sv. Anatum *wba* genes



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

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.

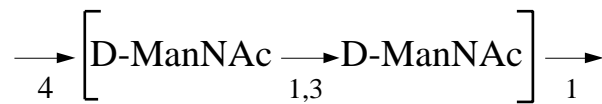
Genetics. 130: 429-443, 1992.

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

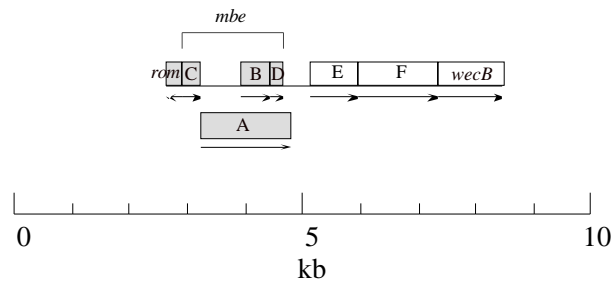
Glycosyl transferases of O-antigen biosynthesis in *Salmonella enterica*: identification and characterisation of transferase genes of groups B, C2 and E1.

Journal of Bacteriology. 175: 3408-3413, 1993.

Salmonella enterica group O:54 sv Borreze O antigen



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



Gene	Product Name	Old Names
wbbE	N-acetylmannosaminotransferase	rfbA
wbbF	protein of unknown function	rfbB
wecB	UDP-N-acetylglucosamine-2-epimerase	rfbC

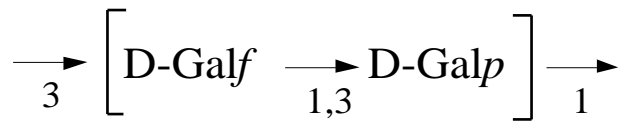
References

1. Keenleyside, W.J. and Whitfield, C.

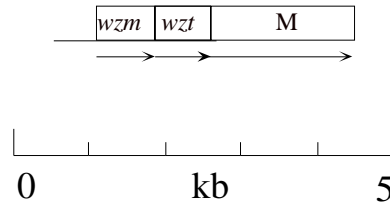
Lateral transfer of *rfb* genes: a mobilizable ColE1-type plasmid carries the *rfb* O:54 (O:54 antigen biosynthesis) gene cluster from *Salmonella enterica* serovar Borreze.

Journal of Bacteriology 177: 5247-5253, 1995.

Serratia marcescens O16 O antigen



S. marcescens O16
wbb genes



Gene	Product Name	Old Names
wbbM	protein of unknown function	rfbF
wzm	ABC-2 type transport system integral membrane protein	rfaA
wzt	ABC-2 type transport system ATP-binding protein	rfbB

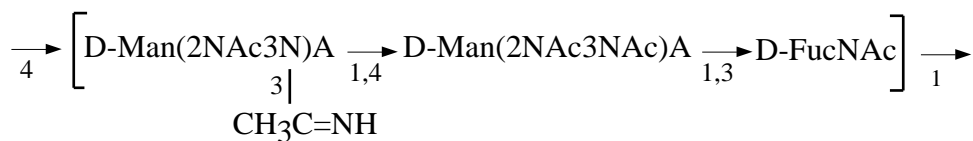
References

1. Szabo, M., Bronner, D. and Whitfield, C.

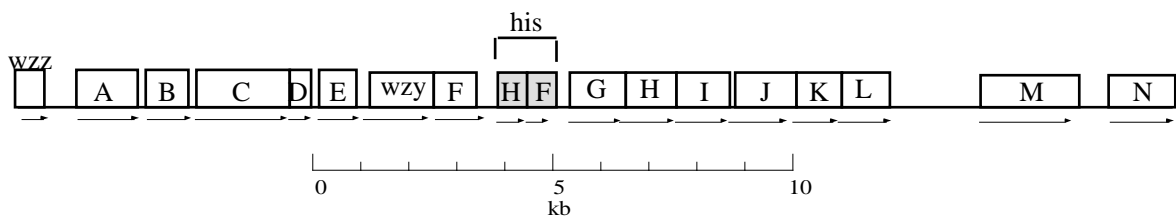
Relationships between *rfb* gene clusters required for biosynthesis of identical D-galactose-containing O antigens in *Klebsiella pneumoniae* serotype O1 and *Serratia marcescens* serotype O16.

Molecular Microbiology 14:505-519, 1994.

Pseudomonas aeruginosa O5 O antigen



P. aeruginosa O5
wbp genes



Gene	Product Name	Old Names
<i>wbpA</i>	putative O-antigen biosynthetic protein	
<i>wbpB</i>	putative O-antigen biosynthetic protein	
<i>wbpC</i>	putative O-antigen biosynthetic protein	
<i>wbpD</i>	putative O-antigen biosynthetic protein	
<i>wbpE</i>	putative O-antigen biosynthetic protein	
<i>wbpF</i>	putative O-antigen biosynthetic protein	
<i>wbpG</i>	putative O-antigen biosynthetic protein	
<i>wbpH</i>	putative O-antigen biosynthetic protein	
<i>wbpI</i>	putative O-antigen biosynthetic protein	
<i>wbpJ</i>	putative O-antigen biosynthetic protein	
<i>wbpK</i>	putative O-antigen biosynthetic protein	
<i>wbpL</i>	UDP-GlcNAc transferase	<i>rfaA</i>
<i>wbpM</i>	putative O-antigen biosynthetic protein	
<i>wbpN</i>	putative O-antigen biosynthetic protein	
<i>wzy</i>	polymerase	<i>rfaC</i>
<i>wzz</i>	flippase	

References

1. Dasgupta, T. Lam, J.S.

Identification of *rfaA*, involved in B-band lipopolysaccharide biosynthesis in *Pseudomonas aeruginosa* serotype O5

Infection and Immunity 63: 1674-1680, 1995.

2. de Kievit, T.R. Dasgupta, T. Schweizer, H. Lam, J.S.

Molecular cloning and characterization of the *rfaC* gene of *Pseudomonas aeruginosa* (serotype O5)

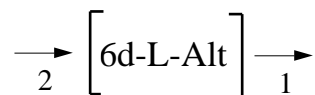
Molecular Microbiology 16: 565-574, 1995.

3. Burrows, L.L. Charter, D.F. Lam, J.S.

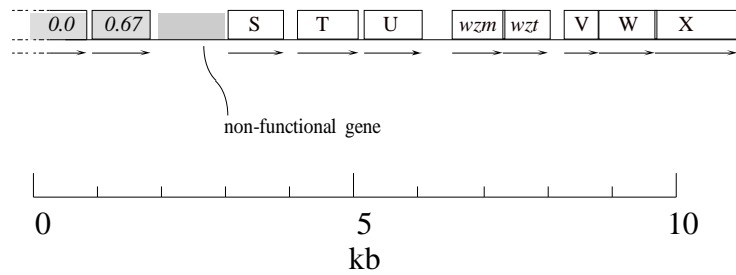
Molecular characterization of the *Pseudomonas aeruginosa* serotype O5 (PAO1) B-band lipopolysaccharide gene cluster.

Molecular Microbiology 22: 481-495.

Yersinia enterocolitica O3 O antigen.



Y. enterocolitica O3 OAg
wbb genes



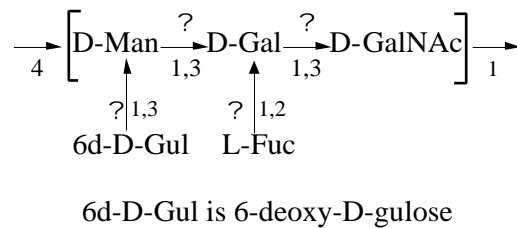
Gene	Product Name	Old Names
wbbS	putative 6-dAlt biosynthetic enzyme 1	rfbA
wbbT	putative glycosyltransferase	rfbB
wbbU	putative glycosyltransferase	rfbC
wbbV	putative 6-dAlt biosynthetic enzyme (2)	rfbF
wbbW	putative 6-dAlt biosynthetic enzyme (3)	rfbG
wbbX	O-antigen biosynthetic protein	rfbH
wzm	ABC-2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

References

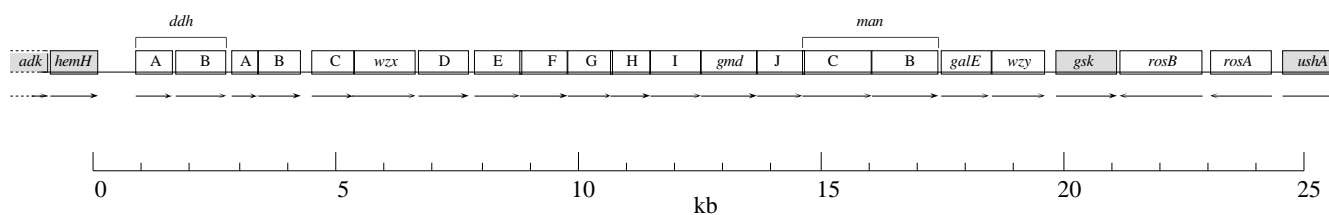
1. Zhang, L., al-Hendy, A., Toivanen, P. and Skurnik, M.

Genetic organisation 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. *Molecular Microbiology* 9:309-21, 1993.

Yersinia enterocolitica O8 O antigen.



Y. enterocolitica O8
wbc genes



Gene	Product Name	Old Names
ddhA	D-glucose-1-phosphate cytidyltransferase	rfbF
ddhB	CDP-glucose 4,6-dehydratase	rfbG
galE	UDP-glucose 4-epimerase	
gmd	GDP-D-mannose 4,6-dehydratase	orf137
manA	α-D-mannose-1-phosphate guanylyltransferase	rfbK
manB	α-D-mannose-1-phosphate guanylyltransferase	rfbM
rosA	putative regulator of O-antigen expression	
rosB	putative regulator of O-antigen expression	
wbcA	putative 6-deoxygulose synthetase (1)	orf4.0
wbcB	putative 6-deoxygulose synthetase (2)	orf4.5
wbcC	putative 6-deoxygulose transferase	orf5.6
wbcD	putative glycosyltransferase	orf7.8
wbcE	putative O-antigen biosynthetic protein	orf8.9
wbcF	putative O-antigen biosynthetic protein	orf9.9
wbcG	putative glycosyltransferase	orf109
wbcH	putative galactoside 2-L-fucosyltransferase	orf118
wbcI	putative galactosyltransferase	rfbP
wbcJ	putative O-antigen biosynthetic protein	orf148
wzx	flippase	rfbX
wzy	polymerase	rfc

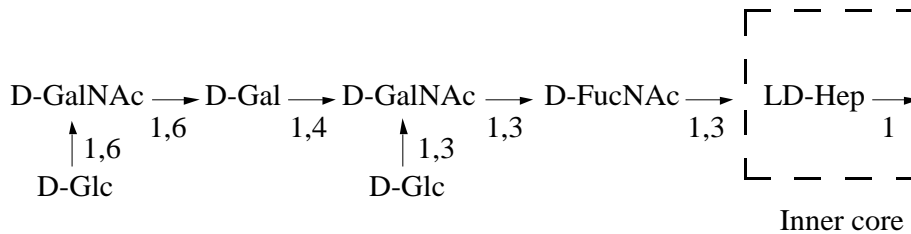
References

- 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.
- 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.
- 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.
Microbiology 142:277-288 (1996) .

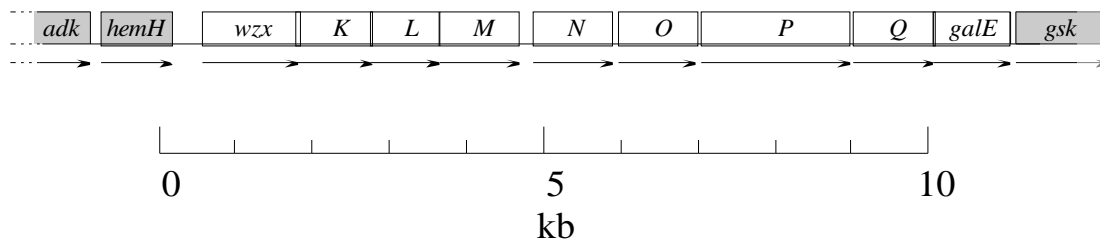
4. Tomshich, S.V., Gorshkova, R.P. and Ovodov, Y.S.

Structural studies on lipopolysaccharide from *Y. enterocolitica* serovar O:8.
 Khim. Prir. Soedin. ###: 657-664 (1987)

Yersinia enterocolitica O3 outer core.



Y. enterocolitica O3
wbc genes



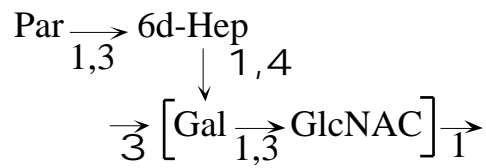
Gene	Product Name	Old Names
galE	UDP-glucose 4-epimerase	trsB
wbcK	putative glycosyltransferase	trsC
wbcL	putative glycosyltransferase	trsD
wbcM	putative glycosyltransferase	trsE
wbcN	putative glycosyltransferase	trsF
wbcO	putative glycosyltransferase	trsG
wbcP	putative FucNAc biosynthetic enzyme	trsH
wbcQ	putative glycosyltransferase	trsA
wzx	flippase	

References

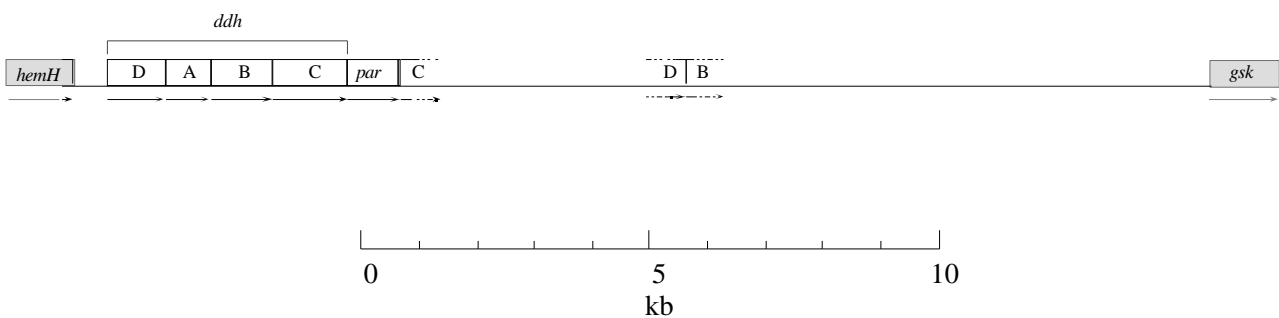
1. Skurnik, M., Venho, R., Toivanen, P. and Alhendy, A.
 A novel locus of *Yersinia enterocolitica* serotype O-3 involved in lipopolysaccharide outer core biosynthesis.
 Molecular Microbiology 17: 575-594, 1995.

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.
 8th European Carbohydrate Symposium, July 2-7, 1995, Sevilla, Spain. Abstracts, B017.

Yersinia pseudotuberculosis IA O antigen



Y. pseudotuberculosis IA wby genes



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

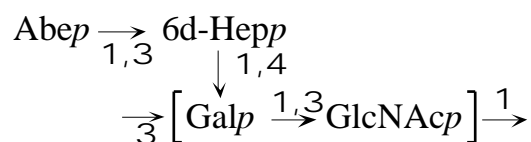
References

1. Hobbs, M. and Reeves, P.R.

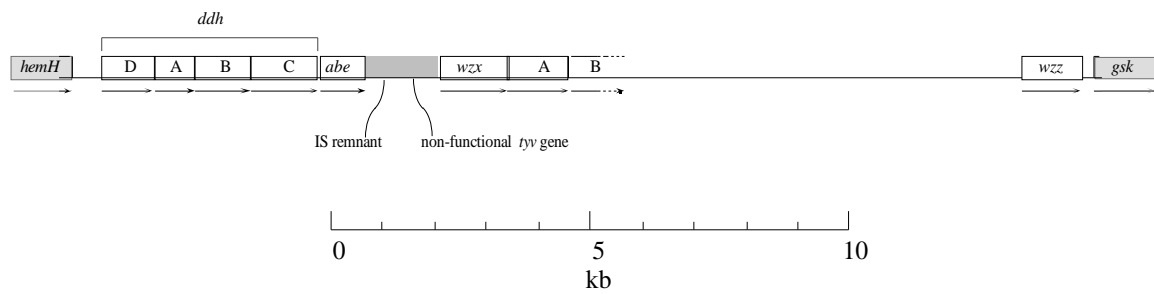
Genetic organisation and evolution of *Yersinia pseudotuberculosis* 3,6-dideoxyhexose biosynthetic genes.

Biochimica et Biophysica Acta. 1245(3):273-7, 1995.

Yersinia pseudotuberculosis IIA O antigen



Y. pseudotuberculosis IIA
wby genes



Gene	Product Name	Old Names
abe	CDP-abequose synthase	rfbJ
ddhA	D-glucose-1-phosphate cytidyltransferase	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-glucoseen reductase	rfbI
wbyA	abequosyltransferase	orf8.7
wbyB	putative O-antigen biosynthetic protein	orf9.8
wzx	flippase	orf7.4
wzy	flippase	rfbX
wzz	chain length determinant	clD

References

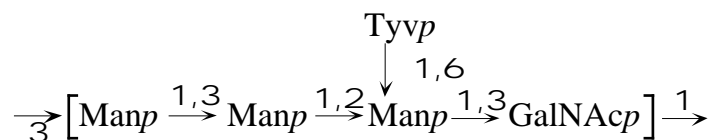
1. Kessler, A.C., Haase, A. and Reeves, P.R.

Molecular analysis of the 3,6-dideoxyhexose pathway genes of *Yersinia pseudotuberculosis* serogroup IIA. *Journal of Bacteriology*. 175: 1412-1422, 1993.

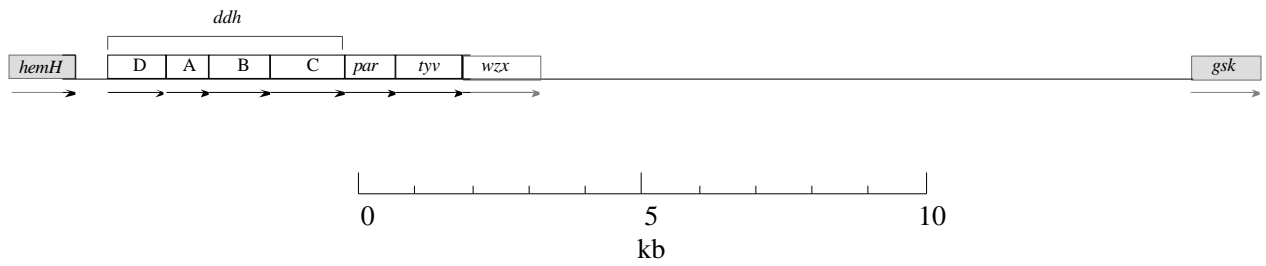
2. Stevenson, G., Kessler, A. and Reeves P.R.

A plasmid-borne O-antigen chain length determinant and its relationship to other chain length determinants. *FEMS Microbiology Letters*. 125:23-30, 1995 .

Yersinia pseudotuberculosis IVA O antigen



Y. pseudotuberculosis IVA
wby genes



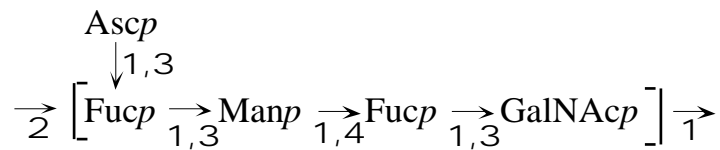
Gene	Product Name	Old Names
ddhA	D-glucose-1-phosphate cytidyltransferase	rfbF
ddhB	CDP-glucose 4,6-dehydratase	rfbG
ddhC	CDP-4-keto-6-deoxy-D-glucose 3-dehydrase	rfbH
ddhD	CDP-6-deoxy-delta3,4-glucoseen reductase	rfbI
prt	paratose synthetase	rfbS
tyv	CDP-paratose epimerase	rfbE
wzx	flippase	rfbX

References

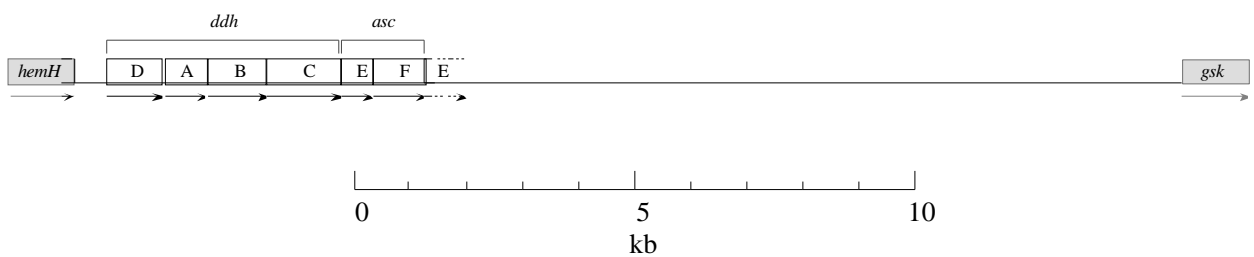
1. Hobbs, M. and Reeves, P.R.

Genetic organisation and evolution of *Yersinia pseudotuberculosis* 3,6-dideoxyhexose biosynthetic genes. *Biochimica et Biophysica Acta*. 1245(3):273-7, 1995.

Yersinia pseudotuberculosis VA O antigen



Y. pseudotuberculosis VA
wby genes



ascE	CDP-3,6-dideoxy-D-glycero-D-glycero-4-hexulose-5-epimerase	ascE
ascF	CDP-3,6-dideoxy-L-glycero-D-glycero-4-hexulose-4-reductase	ascF
ddhA	D-glucose-1-phosphate cytidyltransferase	ascA
ddhB	CDP-glucose 4,6-dehydratase	ascB
ddhC	CDP-4-keto-6-deoxy-D-glucose 3-dehydrase	ascC
ddhD	CDP-6-deoxy-delta3,4-glucose reductase	ascD
wbyE	putative O-antigen biosynthetic protein	asc6.0

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.

Journal of Bacteriology 176:5483-5493, 1994.