Extensive Variation in the O-Antigen Gene Cluster within One *Salmonella enterica* Serogroup Reveals an Unexpected Complex History

Lei Wang,[1]† Kanella Andrianopoulos,[1] Dan Liu,[1] Michel Y. Popoff,[2] and Peter R. Reeves[1]*

*Department of Microbiology, The University of Sydney, Sydney, New South Wales 2006, Australia,[1] and
WHO Collaboration Center for Reference and Research on Salmonella, Unité de Génétique
des Bactéries Intracellulaires, Institut Pasteur, 75724 Paris Cedex 15, France[2]*

The 46 serogroups of *Salmonella enterica* have different O-antigens, and each is thought to have a specific form of the O-antigen cluster. Comparison of the 145 serovars of serogroup B revealed much more intraserogroup genetic diversity than expected. The O27 factor, due to an α 1–6 linkage between O units in place of the more common α 1–2 linkage and previously thought to be due to a converting bacteriophage, is now shown to be due to a $wzy_{\alpha(1-6)}$ gene located within the major gene cluster. Surprisingly a remnant of this gene in all O27$^-$ serovars shows that the ancestor was O27$^+$. There are six distinct gene cluster forms, five apparently derived by a series of deletions and one by an insertion from an ancestral O27$^+$ form present in 57 serovars. The history of the gene cluster and movement between subspecies I and II can be traced. Two of the derivative forms still have a functional $wzy_{\alpha(1-6)}$ gene, while in three it has been inactivated by deletion or insertion. Two of the forms lacking a functional $wzy_{\alpha(1-6)}$ gene have the $wzy_{\alpha(1-2)}$ gene first described for strain LT2 as *rfc*, whereas for the third the *wzy* gene has not been located.

Lipopolysaccharide (LPS), an important component of the outer membrane of gram-negative bacteria, usually consists of three distinct regions: lipid A, core oligosaccharide, and O-specific polysaccharide (O-antigen). The O-antigen, comprising repeats of an O unit of generally two to six sugars, contributes major antigenic variability to the cell surface. On the basis of this variation, *Salmonella enterica* has been divided into 46 O serogroups (19).

The surface O-antigen is subject to intense selection by bacteriophages, which use it as a receptor, and by the host immune system, which may account for the maintenance of many different O-antigen forms within species such as *S. enterica*. The genes for O-antigen synthesis are normally grouped together on the chromosome in a gene cluster which maps close to *gnd* in both *Escherichia coli* and *S. enterica* (22).

S. enterica group B strains have an O-antigen with the basic structure shown in Fig. 1. There are several known variations of this structure. There are two possible sites for glucose addition, the genes for these additions being on phages such as P22 or on the chromosome outside of the main O-antigen gene cluster. The antigenic specificities have been unraveled by cross absorption of antibodies (20). The basic structure confers O factors 4 and 12. The presence of 1–6-linked and 1–4-linked side branch glucose residues is determined by *gtrABC* on phage P22 and homologous genes on the chromosome, which confer epitopes 1 and 12, respectively (14, 24, 27). The addition of an O acetyl group confers epitope 5, and this is determined by the

chromosomal *oafA* gene (9), which is close to but distinct from the main O-antigen gene cluster (25).

Finally, there is variation in the linkage between O units. *S. enterica* serovar Typhimurium has α 1–2 linkages between O units. The *wzy* gene for this linkage is atypical in not being in the main O-antigen gene cluster (2). However, other strains can have an α 1–6 linkage, and this confers the additional epitope O27 (12, 13). The O27 epitope has long been thought to be due to a polymerase gene on a converting phage (see reference 16 for a review), but data presented in this paper support an alternative explanation.

The group D1 antigen differs from that of group B by having tyvelose in place of abequose, and this confers the O9 epitope in place of O4. The two glucose linkages are also known in group D, with O factor 1 conferred by phage P22. The group D2 antigen has the same overall structure as D1 but has a β mannosyl-rhamnose linkage in place of an α mannosyl-rhamnose linkage and an α 1–6 linkage between O units in place of an α 1–2 linkage. The group D2 O-antigen resembles that of group E1 except that E1 lacks the tyvelose side branch. We have shown that D2 strains have the same *wzy* and mannosyl transferase genes as group E1 strains (28) and that group D3 strains, which also have an α 1–6 linkage between O units, have a polymerase gene which differs from that of groups E1 and D2. To our surprise, there was high-level similarity between a part of this gene and part of the *wbaV-wbaU* intergenic segment in group B serovar Typhimurium strain LT2 (3). This clearly identified a remnant *wzy* gene in a group B O-antigen gene cluster and raised the possibility that this was the remnant of an α 1–6 polymerase gene which at some stage had been replaced functionally by the $wzy_{\alpha(1-2)}$ gene found at map position 35.7 min.

In this paper we extend these observations and show that all group B O27$^+$ strains have, within the O-antigen gene cluster,

* Corresponding author. Mailing address: Department of Microbiology, The University of Sydney, Sydney, New South Wales 2006, Australia. Phone: (612) 351 2536. Fax: (612) 351 4571. E-mail: reeves@angis.org.au.
† Present address: Department of Microbiology, College of Life Science, Nankai University, Tianjin 300071, People's Republic of China.

O factors 4, 12$_1$ and 12$_3$

O factors 4, 12$_1$, 12$_3$ and 5

O factors 4, 12$_1$, 12$_3$ and 1

O factors 4, 12$_1$, 12$_3$ and 12$_2$
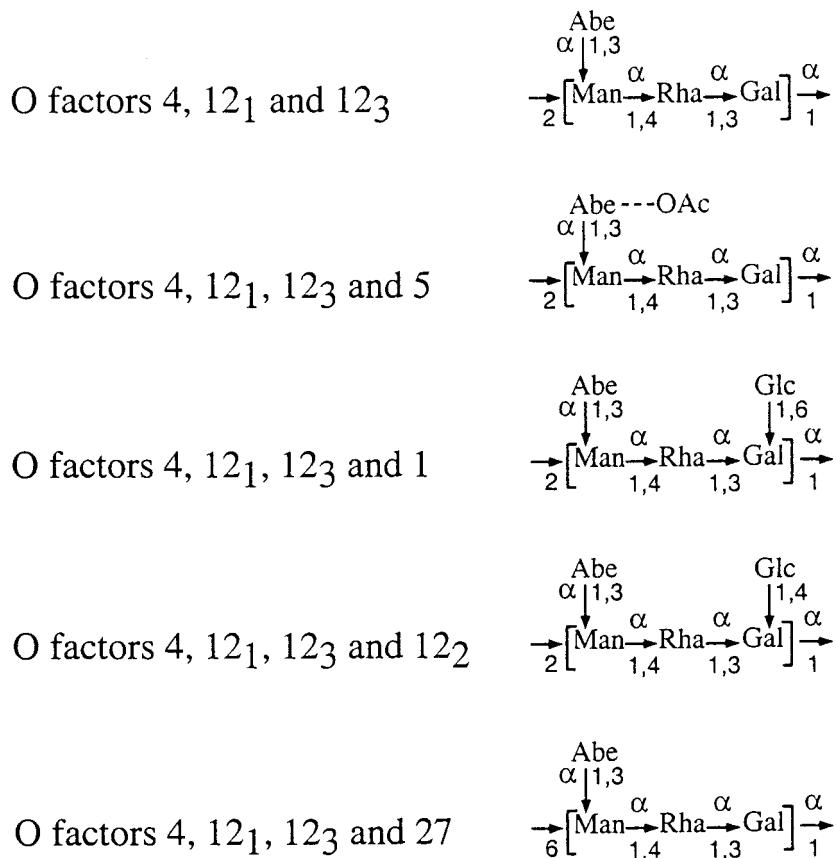
O factors 4, 12$_1$, 12$_3$ and 27

FIG. 1. Basic structure of *S. enterica* group B O-antigen (top) and its variants. Abbreviations: Abe, abequose; Gal, galactose; Glc, glucose; Man, mannose, OAc, O-acetyl; Rha, rhamnose.

the $wzy_{\alpha(1-6)}$ gene first found in group D3. We also show that the O-antigen gene cluster carrying the $wzy_{\alpha(1-6)}$ gene represents the ancestral form of the group B gene cluster.

MATERIALS AND METHODS

Bacterial strains. The *S. enterica* strains listed in Table 1 are the reference strains for the serovars (except LT2 of serovar Typhimurium) and were obtained from the WHO Collaborating Center for Reference and Research on *Salmonella*, Institut Pasteur. The antigenic formula of each *S. enterica* strain was assayed by slide agglutination with specific antisera (antisera were obtained from Bio-Rad). The *S. enterica* serovar Typhimurium $wzy_{\alpha(1-2)}$ mutant strain LV386 was kindly provided by J. Hackett. *E. coli* K-12 strain JM109 was used as the host for cloning.

DNA methods. Chromosomal and plasmid DNA was isolated using the Wizard DNA preparation kits from Promega. PCR, restriction enzyme digestion, ligation, and Southern blotting were done as described by Sambrook et al. (23). Electroporation was performed using a Pulser Controller (Bio-Rad) and the method provided by Bio-Rad. Sequencing was carried out with an Applied Biosystems 377 sequencer.

Other methods. LPS was prepared as described by Franco et al. (5). Analysis of LPS by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and silver staining was performed according to Fomsgaard et al. (4) and Aucken and Pitt (1), respectively.

Bioinformatics. DNA sequences were assembled and edited using the programs Phred, Phrap, and Consed (6). Further analysis was undertaken using programs available through the Australian National Genomic Information Service (ANGIS) at the University of Sydney. Sequence comparisons were analyzed using the Multicomp package (21), which gives pairwise comparisons of DNA and amino acid sequences.

Nucleotide sequence accession numbers. The DNA sequences reported have been deposited in GenBank under accession numbers AY062284 to AY062293, AY064418, and AY064419.

RESULTS

***wbaV-wbaU* intergenic region from Schleissheim, an O27$^+$ serovar.** As discussed above, part of the *wzy* gene of group D3 shows high-level similarity with part of the *wbaV-wbaU* intergenic region of the O-antigen gene cluster of O27$^-$ group B strain LT2 (serovar Typhimurium). To explore the possibility that some group B O27$^+$ strains have an intact *wzy* gene located between *wbaV* and *wbaU*, we first amplified the *wbaV-wbaU* intergenic region from several O27$^+$ strains, using primers 278 (positions 14054 to 14077 of LT2, 5′ end of *wbaV*) and 22 (positions 16940 to 16920 of LT2, middle of *wbaU*) and obtained a 3.7-kb fragment. The PCR product from serovar Schleissheim was cloned into pGEM-T, giving plasmid pPR1843. The insert was sequenced by primer walking from the 5′ end to the start of the *wbaU* gene. A 3,651-bp sequence was obtained, and analysis revealed four open reading frames, two being part of an insertion sequence (IS) (Fig. 2).

The first 983 bp of DNA covers the whole *wbaV* gene except for the first 19 bp, which was the primer sequence (primer 278). It showed 97.8% identity with the Typhimurium *wbaV* gene. The amino acid sequences showed 97.5% identity.

TABLE 1. *S. enterica* group B serovars

Strain no.	Serovar	O-antigens	H antigen		$wzy_{\alpha(1-6)}$ of serovar Schleissheim present	$wzy_{\alpha(1-2)}$ of serovar Typhimurium present	Form[a]
			Phase 1	Phase 2			
1994	Ayinde	1,4,12,27	d	z6	+	−	B(i)
1993	Ayton	1,4,12,27	l,w	z6	+	−	B(i)
2023	Ball	1,4,12,27	y	e,n,x	+	−	B(i)
1491	Bredeney3	1,4,12,27	l,v	1,7	+	−	B(i)
2032	Brezany	1,4,12,27	d	1,6	+	−	B(i)
1490	Budapest	1,4,12,27	g,t	—	+	−	B(i)
2046	Bury	4,12,27	c	z6	+	−	B(i)
1979	Drogana	1,4,12,27	r,i	e,n,z15	+	−	B(i)
2015	Entebbe	1,4,12,27	z	z6	+	−	B(i)
2011	Eppendorf	1,4,12,27	d	1,5	+	−	B(i)
2054	Fortune	4,12,27	z10	z6	+	−	B(i)
2051	Gloucester	1,4,12,(27)	i	l,w	+	−	B(i)
2056	Hallfold	1,4,12,27	c	l,w	+	−	B(i)
2013	Jericho	1,4,12,27	c	e,n,z15	+	−	B(i)
2038	Jos	1,4,12,27	y	e,n,z15	+	−	B(i)
2037	Kamoru	4,12,27	y	z6	+	−	B(i)
2040	Kano	1,4,12,27	l,z13,z28	e,n,x	+	−	B(i)
2074	Kimuenza	1,4,12,27	l,v	e,n,x	+	−	B(i)
2067	Kingston	1,4,12,27	g,s,t	—	+	−	B(i)
2095	Kubacha	1,4,12,27	l,z13,z28	1,7	+	−	B(i)
2079	Kunduchi	1,4,12,27	l,z13,z28	1,2	+	−	B(i)
2081	Legon	1,4,12,27	c	1,5	+	−	B(i)
2077	Ljubljana	4,12,27	k	e,n,x	+	−	B(i)
2063	Maska	1,4,12,27	z41	e,n,z15	+	−	B(i)
2064	Massenya	1,4,12,27	k	1,5	+	−	B(i)
2070	Nakuru	1,4,12,27	a	z6	+	−	B(i)
2097	Neumuenster	1,4,12,27	k	1,6	+	−	B(i)
2101	Remo	1,4,12,27	r	1,7	+	−	B(i)
2103	Sarajane	4,12,27	d	e,n,x	+	−	B(i)
1412	Schleissheim1	4,12,27	b	—	+	−	B(i)
2069	Schwarzengrund	1,4,12,27	d	1,7	+	−	B(i)
2105	Sloterdijk	1,4,12,27	z35	z6	+	−	B(iii)
2106	Southampton	4,12,27	r	z6	+	−	B(i)
1492	Stanley	1,4,[5],12,27	d	1,2	+	−	B(i)
2108	Tafo	1,4,12,27	z35	1,7	+	−	B(i)
2109	Teddington	4,12,27	y	1,7	+	−	B(i)
1984	Thayngen	1,4,12,27	z41	1,2,5	+	−	B(i)
2080	Tinda	1,4,12,27	a	e,n,z15	+	−	B(i)
2115	Trachau	4,12,27	y	1,5	+	−	B(i)
2078	Tripoli	1,4,12,27	b	z6	+	−	B(i)
2119	Uppsala	1,4,12,27	b	1,7	+	−	B(i)
2120	Vellore	1,4,12,27	z10	z35	+	−	B(ii)
2121	Vom	4,12,27	l,z13,z28	e,n,z15	+	−	B(i)
2122	Vuadens (X-98)	4,12,27	z4,z23	z6	+	−	B(i)
2071	Wagenia	1,4,12,27	b	e,n,z15	+	−	B(i)
1995	Wien	1,4,12,27	b	l,w	+	−	B(i)
2123	Wilhelmsburg	4,12,27	z38	—	+	−	B(i)
2125	Yaounde	4,12,27	z35	e,n,z15	+	−	B(i)
2004	II	4,12,27	g,t	—	+	−	B(i)
2006	II	1,4,5,12,27	a	e,n,x	+	−	B(i)
2024	II	1,4,12,27	z	e,n,x	+	−	B(i)
2027	II	1,4,12,27	a	z39	+	−	B(i)
2014	II	1,4,12,27	g,m,s,t	e,n,x	+	−	B(i)
2018	II	1,4,12,27	l,v	z39	+	−	B(i)
2020	II	1,4,12,27	k	1,6	+	−	B(i)
2043	II	1,4,12,27	l,v	e,n,x	+	−	B(i)
2048	II	(1),4,12,27	b	e,n,x	+	−	B(i)
2055	II	1,4,12,27	z	1,5	+	−	B(i)
2005	II	1,4,12,27	e,n,x	1,7	+	−	B(i)
2041	II	4,12,27	i	z35	+	−	B(ii)
1996	Azteca	4,5,12	l,v	1,5	−	−	B(vi)
2072	Travis	4,12	g,z51	1,7	−	−	B(vi)
2110	Tejas	4,12	z36	—	−	−	B(vi)
2111	Tennyson	4,12	g,z51	e,n,z15	−	−	B(vi)
2112	Texas	4,12	k	e,n,z15	−	−	B(vi)

Continued on following page

TABLE 1—*Continued*

Strain no.	Serovar	O-antigens	H antigen		$wzy_{\alpha(1-6)}$ of serovar Schleissheim present	$wzy_{\alpha(1-2)}$ of serovar Typhimurium present	Form[a]
			Phase 1	Phase 2			
2019	II	4,12	z	1,7	−	+	B(iv)
2001	II	4,12	l,z28	—	−	+	B(iv)
2002	II	1,4,12	z29	e,n,x	−	+	B(iv)
2003	II	4,12	d	e,n,x	−	+	B(iv)
2007	II	4,12	b	1,5	−	+	B(iv)
2008	II	4,12	g,z62	—	−	+	B(iv)
2016	II	4,12	g,m,t	z39	−	+	B(iv)
2022	II	4,12	z	z39	−	+	B(iv)
2025	II	4,12	e,n,x	1,2,7	−	+	B(iv)
2026	II	4,12	z39	1,5,7	−	+	B(iv)
2028	II	4,12	l,w	e,n,x	−	+	B(iv)
2044	II	1,4,5,12	f,g,t	z6:z42	−	+	B(iv)
2000	Abony	1,4,5,12	b	e,n,x	−	+	B(v)
1978	Abortusequi	4,12	—	e,n,x	−	+	B(v)
1975	Abortusovis	4,12	c	1,6	−	+	B(v)
1985	Africana	4,12	r,(i)	l,w	−	+	B(v)
1988	Agama	4,12	i	1,6	−	+	B(v)
1977	Agona	4,12	f,g,s	—	−	+	B(v)
1998	Albert	4,12	z10	e,n,x	−	+	B(v)
1992	Altendorf	4,12	c	1,7	−	+	B(v)
1989	Arechavaleta	4,5,12	a	1,7	−	+	B(v)
2049	Banana	4,5,12	m,t	—	−	+	B(v)
2030	Bispebjerg	1,4,12	a	e,n,x	−	+	B(v)
2047	Bissau	4,12	c	e,n,x	−	+	B(v)
2045	Bochum	4,5,12	r	l,w	−	+	B(v)
2031	Bradford	4,12	r	1,5	−	+	B(v)
2017	Brancaster	1,4,12	z29	—	−	+	B(v)
1990	Brandenburg	4,5,12	l,v	e,n,z15	−	+	B(v)
1980	California	4,12	g,m,t	—	−	+	B(v)
2029	Canada	4,12	b	1,6	−	+	B(v)
2060	Chester	4,12	e,h	e,n,x	−	+	B(v)
2042	Clackamas	4,12	l,v	1,6	−	+	B(v)
2050	Coeln	4,5,12	y	1,2	−	+	B(v)
1999	Derby	1,4,5,12	f,g	—	−	+	B(v)
1991	Duisburg	4,12	d	e,n,z15	−	+	B(v)
2039	Eko	4,12	e,h	1,6	−	+	B(v)
1987	Essen	4,12	g,m	—	−	+	B(v)
2057	Farsta	4,12	i	e,n,x	−	+	B(v)
2053	Finaghy	4,12	y	1,6	−	+	B(v)
1997	Fulica	4,5,12	a	—	−	+	B(v)
2058	Fyris	4,5,12	l,v	1,2	−	+	B(v)
2052	Haduna	4,12	l,z13,z28	1,6	−	+	B(v)
2021	Haifa	1,4,5,12	z10	1,2	−	+	B(v)
2035	Hato	4,5,12	g,m,s	—	−	+	B(v)
1976	Heidelberg	1,4,5,12	r	1,2	−	+	B(v)
2033	Hessarekl	4,12	a	1,5	−	+	B(v)
2059	Huettwillen	1,4,12	a	l,w	−	+	B(v)
2034	Indiana	1,4,12	z	1,7	−	+	B(v)
2012	Iruri	1,4,12	z10	1,5	−	+	B(v)
2036	Kaapstad	4,12	e,h	1,7	−	+	B(v)
2076	Kalamu	4,12	z4,z24	—	−	+	B(v)
2075	Kiambu	4,12	z	1,5	−	+	B(v)
2065	Kisangani	1,4,5,12	a	1,2	−	+	B(v)
2062	Koenigstuhl	1,4,12	z	e,n,z15	−	+	B(v)
2073	Lagos	1,4,12	i	1,5	−	+	B(v)
2066	Limete	1,4,12	b	1,5	−	−	B(v)
2084	Loubomo	4,12	z	1,6	−	+	B(v)
2068	Madras	4,5,12	m,t	e,n,z15	−	+	B(v)
2096	Mono	4,12	l,w	1,5	−	+	B(v)
2083	Mons	1,4,12	d	l,w	−	+	B(v)
2082	Mura	1,4,12	z10	l,w	−	+	B(v)
2061	Neftenbach	4,12	z	e,n,x	−	+	B(v)
1986	Paratyphi B	1,4,5,12	b	1,2	−	+	B(v)
1983	Paratyphi B	1,4,5,12	b	—	−	+	B(v)
2098	Pasing	4,12	z35	1,5	−	+	B(v)

Continued on following page

TABLE 1—*Continued*

Strain no.	Serovar	O-antigens	H antigen		$wzy_{\alpha(1-6)}$ of serovar Schleissheim present	$wzy_{\alpha(1-2)}$ of serovar Typhimurium present	Form[a]
			Phase 1	Phase 2			
2099	Preston	1,4,12	z	l,w	−	+	B(v)
1981	Reading	(1),4,5,12	e,h	1,5	−	+	B(v)
2100	Reinickendorf	4,12	l,z28	e,n,x	−	+	B(v)
1982	Saintpaul	1,4,5,12	e,h	1,2	−	+	B(v)
2102	Sandiego	4,12	e,h	e,n,z15	−	+	B(v)
2104	Shubra	4,5,12	z	1,2	−	+	B(v)
2107	Stanleyville	1,4,12	z4,z23	—	−	+	B(v)
2113	Togo	4,12	l,w	1,6	−	+	B(v)
2114	Tokoin	4,12	z10	e,n,z15	−	+	B(v)
2116	Tsevie	4,12	i	e,n,z15	−	+	B(v)
2117	Tudu	4,12	z10	1,6	−	+	B(v)
2118	Tumodi	1,4,12	i	z6	−	+	B(v)
9003	Typhimurium	1,4,[5],12	i	1,2	−	+	B(v)
1974	Tyresoe	4,12	l,z13,z28	1,5	−	+	B(v)
2124	Winneba	4,12	r	1,6	−	+	B(v)

[a] See Fig. 2.

DNA from positions 2398 to 3573 encodes a gene showing 98.2% identity at the amino acid level with the D3 *wzy* gene, shown previously to encode the polymerase for α 1–6 galactosyl-mannose linkage between O units (3). Thus, the *wzy* gene of Schleissheim must have the same function, and we refer to it here as $wzy_{\alpha(1-6)}$. Plasmid pPR1843 was transformed into the *S. enterica* serovar Typhimurium $wzy_{\alpha(1-2)}$ mutant strain LV386 and shown by Western blotting to confer the ability to produce LPS with long-chain O-antigen of O27$^+$ (data not shown), indicating that the $wzy_{\alpha(1-6)}$ gene is functional.

DNA from positions 993 to 2209 is an IS element, with inverted repeats of 40 and 39 bp containing 16 mismatches. It shows 73% identity along its entire length with IS*1617* of *Yersinia pestis* plasmid pCD1 (18), which has two putative transposase genes, *orfA* and *orfB*. DNA from positions 1074 to 1340 encodes a protein showing 86% identity and 94% similarity at the amino acid level with OrfA of plasmid pCD1. DNA from positions 1697 to 2173 encodes a protein showing 70% identity and 84% similarity with the C-terminal half of OrfB of plasmid pCD1.

Presence and location of $wzy_{\alpha(1-6)}$ in other group B serovars. The above results show that, in serovar Schleissheim at least, the *wzy* gene encoding the O27 factor is within the main O-antigen gene cluster and is not located on a phage. To see how general this phenomenon is, we carried out PCR on the type strains for all 145 serogroup B serovars using two sets of oligonucleotide primers. They were all tested for their O-antigens, and 60 were confirmed to have O27.

Primers 2138 and 2139 bind to positions 2447 to 2464 and positions 3474 to 3457, respectively, of Schleissheim DNA, within the $wzy_{\alpha(1-6)}$ gene (Fig. 2). PCRs produced a DNA band from all of the other 59 O27$^+$ serovars of the same size as in Schleissheim, indicating that these 60 serovars have a functional $wzy_{\alpha(1-6)}$ gene. None of the O27$^-$ serovars gave any PCR product.

To determine if the $wzy_{\alpha(1-6)}$ gene is also located in the same position, between *wbaV* and *wbaU,* as in serovar Schleissheim, we then carried out PCR on all serovars using primers 2136 and 2137, which bind, respectively, to DNA of the 3′ end of

wbaV (positions 668 to 685 of Schleissheim or 14746 to 14763 of Typhimurium) and DNA of the 5′ end of *wbaU* (positions 3626 to 3608 of Schleissheim or 15435 to 15418 of Typhimurium). Altogether, 57 of the 60 O27$^+$ serovars gave a band of the same size as in Schleissheim, indicating that these serovars have the $wzy_{\alpha(1-6)}$ gene and the IS sequence located in the same position as in Schleissheim [Fig. 2, form B(i)]. Two O27$^+$ serovar strains with $wzy_{\alpha(1-6)}$ (Vellore and II 4,12,27:i:z35) gave a smaller fragment of about 1.9 kb, and serovar Sloterdijk gave a band of about 4 kb. Of the 73 subspecies I O27$^-$ serovars, 68 gave a band of the same size as that in Typhimurium (about 0.75 kb), and five (Travis, Tejas, Tennyson, Texas, and Azteca) gave a band of about 1.8 kb. All 12 subspecies II O27$^-$ serovars gave a band of about 0.8 kb.

Sequences of *wbaV*-*wbaU* intergenic regions from serovars Vellore, II 4,12,27:i:z35, and Sloterdijk. As shown above by PCR, these three serovars have the $wzy_{\alpha(1-6)}$ gene, but the *wbaV*-*wbaU* intergenic regions are not of the same size as those of the other O27$^+$ serovars. These regions were sequenced from the DNA fragments amplified using primer pair 2136/2137.

Serovars Vellore and II 4,12,27:i:z35 [Fig. 2, form B(ii)] gave 1,910-bp sequences that showed 99.9% identity. The 5′ end (positions 1 to 456) and 3′ end (positions 459 to 1910) showed 94.3 and 98.0% DNA identity with DNA from positions 703 to 1157 and from positions 2156 to 3608, respectively, of form B(i). The DNA from positions 1074 to 2173 encodes the two transposase genes in Schleissheim, and this region was largely absent in serovars II 4,12,27:i:z35 and Vellore.

Serovar Sloterdijk gave a 4,115-bp sequence [Fig. 2, form B(iii)]. Comparison with form B(i) showed that serovar Sloterdijk has an IS*3* of 1,213 bp inserted between positions 2006 and 2007 within the transposase *orfB* of form B(i). *orfB* itself is highly similar to corresponding DNA of form B(i), with identity levels of 94.4 and 97.4% for the segments upstream and downstream, respectively, of IS*3*. The IS*3* sequence showed 84% identity along its entire length with the IS*3* of *E. coli* plasmid pABV5 (26), with one and two differences in the 5′ and 3′ inverted repeats, respectively. IS*3* has a single trans-
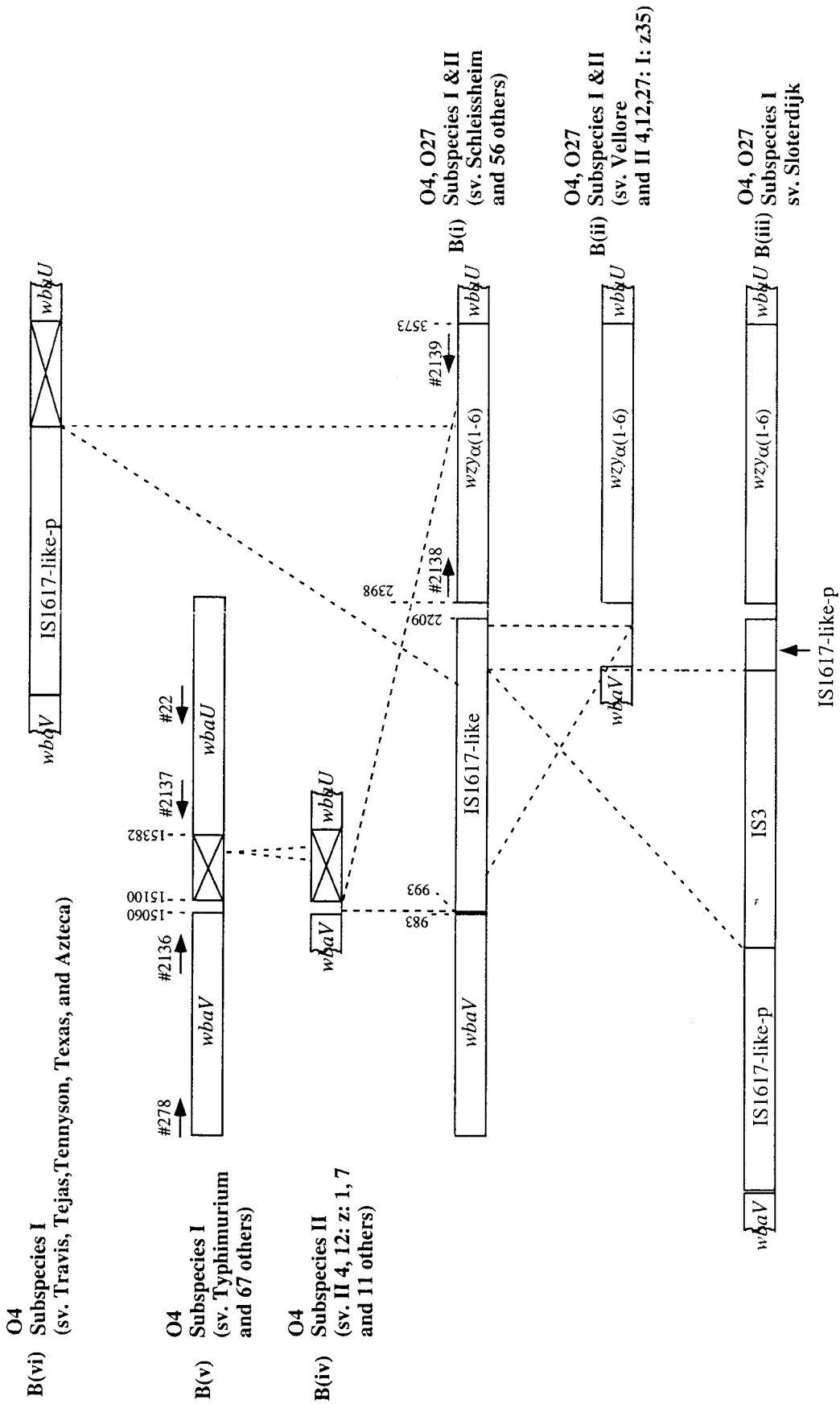
FIG. 2. Six different forms of S. enterica group B O-antigen gene clusters and their relationship. Positions are based on reference 8 for the Typhimurium form and on the sequence reported in this paper for the Schleissheim form. Oligonucleotide primers are indicated by arrows. The box filled with an X indicates a remnant $wzy_{\alpha(1-6)}$ gene. Note that the wzy gene for the three O27$^-$ forms except for serovar Limete is located outside of the main O-antigen gene cluster and is not shown here. The suffix -p indicates that only part of an IS element is present.

posase gene (26), and the deduced amino acid sequence of the gene in Sloterdijk IS*3* shows 87% identity and 91% similarity to that of the *E. coli* IS*3*.

Thus, all 60 O27$^+$ group B serovars have the full $wzy_{\alpha(1-6)}$ gene located between *wbaV* and *wbaU*.

Sequences of *wbaV-wbaU* intergenic regions from serovars Travis, Tejas, Tennyson, Texas, and Azteca. PCR tests showed that these five subspecies I O27$^-$ serovars [Fig. 2, form B(vi)] do not have the $wzy_{\alpha(1-6)}$ gene but have a *wbaV-wbaU* intergenic region of different size from that of form B(v) found in most subspecies I O27$^-$ serovars (see above). The intergenic region was PCR amplified using primer pair 2136/2137 from all five serovars and sequenced. The five sequences are 1,832 to 1,834 bp in length and are highly similar, with pairwise identity levels ranging from 99.5 to 99.9%. The 5′ (positions 1 to 1266 of serovar Tejas) and 3′ (positions 1267 to 1834 of serovar Tejas) segments of this DNA show 96.6 and 98.2% identity with the DNA from positions 703 to 1982 and from positions 3041 to 3608, respectively, of form B(i). Thus, there is a deletion in the O-antigen gene cluster of these five serovars, which includes the second transposase gene and the 5′ part of the $wzy_{\alpha(1-6)}$ gene.

Sequence of *wbaV-wbaU* intergenic region of serovar II 4,12: z:1,7. None of the 12 subspecies II O27$^-$ serovars has the $wzy_{\alpha(1-6)}$ gene, based on PCR (see above), but they all have a *wbaV-wbaU* intergenic region of about 0.8 kb, which is different in size from the corresponding region for most subspecies I O27$^-$ serovars. We sequenced this 0.8-kb fragment from serovar II 4,12:z:1,7 and compared it to the corresponding sequences of serovars Typhimurium [form B(v)] and Schleissheim [form B(i)].

Compared to form B(i), the intergenic regions of serovars II 4,12:z:1,7 [Fig. 2, form B(iv)] and Typhimurium both have a deletion from positions 996 to 3182 of Schleissheim (Fig. 2). Typhimurium has a second deletion from positions 3406 to 3514 of Schleissheim (Fig. 2).

Distribution of the $wzy_{\alpha(1-2)}$ gene. It has long been known that serovar Typhimurium has a *wzy* gene far from the O-antigen gene cluster, at a locus first named *rfc* (rough C) (2, 15). This is now seen as atypical, with the distant $wzy_{\alpha(1-2)}$ gene replacing the $wzy_{\alpha(1-6)}$ gene in the O-antigen gene cluster. In order to determine the distribution of the $wzy_{\alpha(1-2)}$ gene in serogroup B, we carried out PCR against all serovars using primers 2134 and 2135, which are specific to $wzy_{\alpha(1-2)}$ of serovar Typhimurium (positions 507 to 524 and 1243 to 1226 of GenBank entry STYRFC, respectively). None of the O27$^+$ serovars gave a band (Table 1). Of the 68 O27$^-$ subspecies I serovars, which have the 0.75-kb *wbaV-wbaU* intergenic region [form B(v) in Fig. 2], all except serovar Limete gave a PCR band of the same size as in serovar Typhimurium.

Five O27$^-$ subspecies I serovars in form B(vi) which have a larger *wbaV-wbaU* intergenic region did not give a PCR band, and the same applied to the 12 form B(iv) O27$^-$ subspecies II serovars (Table 1). We then carried out Southern blotting using as a probe the $wzy_{\alpha(1-2)}$ gene, which was PCR amplified from serovar Typhimurium using primers 2134 and 2135, and *Eco*RI-digested chromosomal DNA. The probe did not hybridize with DNA of form B(vi) or Limete, but for the 12 O27$^-$ subspecies II serovars, the probe hybridized to a band of ap-

proximately 9.5 kb, about 4 kb larger than that of Typhimurium (data not shown).

DNA from a few O27$^+$ serovars was also included in the Southern blotting experiment (data not shown), and it also did not hybridize to the probe. This confirms that the $wzy_{\alpha(1-2)}$ gene of serovar Typhimurium is not present in the O27$^+$ serovars.

The above results indicate that neither of the two identified *wzy* genes of group B *S. enterica* strains is present in serovars of form B(vi), but SDS-PAGE and silver staining of the LPS show that all produce complete O-antigens (data not shown). Thus, they must have a different *wzy* gene located outside of the main O-antigen gene cluster. The serovar Limete isolate is a specific case, as it lacks both the *wzy* genes and does not produce long-chain O-antigen.

There is no simple way to find the *wzy* gene for those lacking both known *wzy* genes. However, we looked at the site where the $wzy_{\alpha(1-2)}$ gene of serovar Typhimurium is located in serovars of both subspecies I and II representing each of the forms in Fig. 2. Primers 4400 and 4401, which bind to bp 788 to 815 and 1118 to 1090 3′ and 5′ to the $wzy_{\alpha(1-2)}$ gene of serovar Typhimurium (http://genome.wustl.edu/gsc/Projects/S.typhimurium/), respectively, were used to do PCR on serovars Schleissheim, II 1,4,12,27:i,v:z39, Vellore, II 4,12,27:i:z35, Sloterdijk, II 4,12:z:1,7, Travis, and Typhimurium. Serovars Schleissheim, II 1,4,12,27:i,v:z39, Vellore, II 4,12,27:i:z35, Sloterdijk, and Travis produced a PCR fragment of about 600 bp, serovar II 4,12:z:1,7 produced a fragment of 3,000 bp, and, as expected, serovar Typhimurium produced a fragment of about 3,100 bp (data not shown).

We then sequenced the PCR products of serovars Schleissheim and II 4,12:z:1,7. Comparison of the sequences with the $wzy_{\alpha(1-2)}$ region of serovar Typhimurium (http://genome.wustl.edu/gsc/Projects/S.typhimurium/) showed that serovar Schleissheim lacks the serovar Typhimurium DNA including the $wzy_{\alpha(1-2)}$ gene and 863 and 481 bp 5′ and 3′, respectively, to the $wzy_{\alpha(1-2)}$ gene, while serovar II 4,12:z:1,7 has a *wzy* gene 94.1% identical to that of serovar Typhimurium (72 nucleotide substitutions, including 21 nonsynonymous). We conclude that serovar II 4,12:z:1,7 has a functional $wzy_{\alpha(1-2)}$ gene. The $wzy_{\alpha(1-2)}$ genes of both serovars Typhi (http://www.sanger.ac.uk/Projects/S_typhi) and Paratyphi A (http://genome.wustl.edu/gsc) are 99.8% identical to that of serovar Typhimurium.

Subspecies II serovars have the same O-antigen gene cluster structure as subspecies I serovars. The results reported above indicate that the $wzy_{\alpha(1-2)}$ gene of subspecies I is present in subspecies II serovars but the *wbaV-wbaU* intergenic region is different in the two subspecies. To determine if there is any other difference within the group B O-antigen gene clusters of these two subspecies, we carried out adjacent gene PCR for all O-antigen genes of serovars Typhimurium (O27$^-$), Schleissheim (O27$^+$), II 4,12:i,z28:$-$ (O27$^-$), and II 1,4,5,12,27:a:e,n,x (O27$^+$). PCR primers were based on the O-antigen sequence of Typhimurium (8) and, for the $wzy_{\alpha(1-6)}$ gene, on that of Schleissheim. All four strains gave the same PCR results for regions other than the *wbaV-wbaU* intergenic region. Thus, there is no difference between subspecies in the presence or order of O-antigen genes.

DISCUSSION

We have determined the composition of the O-antigen gene cluster for the type strains of all *S. enterica* group B serovars, either by sequence or by using PCR to define members with a particular form. This detailed study of one serogroup, all by definition carrying the same basic O-antigen structure, has revealed quite unexpected complexity. However, there are very clear relationships between the forms found, and it is possible to put forward a realistic hypothesis on the course of events.

There are six different forms of the group B gene cluster, but most fit into one of three patterns in terms of functional gene content. (i) All O27$^+$ serovars have the appropriate $wzy_{\alpha(1-6)}$ gene between *wbaV* and *wbaU*, together with all or part of a variant of IS*1617* found on a *Y. pestis* plasmid. (ii) Most serovars with the α 1–2 linkage between O units have the $wzy_{\alpha(1-2)}$ gene outside of the main gene cluster at the same locus as found long ago in serovar Typhimurium strain LT2. (iii) Form B(vi) serovars lack both *wzy* genes. The relationships of the six forms seem clear, as the other five are readily derived from the Schleissheim form (see Fig. 2). The Schleissheim form has a full $wzy_{\alpha(1-6)}$ gene between the *wbaV* and *wbaU* genes, and a remnant of this gene is located in the same position in all O27$^-$ serovars. This indicates that the original group B O-antigen gene cluster had the $wzy_{\alpha(1-6)}$ gene in its current position.

The O27$^-$ strains can all be derived from the B(i) form by deletion or insertion. There are two groups of O27$^+$ strains which have undergone deletion or insertion in the *wbaN-wbaU* intergenic region while retaining the $wzy_{\alpha(1-6)}$ gene. One deletion, which removed most of IS*1617*, generated form B(i). An IS*3* insertion was inserted into one of the transposase genes (*orfB*) of IS*1617* to generate form B(iii). The B(iv) O27$^-$ form found only in subspecies II strains appears to have arisen by a major deletion of most of the $wzy_{\alpha(1-6)}$ gene together with the IS*1617* element. This was followed by a second deletion of more of the $wzy_{\alpha(1-6)}$ gene to give form B(v) found in most O27$^-$ subspecies I strains. An independent deletion in the *wbaV-wbaU* intergenic region including parts of IS*1617* and $wzy_{\alpha(1-6)}$ gave rise to form B(vi), also found only in subspecies I.

The O27 epitope was reported many years ago to be due to the presence of a bacteriophage. However, for the reference strains of all O27$^+$ serovars of both subspecies I and II, it now appears that the polymerase gene is in the main O-antigen gene cluster. The question arises as to the observations made by Le Minor and coworkers (10, 11). The extensive studies on phage conversion leave no doubt that the converting phage existed. However, a converting phage was not consistently isolated, only some serovars giving a converting phage, and most work was done on a phage first isolated from a serovar Schwarzengrund strain but generally studied in serovar Typhimurium and other strains.

The most plausible explanation is that the converting phage was generated by chance by a recombination event in which the $wzy_{\alpha(1-6)}$ gene in the Schwarzengrund strain was picked up by a temperate phage which was released by lysis. Such a phage could lysogenize the recipient strain, making it O27$^+$ and generating the strain carrying the converting phage then studied. One of the two transposases near the $wzy_{\alpha(1-6)}$ gene could have provided the means for translocation of the gene to a phage.

This explanation would account for the low frequency of conversion by phage from the primary O27$^+$ strains. The O27$^+$ strains generated in those experiments are no longer available. However, it now seems clear that they did not demonstrate, as believed at the time, a naturally occurring converting phage, but rather the ease with which such a phage could arise in the laboratory under selection for transfer of one epitope.

The finding that the O27$^+$ and O27$^-$ forms differ in the genes present in their O-antigen gene clusters has implications for subdivision into serogroups and serovars. Isolates differing only in the presence or absence of O27 were placed in the same serovar on the basis that the O27 epitope was phage encoded. On this basis, serovars previously distinguished only by the O27 epitope were combined. For example, serovars Abony and Abortusbovis were combined in 1983 because they differed only with regard to epitope 27 (1,4,5,12:b:e,n,x and 1,4,5,12,27:b:e,n,x, respectively). Because Abony was the first described, its name was retained, and serovar Abortusbovis was withdrawn. Serovar Abony is shown in the current Kauffman-White scheme as 1,4,5,12,27:b:e,n,x, although the type strain which we studied is 1,4,12:b:e,n,x. There is clearly a need to review this part of the Kauffman-White scheme, but no changes are proposed at this time.

Concluding comments. This study started with the aim of seeing if any of the O27$^+$ group B strains had the *wzy* gene found in D3 strains, as a remnant of it was present in the sequenced LT2 O-antigen gene cluster. When the first O27$^+$ strain studied was found to contain the D3 polymerase, all group B serovars were tested, and we found that it was present in all O27$^+$ strains. The study further showed that the O27$^+$ form, far from being a phage-encoded variant of the typical O group B form, is in reality the ancestral form.

The study revealed some interesting variation within both O27$^-$ and O27$^+$ serovars. There is a group of serovars that have lost the O27 polymerase but do not have the α 1–2 polymerase found in most other O27$^-$ serovars. They have long-chain O-antigen and so must have a polymerase, but as yet they have not been shown to have the α 1–2 linkage, as has generally been assumed, but only to lack the α 1–6 linkage that confers O27 specificity. Clearly more work is needed to clarify this. It is also interesting that the 12 O27$^-$ subspecies II serovars have their own characteristic form of the O-antigen gene cluster, which appears to be ancestral to the form found in the major group of subspecies I O27$^-$ serovars that includes serovar Typhimurium with strain LT2.

The $wzy_{\alpha(1-2)}$ gene from the 12 subspecies II strains [form B(iv)] differed at 6% of bases from the three sequenced $wzy_{\alpha(1-2)}$ genes of subspecies I strains. Furthermore, the band sizes on Southern blot are different between the two subspecies but consistent within them. The association of specific forms of O-antigen cluster with specific forms of *wzy* locus implies either that they have been transferred together or that all or most serovars arose by gain or loss of genes other than these for the O-antigen. The first possibility seems unlikely, as the $wzy_{\alpha(1-2)}$ gene locus and O-antigen gene cluster are 748 kb apart, and the second is surprising, as serovars Typhi and Typhimurium, for example, are quite divergent.

To the best of our knowledge, this is the first report on the extent of variation within a gene cluster for such an antigen. The diversity revealed has allowed us to follow in detail the

histories of the gene clusters for one serogroup. It remains to be seen if such a high level of variation is at all typical. It certainly provides potential for further subdivision of serovars by molecular means.

This is unlikely to be the full story of the evolution of serogroup B. The presence of a transposable element next to its *wzy* gene suggests that the ancestral B(i) form itself arose by transposon-mediated recombination, as was postulated for group D2 (28). In that case there were very plausible parent forms in D1 and E1, with the transposable element seen only in the presumed derivative D2 form (28). The transposable elements seen in O-antigen gene clusters are generally adjacent to sites where there is evidence for recombination or where one can reasonably explain recombination. Thus, in *Yersinia pseudotuberculosis* there is a remnant IS*630* next to an apparent recombination site for replacement of the *prt* and *tyv* genes by an *abe* gene, with a remnant *tyv* gene present (7). In *Vibrio cholerae* a defective IS*1358* is found adjacent to *wbeT*, a gene that is nonessential for O-antigen production but modifies the specificity and could clearly have been added later (17).

The transposable element for B(i) provides a clue, but it remains to be seen if we will ever find a precursor of the B(i) form. However, it seems unlikely that it would be in *S. enterica*, as most O-antigens are probably known and those with extensive serological cross-reaction (B, D1, D2, D3, C2, and E1) have been sequenced.

ACKNOWLEDGMENT

REFERENCES

1. **Aucken, H., and T. Pitt.** 1991. Serological relationships of the O-antigens of *Klebsiella pneumoniae* O5, *Escherichia coli* O8 and a new serotype of *Serratia marcescens*. FEMS Microbiol. Lett. **80:**93–98.
2. **Collins, L. V., and J. Hackett.** 1991. Molecular cloning, characterization, and nucleotide sequence of the *rfc* gene, which encodes an O-antigen polymerase of *Salmonella typhimurium*. J. Bacteriol. **173:**2521–2529.
3. **Curd, H., D. Liu, and P. R. Reeves.** 1998. Relationships among the O-antigen gene clusters of *Salmonella enterica* groups B, D1, D2, and D3. J. Bacteriol. **180:**1002–1007.
4. **Fomsgaard, A., M. Freudenberg, and C. Galanos.** 1990. Modification of the silver staining technique to detect lipopolysaccharide in polyacrylamide gels. J. Clin. Microbiol. **28:**2627–2631.
5. **Franco, V. A., D. Liu, and P. R. Reeves.** 1998. The Wzz (Cld) protein in *Escherichia coli*: amino acid sequence variation determines O-antigen chain length specificity. J. Bacteriol. **180:**2670–2675.
6. **Gordon, D., C. Abajian, and P. Green.** 1998. CONSED—a graphical tool for sequence finishing. Genome Res. **8:**195–202.
7. **Hobbs, M., and P. R. Reeves.** 1995. Genetic organization and evolution of *Yersinia pseudotuberculosis* 3,6-dideoxyhexose biosynthetic genes. Biochim. Biophys. Acta **1245:**273–277.
8. **Jiang, X. M., B. Neal, F. Santiago, S. J. Lee, L. K. Romana, and P. R. Reeves.** 1991. Structure and sequence of the *rfb* (O-antigen) gene cluster of *Salmonella* serovar typhimurium (strain LT2). Mol. Microbiol. **5:**695–713.
9. **Kishi, K., and S. Iseki.** 1973. Genetic analysis of the O-antigens in *Salmonella*. II. Heredity of the O-antigens 4, 5 and 9 of *Salmonella*. Jpn. J. Genet. **48:**133–136.
10. **Le Minor, L.** 1962. Conversion par lysogénisation de quelques sérotypes de Salmonella des groups A, B et D normalement dépourvues du facteur O27 eu cultures 27 positives. Ann. Inst. Pasteur Microbiol. **103:**684–706.
11. **Le Minor, L., S. Le Minor, and P. Nicolle.** 1961. Conversion de cultures de *Salmonella Schwarzengrund* et *Salmonella Bredeney* dépourvues de l'antigéue 27 eu cultures 27 positive par le lysogénisation. Ann. Inst. Pasteur Microbiol. **101:**571–589.
12. **Le Minor, L., and A. M. Staub.** 1966. Serologic study of the O27 factors of *Salmonella*. Ann. Inst. Pasteur Microbiol. **110:**834–848.
13. **Lindberg, A. A., C. G. Helleqvist, G. Bagdian-Motta, and P. H. Mäkelä.** 1978. Lipopolysaccharide modification accompanying antigenic conversion by phage P27. J. Gen. Microbiol. **107:**279–287.
14. **Mäkelä, P. H.** 1973. Glucosylation of lipopolysaccharide in *Salmonella*: mutants negative for O-antigen factor 122. J. Bacteriol. **116:**847–856.
15. **Mäkelä, P. H., and B. A. D. Stocker.** 1969. Genetics of polysaccharide synthesis. Annu. Rev. Genet. **3:**291–322.
16. **Mäkelä, P. H., and B. A. D. Stocker.** 1984. Genetics of lipopolysaccharide, p. 59–137. *In* E. T. Rietschel (ed.), Handbook of endotoxin, vol. I. Elsevier Science Publishers., Amsterdam, The Netherlands.
17. **Manning, P. A., U. H. Stroeher, and R. Morona.** 1993. Molecular basis for O-antigen biosynthesis in *Vibrio cholerae* O1: Ogawa-Inaba switching, p. 77–94. *In* I. K. Wachsmuth, P. Blake, and O. Olsvik (ed.), *Vibrio cholerae* and cholera. American Society for Microbiology, Washington, D.C.
18. **Perry, R. D., S. C. Straley, J. D. Fetherston, D. J. Rose, J. Gregor, and F. R. Blattner.** 1998. DNA sequencing and analysis of the low-Ca^{2+}-response plasmid pCD1 of *Yersinia pestis* KIM5. Infect. Immun. **66:**4611–4623.
19. **Popoff, M. Y., and L. L. Minor.** 1997. Antigenic formulas of the *Salmonella* serovars, 7th revision. W.H.O. Collaborating Centre for Reference and Research on *Salmonella*. Institut Pasteur, Paris, France.
20. **Popoff, M. Y., and L. L. Minor.** 1997. Guidelines for the preparation of *Salmonella* antisera. W.H.O. Collaborating Centre for Reference and Research on Salmonella. Institut Pasteur, Paris, France
21. **Reeves, P. R., L. Farnell, and R. Lan.** 1994. MULTICOMP: a program for preparing sequence data for phylogenetic analysis. Comput. Appl. Biol. Sci. **10:**281–284.
22. **Reeves, P. R., M. Hobbs, M. Valvano, M. Skurnik, C. Whitfield, D. Coplin, N. Kido, J. Klena, D. Maskell, C. Raetz, and P. Rick.** 1996. Bacterial polysaccharide synthesis and gene nomenclature. Trends Microbiol. **4:**495–503.
23. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
24. **Sanderson, K. E., A. Hessel, and K. E. Rudd.** 1995. Genetic map of *Salmonella typhimurium*, edition VIII. Microbiol. Rev. **59:**241–303.
25. **Slauch, J. M., A. A. Lee, M. J. Mahan, and J. J. Mekalanos.** 1996. Molecular characterization of the *OafA* locus responsible for acetylation of *Salmonella typhimurium* O-antigen: OafA is a member of a family of integral membrane trans-acylases. J. Bacteriol. **178:**5904–5909.
26. **Timmerman, K. P., and C. P. Tu.** 1985. Complete sequence of IS*3*. Nucleic Acids Res. **13:**2127–2139.
27. **Vander Byl, C., and A. M. Kropinski.** 2000. Sequence of the genome of *Salmonella* bacteriophage P22. J. Bacteriol. **182:**6472–6481.
28. **Xiang, S. H., M. Hobbs, and P. R. Reeves.** 1994. 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. J. Bacteriol. **176:**4357–4365.