

SUPPLEMENTARY DATA

Table S1. List of sequences from *E. coli* O-antigen biosynthesis gene clusters and SSI reference strains used in this study.

Table S2. Results of the pfam search of glycosyltransferases at *E. coli* O-antigen biosynthesis gene clusters.

Figure S1. The O-antigen biosynthesis gene clusters of all 184 *E. coli* O serogroups. One hundred eight O serogroups, O-AGC sequences of which were determined in this study, were marked with asterisks. Others were obtained from public databases. O-AGCs of five O serogroups (O26, O55, O103, O111 and O157) were extracted from complete genome sequences published (O26:AP010953.1, O55:CP001846.1, O103:AP010958.1, O111:AP010960.1 and O157:BA000007.2). Genes truncated by the length limitation of sequences used are indicated with parentheses.

Figure S2. Symmetric matrices of pairwise distances within *wzx*, *wzy*, *wzm* and *wzt* genes. Distances between amino acid sequences were calculated using ProtDist in Phylip package (ver 3.696) (Felsenstein, J. Department of Genome Sciences. University of Washington, Seattle; 2005) with JTT model from an amino acid alignment generated by MAFFT (v7.164) with L-INS-i option.

Figure S3. Forty-six O-antigen biosynthesis gene clusters inside ST10 and its close relatives. O-AGC sequences determined in this study were marked with asterisks. Others were obtained from public databases.

Figure S4. O-antigen biosynthesis gene clusters shared by *E. coli* and *Shigella* strains. Sequences of the *Shigella* O-antigen biosynthesis gene clusters were obtained from databases (see reference; Liu B. et al., FEMS Microbiol Rev, 32: 627-653, 2008). DNA sequence identities (%) between group members are indicated in each group. Two groups that were identified in this study are indicated by asterisk.

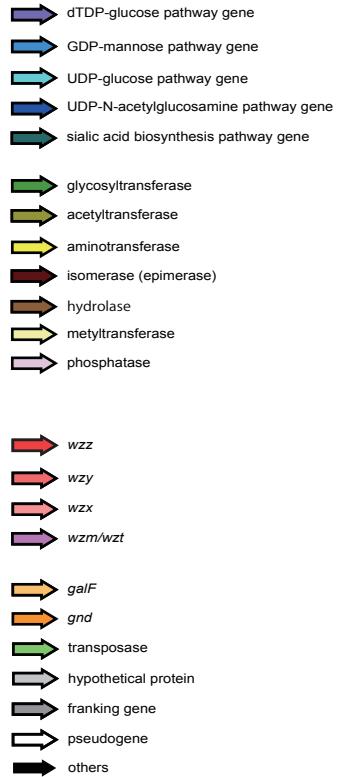
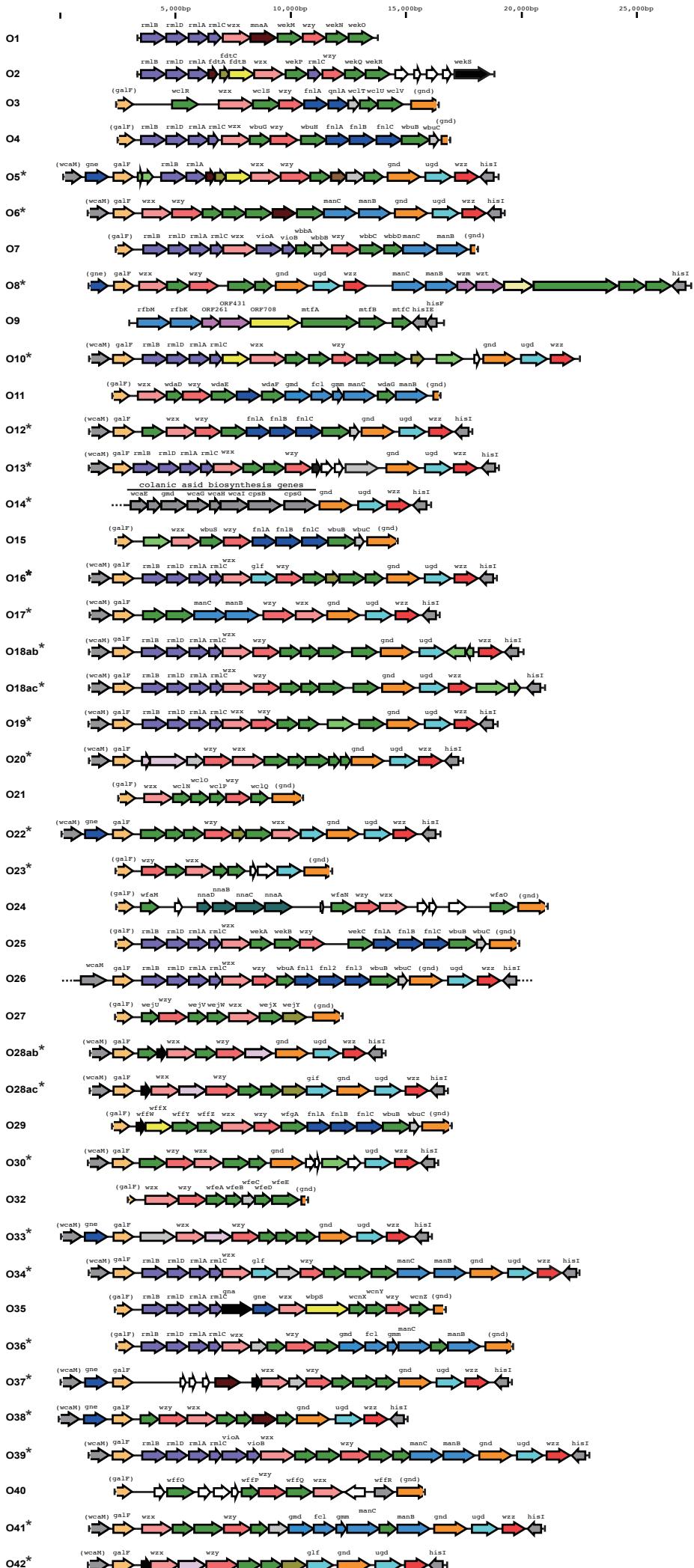


Fig. S1 (page 1)

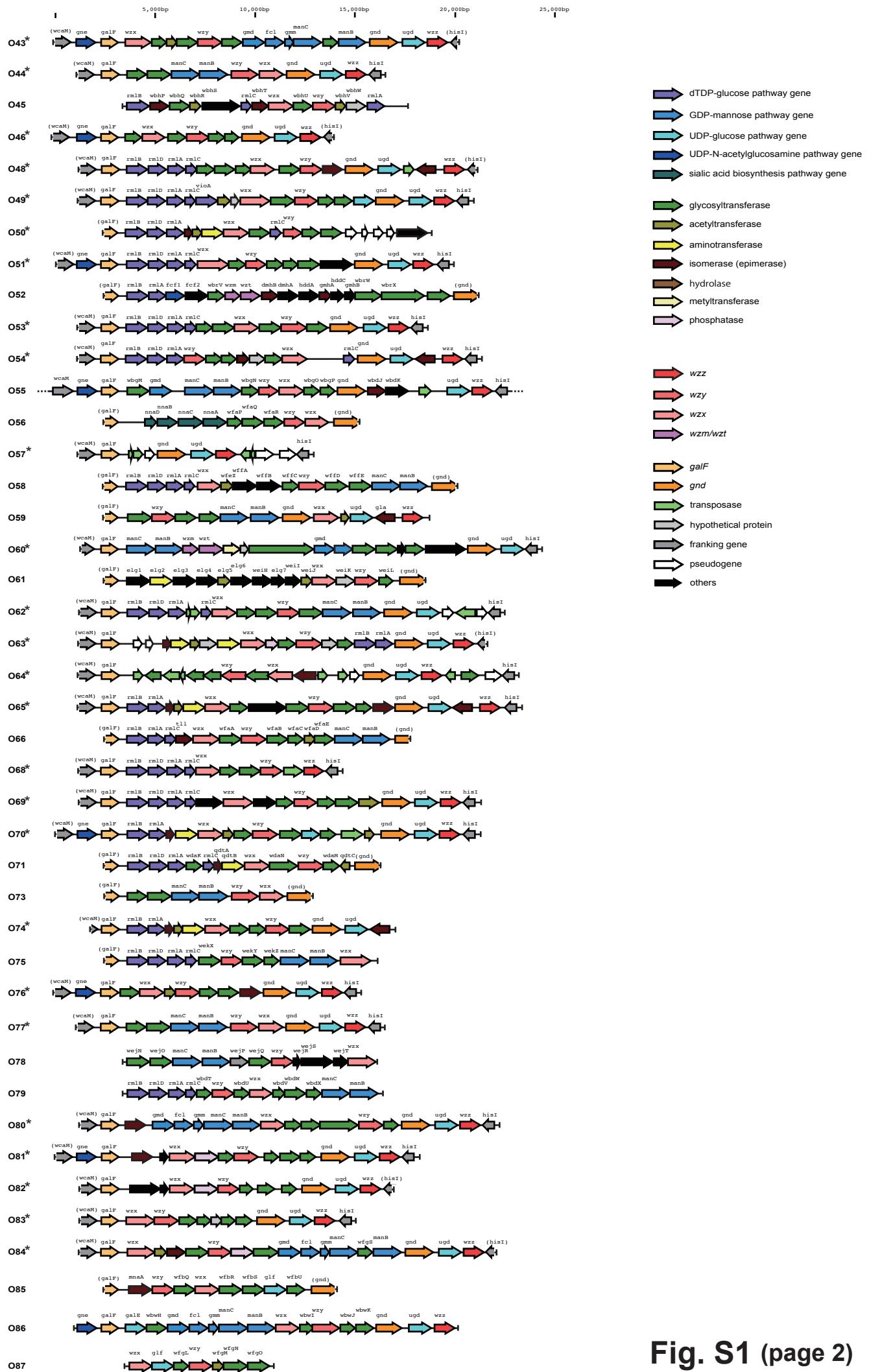


Fig. S1 (page 2)

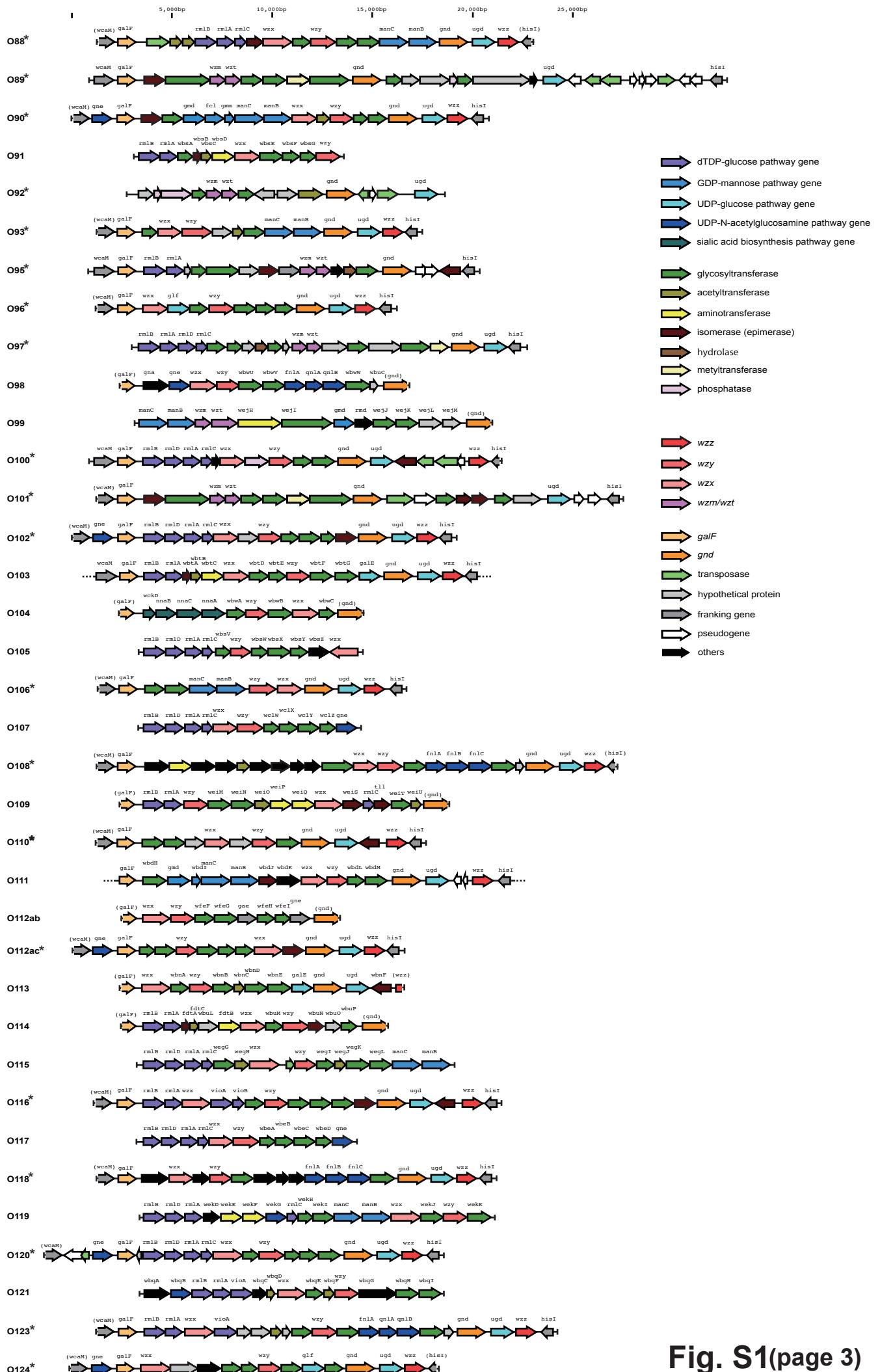


Fig. S1(page 3)

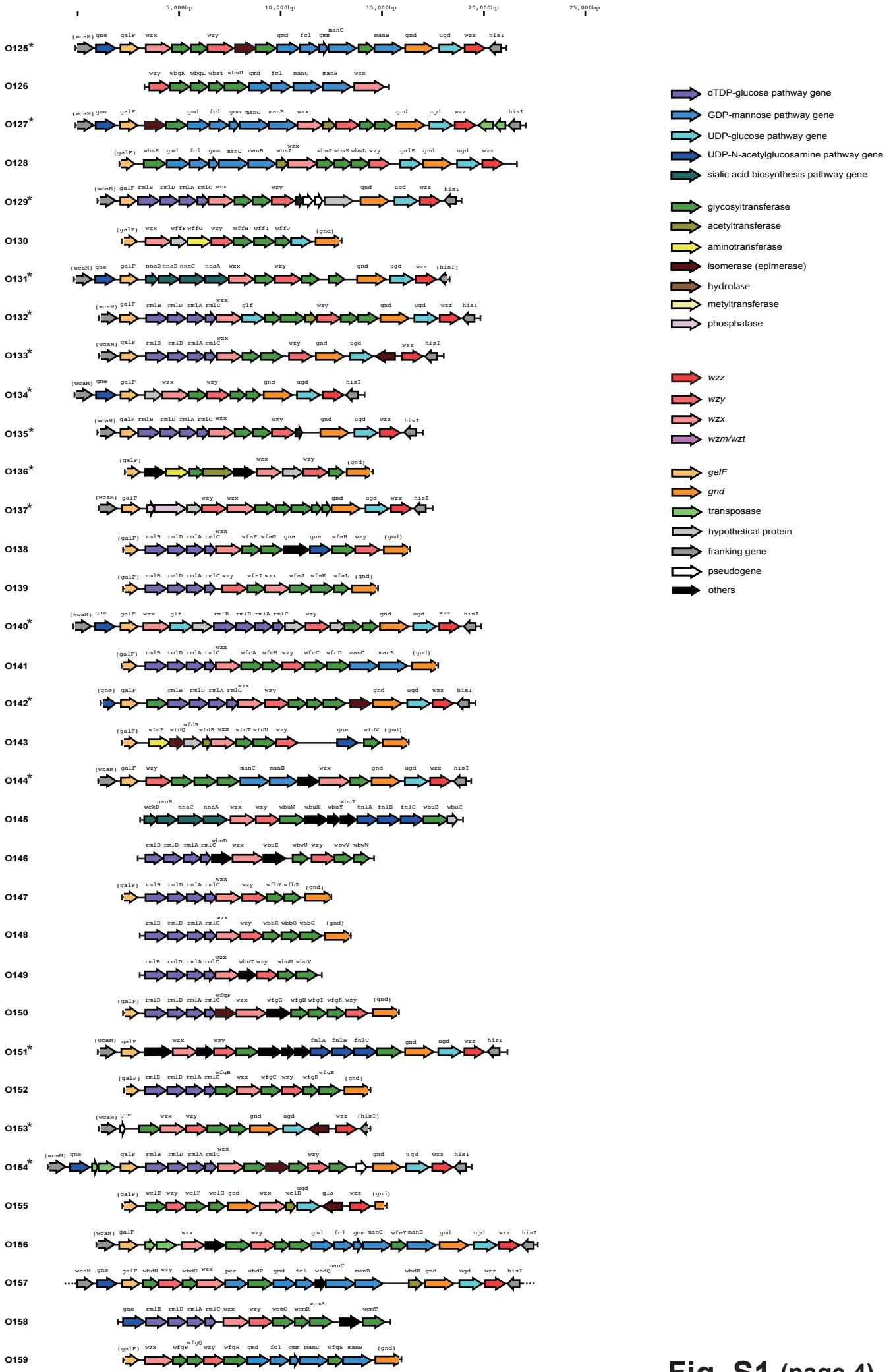


Fig. S1 (page 4)

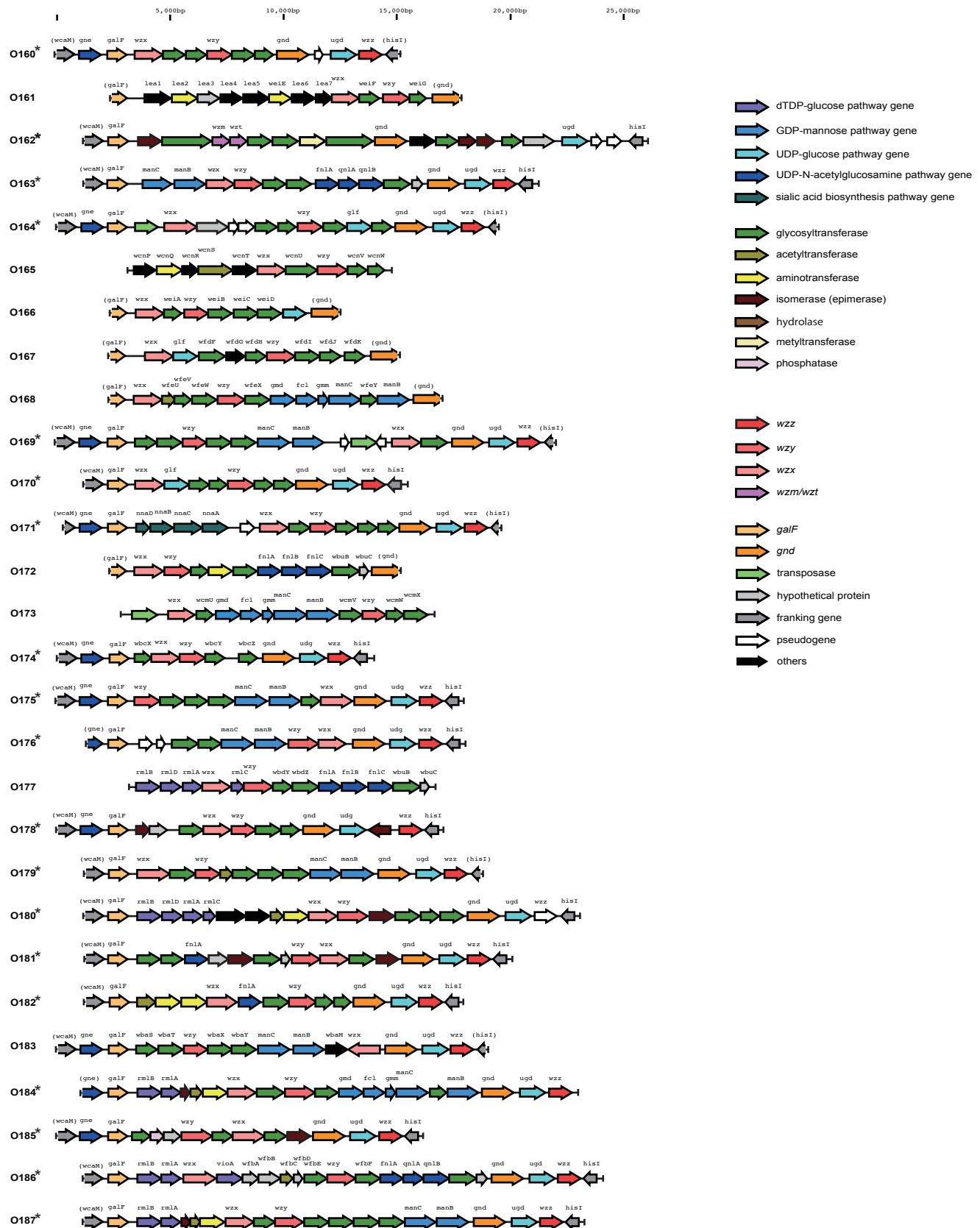


Fig. S1 (page 5)

Wzx

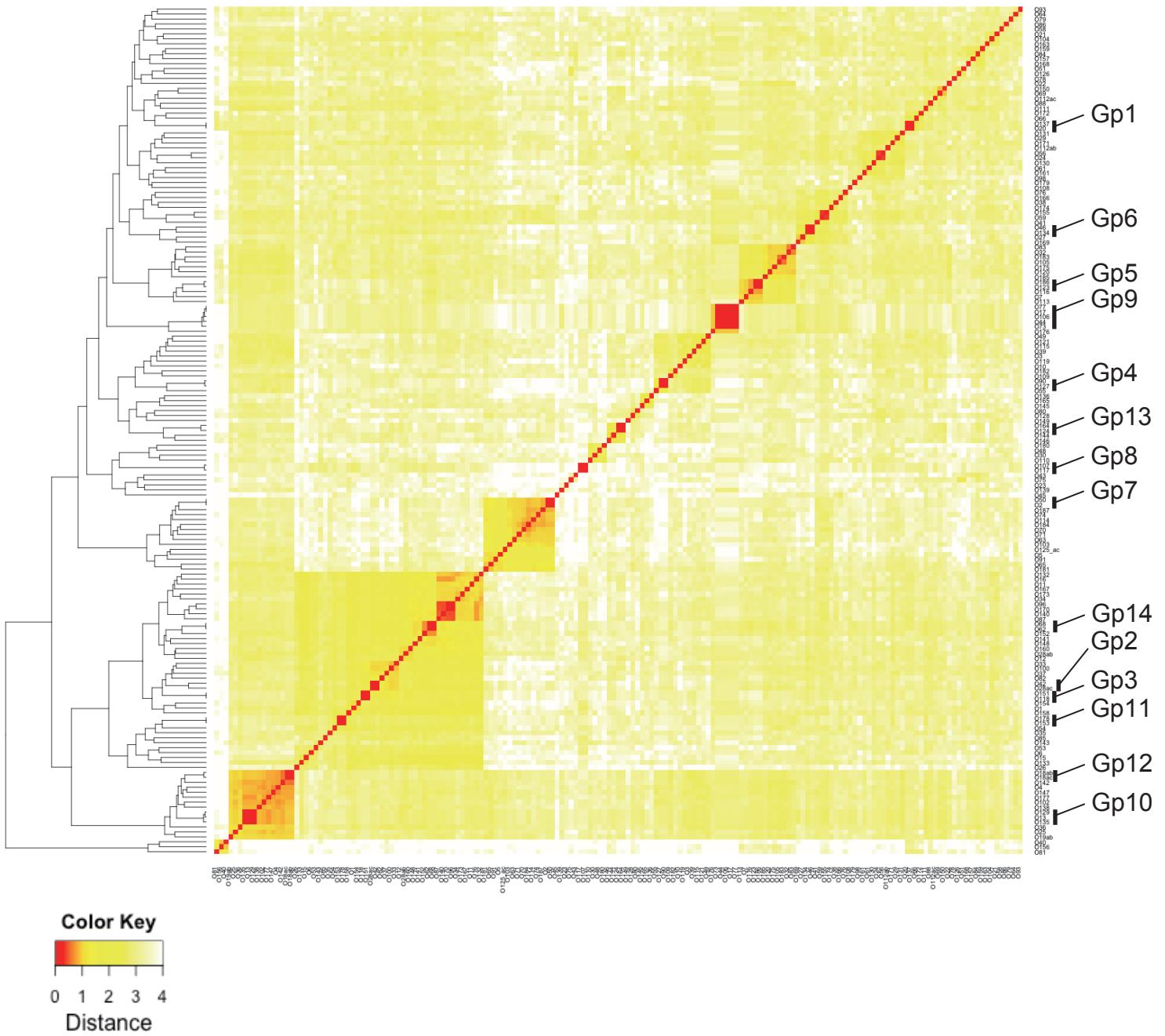
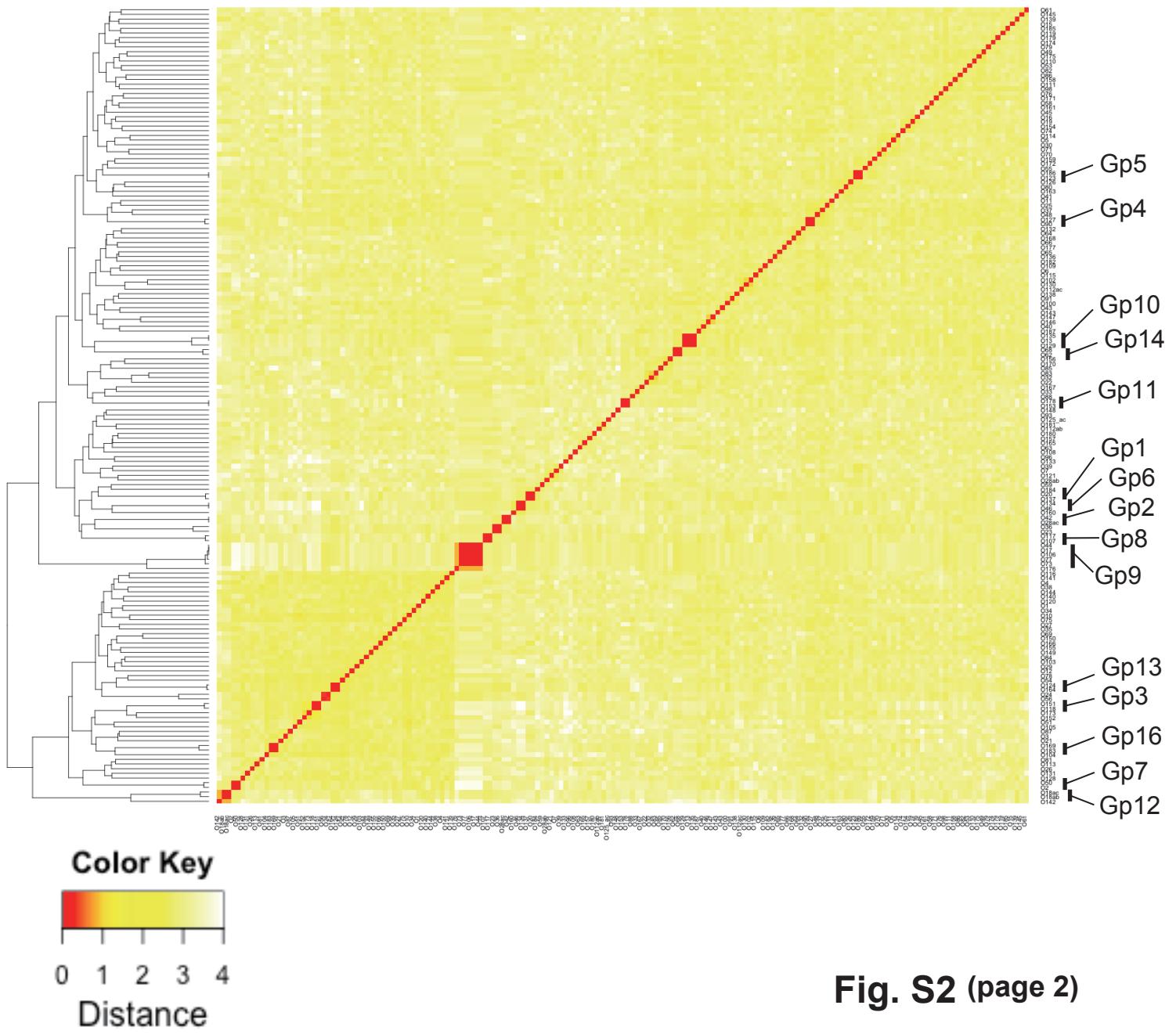
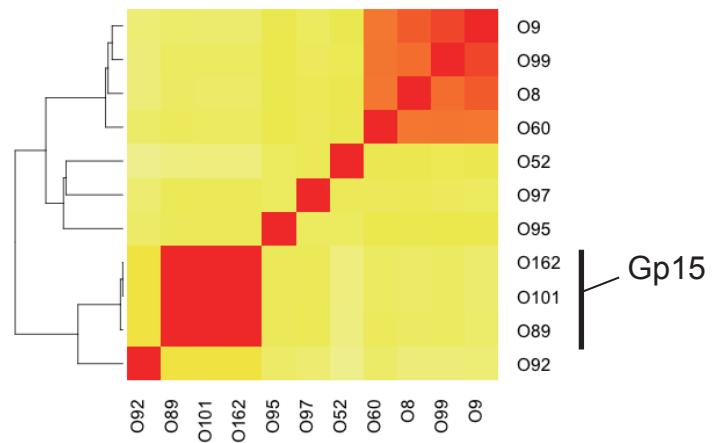


Fig. S2 (page 1)

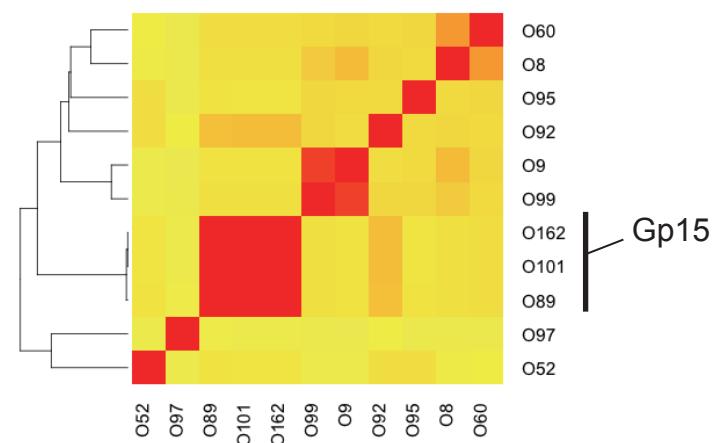
Wzy



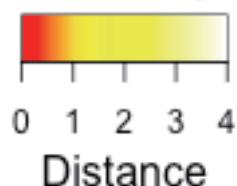
Wzm



Wzt



Color Key



0 1 2 3 4

Distance

Fig. S2 (page 3)

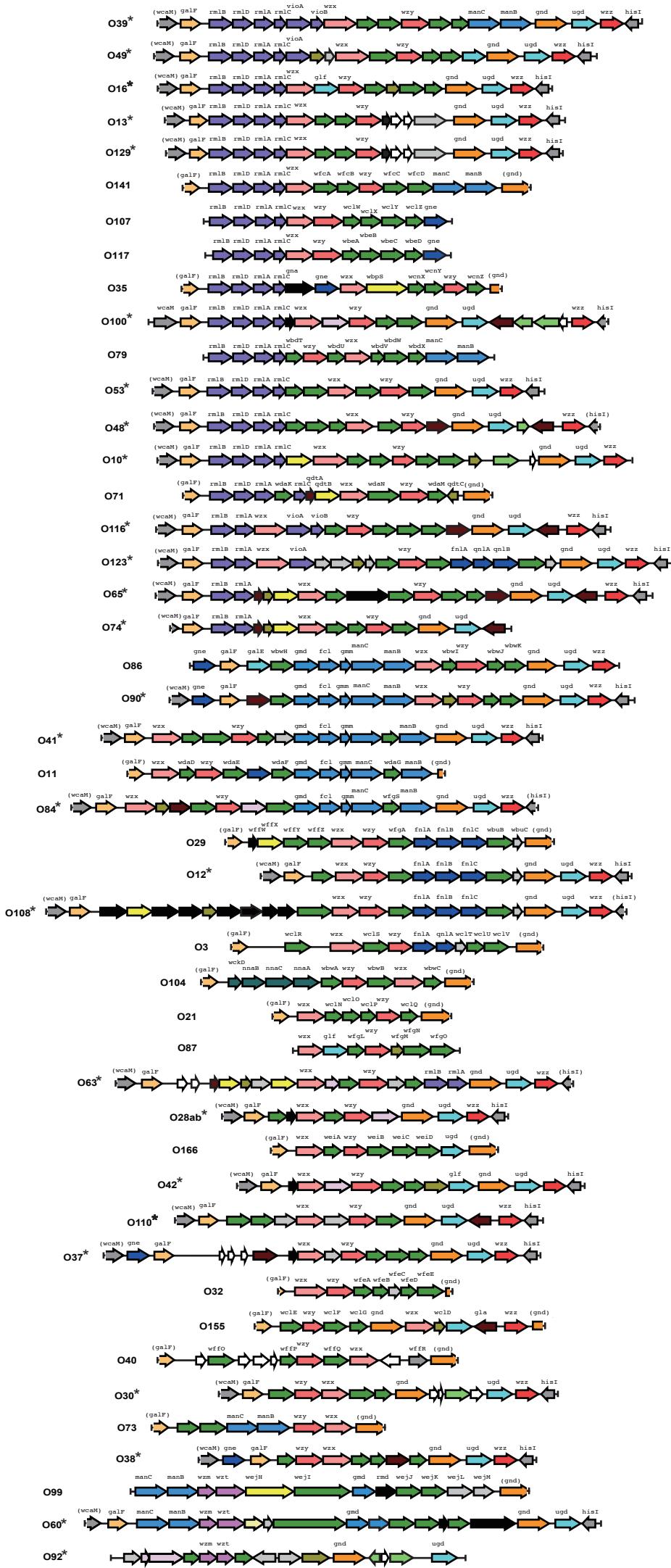


Fig. S3

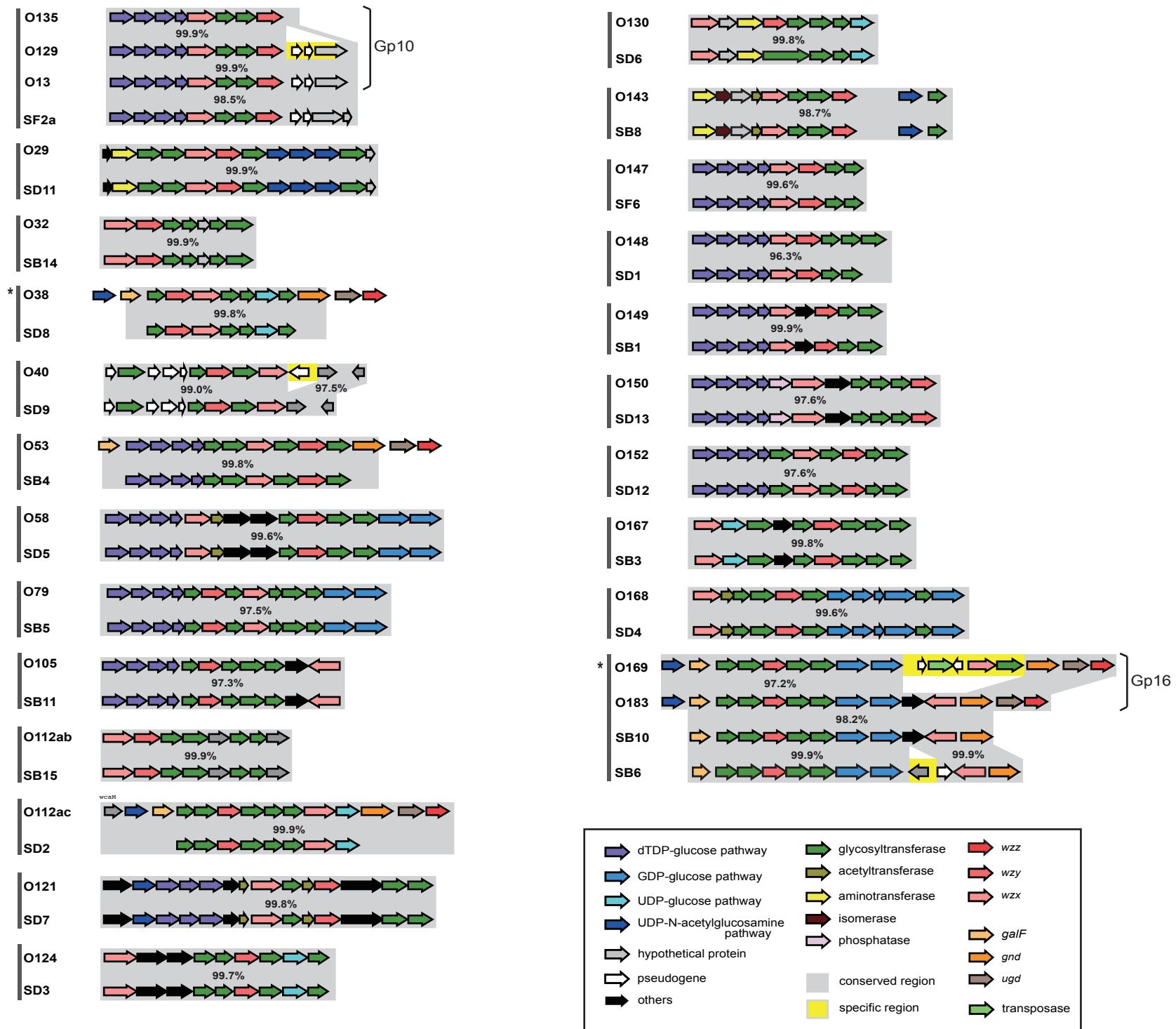


Fig. S4

Table S1. List of sequences from *E. coli* O-antigen biosynthesis gene clusters and SSI reference strains used in this study.

O serogroup	Accession no. of O-AGC sequence	Size*	Type of processing genes	Reference	Strain ID of SSI reference strains
O1	GU299791	10301 bp	wzx/wzy	Li D. et al. J. Microbiol. Methods 82 (1), 71-77 (2010)	U5-41
O2	GU299792	15394 bp	wzx/wzy	Li D. et al. J. Microbiol. Methods 82 (1), 71-77 (2010)	U9-41
O3	EU694097	14049 bp	wzx/wzy	Ren Y. et al. J. Microbiol. Methods 75 (2), 329-334 (2008)	U14-41
O4	AY568960	14433 bp	wzx/wzy	D'Souza J. M. et al. FEMS Microbiol. Lett. 244 (1), 27-32 (2005)	U4-41
O5	AB811596	19014 bp	wzx/wzy	in this study	U1-41
O6	AB811597	18341 bp	wzx/wzy	in this study	Bi7458-41
O7	AF125322	15943 bp	wzx/wzy	Marolda C. L. et al. Microbiology 145 (Pt 9), 2485-2495 (1999)	Bi7509-41
O8	AB811598	26523 bp	wzm/wzt	in this study	93404-41
O9	D43637	13316 bp	wzm/wzt	Kido N. et al. J. Bacteriol. 177 (8), 2178-2187 (1995)	Bi316-42
O10	AB811599	21154 bp	wzx/wzy	in this study	Bi8337-41
O11	HQ388393	14174 bp	wzx/wzy	Yang L. et al. FEMS Immunol. Med. Microbiol. 61 258-268 (2011)	Bi623-42
O12	AB811600	16805 bp	wzx/wzy	in this study	Bi626-42
O13	AB972413	18293 bp	wzx/wzy	in this study	Su4321-41
O14	AB972414	16727 bp	-	in this study	Su4411-41
O15	AY647261	11893 bp	wzx/wzy	Beutin L. et al. J. Clin. Microbiol. 43 (2), 703-710 (2005)	F7902-41
O16	AB811601	17903 bp	wzx/wzy	in this study	F11119-41
O17	AB812084	15422 bp	wzx/wzy	in this study	K12a
O18ab	AB811602	18945 bp	wzx/wzy	in this study	F10018-41
O18ac	AB811603	19922 bp	wzx/wzy	in this study	D-M3219-54
O19	AB811604	17909 bp	wzx/wzy	in this study	F8188-41
O20	AB811605	16391 bp	wzx/wzy	in this study	P7a
O21	EU694098	8041 bp	wzx/wzy	Ren Y. et al. J. Microbiol. Methods 75 (2), 329-334 (2008)	E19a
O22	AB811606	16589 bp	wzx/wzy	in this study	E14a
O23	AB811607	9332 bp	wzx/wzy	in this study	E39a
O24	DQ220292	18772 bp	wzx/wzy	Cheng J. et al. Curr. Microbiol. 53 (6), 470-476 (2006)	E41a
O25	GU014554	17547 bp	wzx/wzy	Wang Q. et al. J. Clin. Microbiol. 48 (6), 2066-2074 (2010)	E47a
O26	AF529080	13270 bp	wzx/wzy	D'Souza J. M. et al. Gene 297 (1-2), 123-127 (2002)	H311b
O27	GU014555	9510 bp	wzx/wzy	Wang Q. et al. J. Clin. Microbiol. 48 (6), 2066-2074 (2010)	F9884-41
O28ab	AB811608	14024 bp	wzx/wzy	in this study	K1a
O28ac	AB811609	15777 bp	wzx/wzy	in this study	Kattwijk
O29	EU294173	14249 bp	wzx/wzy	Wang Q. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	1085-87
O30	AB811610	15355 bp	wzx/wzy	in this study	P2a
O32	EU296410	7286 bp	wzx/wzy	Wang Q. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	P6a
O33	AB811611	16305 bp	wzx/wzy	in this study	E40
O34	AB811612	21470 bp	wzx/wzy	in this study	H304
O35	FJ940774	14060 bp	wzx/wzy	Liu B. et al. Vet. Microbiol. 142 (3-4), 373-378 (2010)	E77a
O36	AB811613	17217 bp	wzx/wzy	in this study	H502a
O37	AB811614	19104 bp	wzx/wzy	in this study	H510c
O38	AB811615	15239 bp	wzx/wzy	in this study	F11621-41
O39	AB811616	21940 bp	wzx/wzy	in this study	H7
O40	EU296417	13698 bp	wzx/wzy	Wang Q. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	H316
O41	AB811617	19935 bp	wzx/wzy	in this study	H710C
O42	AB811618	16901 bp	wzx/wzy	in this study	P11a
O43	AB811619	19971 bp	wzx/wzy	in this study	Bi7455-41
O44	AB811620	15423 bp	wzx/wzy	in this study	H702c
O45	AY771223	14483 bp	wzx/wzy	DebRoy C. et al. Appl. Environ. Microbiol. 71 (8), 4919-4924 (2005)	H61
O46	AB811621	13542 bp	wzx/wzy	in this study	P1c
O48	AB811622	19477 bp	wzx/wzy	in this study	U8-41
O49	AB811623	19664 bp	wzx/wzy	in this study	U12-41
O50	AB811624	16181 bp	wzx/wzy	in this study	U18-41
O51	AB812020	20072 bp	wzx/wzy	in this study	U19-41
O52	AY528413	18900 bp	wzm/wzt	Feng L. et al. J. Bacteriol. 186 (14), 4510-4519 (2004)	U20-41
O53	AB812021	17467 bp	wzx/wzy	in this study	Bi7327-41
O54	AB812085	20197 bp	wzx/wzy	in this study	Su3972-41
O55	AF461121	27730 bp	wzx/wzy	Wang Q. et al. J. Bacteriol. 184 (10), 2620-2625 (2002)	Su3912-41
O56	DQ220293	13075 bp	wzx/wzy	Cheng J. et al. Curr. Microbiol. 53 (6), 470-476 (2006)	Su3684-41
O57	AB972415	11798 bp	-	in this study	F8198-41
O58	EU294175	17770 bp	wzx/wzy	Wang Q. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	F8962-41
O59	AY654590	16573 bp	wzx/wzy	Guo H. et al. FEMS Microbiol. Lett. 248 (2), 153-161 (2005)	F9095-41
O60	AB812022	22813 bp	wzm/wzt	in this study	F10167a-41
O61	GU220362	15464 bp	wzx/wzy	Li X. et al. Carbohydr. Res. 345 (11), 1581-1587 (2010)	F10167a-42
O62	AB812023	21256 bp	wzx/wzy	in this study	F10524-41
O63	AB812024	19338 bp	wzx/wzy	in this study	F10598-41
O64	AB812025	22017 bp	wzx/wzy	in this study	K6b
O65	AB812026	22182 bp	wzx/wzy	in this study	K11a
O66	DQ069297	15137 bp	wzx/wzy	Cheng J. J. et al. Microbiol. 45 (1), 69-74 (2007)	P1a
O68	AB812027	13294 bp	wzx/wzy	in this study	P7d
O69	AB812028	20114 bp	wzx/wzy	in this study	P9b
O70	AB812029	20765 bp	wzx/wzy	in this study	P9c
O71	GU445927	13743 bp	wzx/wzy	Hu B. et al. FEMS Immunol. Med. Microbiol. 59 (2), 161-169 (2010)	P10a
O73	DQ000313	10486 bp	wzx/wzy	Feng L. et al. Microbiology 153 (1), 139-147 (2007)	P12a

O74	AB812030	15182 bp	wzx/wzy	in this study	E3a
O75	GU299795	12945 bp	wzx/wzy	Li D. et al. J. Microbiol. Methods 82 (1), 71-77 (2010)	E3b
O76	AB812031	15443 bp	wzx/wzy	in this study	E5d
O77	AB972416	15419 bp	wzx/wzy	in this study	E10
O78	FJ940775	12655 bp	wzx/wzy	Liu B. et al. Vet. Microbiol. 142 (3-4), 373-378 (2010)	E38
O79	EU294162	12892 bp	wzx/wzy	Liu B. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	E39
O80	AB812032	21073 bp	wzx/wzy	in this study	E71
O81	AB812033	18282 bp	wzx/wzy	in this study	H5
O82	AB812034	15306 bp	wzx/wzy	in this study	H14
O83	AB812035	13926 bp	wzx/wzy	in this study	H17a
O84	AB812036	20537 bp	wzx/wzy	in this study	H19
O85	GU299798	11203 bp	wzx/wzy	Perepelov A. V. et al. Innate Immunity 17(2), 164-173 (2011)	H23
O86	AY220982	19288 bp	wzx/wzy	Guo H. et al. Appl. Environ. Microbiol. 71 (12), 7995-8001 (2005)	H35
O87	EU294177	7267 bp	wzx/wzy	Liu B. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	H40
O88	AB812037	21331 bp	wzx/wzy	in this study	H53
O89	AB812038	32654 bp	wzm/wzt	in this study	H68
O90	AB812039	20886 bp	wzx/wzy	in this study	H77
O91	AY035396	10196 bp	wzx/wzy	Perelle S. et al. J. Appl. Microbiol. 93 (5), 758-764 (2002)	H307b
O92	AB812040	16295 bp	wzm/wzt	in this study	H308a
O93	AB812041	16390 bp	wzx/wzy	in this study	2885-2
O95	AB812042	20111 bp	wzm/wzt	in this study	H311a
O96	AB812043	14195 bp	wzx/wzy	in this study	H319
O97	AB812044	19827 bp	wzm/wzt	in this study	H320a
O98	DQ180602	14333 bp	wzx/wzy	Cunneen M. M. and Reeves P. R. Mol. Biol. Evol. 24 (6), 1355-1365 (2007)	H501d
O99	FJ940773	17685 bp	wzm/wzt	Perepelov A. V. et al. FEMS Immunol. Med. Microbiol. 57 (1), 80-87 (2009)	H504c
O100	AB812045	20808 bp	wzx/wzy	in this study	H509a
O101	AB812046	26407 bp	wzm/wzt	in this study	H510a
O102	AB812047	19219 bp	wzx/wzy	in this study	H511
O103	AY532664	12033 bp	wzx/wzy	Fratamico P. M. et al. Can. J. Microbiol. 51 (6), 515-522 (2005)	H515b
O104	AF361371	12105 bp	wzx/wzy	Wang L. et al. Gene 270 (1-2), 231-236 (2001)	H519
O105	EU294171	11106 bp	wzx/wzy	Wang Q. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	H520b
O106	AB972417	15419 bp	wzx/wzy	in this study	H521a
O107	EU694095	10884 bp	wzx/wzy	Wang Q. et al. FEMS Immunol. Med. Microbiol. 55 (1), 47-54 (2009)	H705
O108	AB812048	25592 bp	wzx/wzy	in this study	H708b
O109	HM485572	16597 bp	wzx/wzy	Perepelov A. V. et al. FEMS Immunol. Med. Microbiol. 61 (1), 47-53 (2011)	H709c
O110	AB812049	16588 bp	wzx/wzy	in this study	H711c
O111	AF078736	14516 bp	wzx/wzy	Bastin D. A. and Reeves P. R. Gene 164 (1), 17-23 (1995)	Stoke W
O112ab	EU296413	10812 bp	wzx/wzy	Liu B. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	32W
O112ac	AB812050	16668 bp	wzx/wzy	in this study	Guanabara 1685
O113	AF172324	14263 bp	wzx/wzy	Paton A. W. and Paton J. C. Infect. Immun. 67 (11), 5930-5937 (1999)	6182-50
O114	AY573377	13272 bp	wzx/wzy	Feng L. et al. J. Clin. Microbiol. 42 (8), 3799-3804 (2004)	26w
O115	GU068041	15588 bp	wzx/wzy	Wang Q. et al. Mol. Cell. Probes 24 (5), 286-290 (2010)	27w
O116	AB812051	20297 bp	wzx/wzy	in this study	28w
O117	DQ465247	10886 bp	wzx/wzy	Liu Y. et al. Mol. Cell. Probes 21 (4), 295-302 (2007)	30w
O118	AB972418	20533 bp	wzx/wzy	in this study	31w
O119	GQ499368	17610 bp	wzx/wzy	Liu B. et al. Vet. Microbiol. 142 (3-4), 373-378 (2010)	34w
O120	AB812052	19924 bp	wzx/wzy	in this study	35w
O121	AY208937	15155 bp	wzx/wzy	Fratamico P. M. et al. J. Clin. Microbiol. 41 (7), 3379-3383 (2003)	39w
O123	AB972419	22904 bp	wzx/wzy	in this study	43w
O124	AB972420	17893 bp	wzx/wzy	in this study	Ew227
O125	AB812053	21385 bp	wzx/wzy	in this study	Ew2129-54
O126	DQ465248	11783 bp	wzx/wzy	Liu Y. et al. Mol. Cell. Probes 21 (4), 295-302 (2007)	E611
O127	AB812054	22349 bp	wzx/wzy	in this study	4932-53
O128	AY217096	19013 bp	wzx/wzy	Shao J. et al. FEBS Lett. 553 (1-2), 99-103 (2003)	56-54
O129	AB972421	18333 bp	wzx/wzy	in this study	178-54
O130	EU296421	10990 bp	wzx/wzy	Liu B. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	Ew4866-53
O131	AB812055	18140 bp	wzx/wzy	in this study	S239
O132	AB812056	18692 bp	wzx/wzy	in this study	N87
O133	AB812057	16946 bp	wzx/wzy	in this study	N282
O134	AB812058	14265 bp	wzx/wzy	in this study	4370-53
O135	AB972422	16164 bp	wzx/wzy	in this study	Coli Pecs
O136	AB812059	12191 bp	wzx/wzy	in this study	1111-55
O137	AB972423	16392 bp	wzx/wzy	in this study	RVC1787
O138	DQ109551	14139 bp	wzx/wzy	Wang L. et al. Vet. Microbiol. 111 (3-4), 181-190 (2005)	CDC62-57
O139	DQ109552	12507 bp	wzx/wzy	Wang L. et al. Vet. Microbiol. 111 (3-4), 181-190 (2005)	CDC63-57
O140	AB812060	20161 bp	wzx/wzy	in this study	CDC149-51
O141	DQ868765	15601 bp	wzx/wzy	Han W. et al. Appl. Environ. Microbiol. 73 (12), 4082-4088 (2007)	RVC2907
O142	AB812061	18883 bp	wzx/wzy	in this study	C771
O143	EU294164	14105 bp	wzx/wzy	Liu B. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	4608-58
O144	AB812062	18384 bp	wzx/wzy	in this study	1624-56
O145	AY647260	15475 bp	wzx/wzy	Feng L. et al. J. Bacteriol. 187 (2), 758-764 (2005)	E1385-3
O146	DQ465249	11888 bp	wzx/wzy	Liu Y. et al. Mol. Cell. Probes 21 (4), 295-302 (2007)	CDC2959-54
O147	DQ868766	10319 bp	wzx/wzy	Han W. et al. Appl. Environ. Microbiol. 73 (12), 4082-4088 (2007)	G1253
O148	DQ167407	10241 bp	wzx/wzy	Feng L. et al. Microbiology 153 (1), 139-147 (2007)	E519-66
O149	DQ868764	8729 bp	wzx/wzy	Han W. et al. Appl. Environ. Microbiol. 73 (12), 4082-4088 (2007)	D616
O150	EU294168	13552 bp	wzx/wzy	Liu B. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	1935
O151	AB972424	20446 bp	wzx/wzy	in this study	880-67

O152	EU294170	11575 bp	wzx/wzy	Liu B. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	1184-68
O153	AB812063	12998 bp	wzx/wzy	in this study	14097
O154	AB812064	20962 bp	wzx/wzy	in this study	E1541-68
O155	AY657020	12755 bp	wzx/wzy	Guo H. et al. FEMS Microbiol. Lett. 248 (2), 153-161 (2005)	E1529-68
O156	AB812065	21656 bp	wzx/wzy	in this study	E1585-68
O157	AF061251	14002 bp	wzx/wzy	Wang L and Reeves P. R. Infect. Immun. 66 (8), 3545-3551 (1998)	A2
O158	GU068044	14597 bp	wzx/wzy	Wang Q. et al. Mol. Cell. Probes 24 (5), 286-290 (2010)	E1020-72
O159	EU294176	13749 bp	wzx/wzy	Liu B. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	E2476-72
O160	AB812066	14841 bp	wzx/wzy	in this study	E110-69
O161	GU220361	15878 bp	wzx/wzy	Li X. et al. Carbohydr Res. 19;345(11):1581-1587	E110-69
O162	AB812067	25001 bp	wzm/wzt	in this study	10B-1
O163	AB812068	20184 bp	wzx/wzy	in this study	SN3B-1
O164	AB972425	19387 bp	wzx/wzy	in this study	SC647
O165	GU068045	11292 bp	wzx/wzy	Wang Q. et al. Mol. Cell. Probes 24 (5), 286-290 (2010)	E78634
O166	GU299794	10140 bp	wzx/wzy	Liu B. Microbiology 156 (6), 1642-1649 (2010)	3866-54
O167	EU296408	12864 bp	wzx/wzy	Liu B. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	E10702
O168	EU296403	14804 bp	wzx/wzy	Liu B. et al. FEMS Microbiol. Rev. 32 (4), 627-653 (2008)	E10710
O169	AB812069	21813 bp	wzx/wzy	in this study	1792-54
O170	AB812070	14143 bp	wzx/wzy	in this study	745-56
O171	AB812071	18684 bp	wzx/wzy	in this study	244-54
O172	AY545992	12850 bp	wzx/wzy	Guo H. et al. J. Appl. Microbiol. 97 (1), 181-190 (2004)	3288-85
O173	GU068046	13697 bp	wzx/wzy	Wang Q. et al. Mol. Cell. Probes 24 (5), 286-290 (2010)	L119B-10
O174	AB812072	14063 bp	wzx/wzy	in this study	C8/55
O175	AB812073	17948 bp	wzx/wzy	in this study	C12/55
O176	AB812074	17248 bp	wzx/wzy	in this study	E29518
O177	DQ008593	13198 bp	wzx/wzy	Beutin L. et al. J. Clin. Microbiol. 43 (10), 5143-5149 (2005)	E40874
O178	AB812075	17202 bp	wzx/wzy	in this study	E54071
O179	AB812076	17188 bp	wzx/wzy	in this study	E43478
O180	AB812077	21994 bp	wzx/wzy	in this study	86-381
O181	AB812078	19016 bp	wzx/wzy	in this study	92-1250
O182	AB812079	16810 bp	wzx/wzy	in this study	99-1287
O183	AB627352	18026 bp	wzx/wzy	Iguchi A. et al. J. Clin. Microbiol. 49 (10), 3678-3680 (2011)	99-2442
O184	AB812080	22184 bp	wzx/wzy	in this study	99-4473
O185	AB812081	16331 bp	wzx/wzy	in this study	99-6301
O186	AB812082	22931 bp	wzx/wzy	in this study	182-02
O187	AB812083	22170 bp	wzx/wzy	in this study	559-59

* registered sequence size including the O-AGC and in some cases its flanking regions.

Table S2. Results of the pfam search of glycosyltransferases at *E. coli* O-AGCs.

pfam ID	accession	no. of hit
Glycos_transf_2	PF00535	255
Glycos_transf_1	PF00534	232
Glyco_transf_4	PF13439	58
Glyco_transf_4_2	PF13477	47
Glyco_transf_1_4	PF13692	40
Glyco_transf_4_4	PF13579	31
Glyco_transf_2_3	PF13632	7
Glyco_transf_11	PF01531	7
Glyco_transf_25	PF01755	6
Glyco_transf_52	PF07922	5
Glyco_tranf_2_3	PF13641	4
Glyco_tran_WecB	PF03808	4
Glyco_tranf_2_2	PF10111	3
Glyco_tran_WbsX	PF14307	3
Gly_transf_sug	PF04488	3
Glyco_tran_28_C	PF04101	2
Glyco_transf_1_2	PF13524	2
Galactosyl_T	PF01762	1
Glycos_transf_4	PF00953	1
Glyco_tranf_2_4	PF13704	1
Glyco_tranf_2_5	PF13712	1
Glyco_transf_21	PF13506	1
Glyco_transf_5	PF08323	1
Glyco_transf_6	PF03414	1
Glyco_transf_7C	PF02709	1

Pfam search (ver. 26.0) (Finn RD et al. Nucleic Acids Res 38: D211–222, 2010) was performed with an E-value threshold of 0.01.