

lementary Table 4

Sequence	LOS classes												Cluster A GT-42				Cluster B GT-42			NeuB	Clade
	I	II	III	IV	VI	VII	VIII	IX	X	XI	XII	cstI (G1)	cstIII (G2)	cstVII (G3)	cstII (G4)	cstV (G5)	cstVI (G6)	cstIV (G7)			
ERR083880	0	0	100	0	0	6	0	0	23	17	0	0	0	0	0	1	0	0	0	1	
ERR083884	0	0	100	0	0	6	6	0	23	17	0	0	0	0	0	1	0	0	0	1	
ERR083912	14	0	9	0	0	0	94	38	0	8	0	0	0	1	0	0	0	1	1		
ERR083913	0	0	100	0	0	6	6	0	23	17	0	0	0	0	0	1	0	0	0	1	
ERR083924	0	0	100	0	0	6	6	0	23	17	0	0	0	0	0	1	0	0	0	1	
ERR083928	0	0	100	0	7	6	6	0	23	17	0	0	0	0	0	1	0	0	0	1	
ERR083953	0	0	100	0	0	6	6	0	23	17	0	0	0	0	0	1	0	0	0	1	
ERR083996	0	0	100	0	0	6	0	0	15	0	0	0	0	0	0	1	0	0	0	1	
ERR084002	0	92	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	
ERR084011	14	0	0	0	0	0	6	0	0	0	0	0	0	1	0	0	0	1	1		
ERR084042	0	0	91	0	0	6	6	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR084055	0	88	0	0	0	0	6	0	0	0	0	0	0	0	0	0	1	0	0	1	
ERR084056	0	92	0	0	0	0	6	0	0	0	0	0	0	0	0	0	1	0	0	1	
ERR084059	0	0	91	0	0	6	6	0	23	17	0	0	0	0	0	1	0	0	0	1	
ERR084071	0	100	0	0	0	0	6	0	0	0	0	0	0	0	0	0	1	0	0	1	
ERR084082	0	0	100	0	0	6	6	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR084083	0	0	91	0	0	6	6	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR084085	0	0	100	0	0	6	6	0	23	17	0	0	0	0	0	1	0	0	0	1	
ERR084115	0	0	91	0	0	6	6	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR084124	0	0	100	0	0	6	6	0	23	17	0	0	0	0	0	1	0	0	0	1	
ERR084134	0	100	0	0	0	0	6	0	0	0	0	0	0	0	0	0	1	0	0	1	
ERR084148	0	100	0	0	0	0	6	0	0	0	0	0	0	0	0	0	1	0	0	1	
ERR108323	0	100	0	0	0	0	6	0	0	0	0	0	0	0	0	0	1	0	0	1	
ERR1147218	0	92	0	0	0	0	6	0	0	0	0	0	0	0	0	0	1	0	0	1	
ERR1147219	0	88	0	0	0	0	6	0	0	0	0	0	0	0	0	0	1	0	0	1	
ERR1163389	0	0	91	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163390	0	0	91	0	0	6	0	0	15	8	0	0	0	0	0	1	0	0	0	1	
ERR1163391	0	0	91	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163392	0	96	0	0	0	0	6	0	0	0	0	0	0	0	0	0	1	0	0	1	
ERR1163393	0	8	9	0	0	6	0	38	0	0	0	0	0	1	0	0	0	1	1		
ERR1163397	0	0	91	0	0	6	0	0	15	8	0	0	0	0	0	1	0	0	0	1	
ERR1163398	0	0	91	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163399	0	0	91	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163400	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163401	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163403	0	0	91	0	0	6	6	0	15	8	0	0	0	0	0	1	0	0	0	1	
ERR1163404	0	0	100	0	0	6	0	0	15	17	0	0	0	0	0	1	0	0	0	1	
ERR1163405	0	0	91	0	0	6	6	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163406	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163407	0	0	91	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163410	0	0	91	0	0	6	6	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163412	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163415	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163418	0	0	100	0	0	6	6	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163419	0	0	91	0	0	6	6	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163420	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163421	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163422	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163423	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163426	0	0	91	0	0	6	6	0	15	8	0	0	0	0	0	1	0	0	0	1	
ERR1163428	0	0	100	0	0	6	6	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163429	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163430	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163431	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163435	0	0	91	0	0	6	19	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163436	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163437	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163438	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163439	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163440	0	0	100	0	0	6	6	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163442	0	0	100	0	7	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163444	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163445	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163446	0	0	91	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163447	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163451	0	8	9	0	0	6	0	38	0	0	0	0	0	1	0	0	0	1	1		
ERR1163452	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163453	0	0	91	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163454	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163455	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163458	0	0	91	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	
ERR1163460	0	0	91	0	0	6	0	0	23	8	0	0	0	0	0	1	0	0	0	1	

ERR1163574	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163575	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163576	0	0	91	0	0	0	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163577	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163578	0	0	100	0	0	0	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163579	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163580	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163581	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163582	0	0	91	0	0	0	0	0	0	8	0	0	0	0	0	0	1	0	0	1
ERR1163583	0	0	91	0	0	0	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163584	0	0	73	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163585	0	0	91	0	0	0	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163587	0	0	82	0	0	0	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163588	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163589	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163590	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163591	0	0	91	0	0	0	0	0	0	8	0	0	0	0	0	0	1	0	0	1
ERR1163610	14	100	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	1	0	1
ERR1163618	0	0	91	0	0	6	6	0	15	8	0	0	0	0	0	0	1	0	0	1
ERR1163619	0	0	91	0	0	6	6	0	15	8	0	0	0	0	0	0	1	0	0	1
ERR1163620	0	0	91	0	0	6	6	0	23	8	0	0	0	0	0	0	1	0	0	1
ERR1163621	0	0	91	0	0	6	6	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163623	0	0	91	0	0	6	0	38	15	8	0	0	0	0	0	0	1	0	1	1
ERR1163625	0	0	91	0	0	0	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163627	0	0	82	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163633	0	0	82	0	0	0	0	0	0	8	0	0	0	0	0	0	1	0	0	1
ERR1163634	0	96	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	1	0	1
ERR1163635	0	0	91	0	0	6	0	0	15	8	0	0	0	0	0	0	1	0	0	1
ERR1163637	0	0	91	0	0	6	6	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163641	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163642	0	0	91	0	0	0	0	0	0	8	0	0	0	0	0	0	1	0	0	1
ERR1163644	0	0	91	0	0	6	6	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163646	29	0	64	0	0	0	19	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163648	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163649	0	0	0	7	0	0	6	38	0	0	0	0	0	1	0	0	0	0	1	1
ERR1163654	0	96	0	0	0	0	0	38	0	0	0	1	0	0	0	0	0	1	1	1
ERR1163656	0	88	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	1	0	1
ERR1163663	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163664	0	88	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	1	0	1
ERR1163665	0	0	91	0	0	6	6	0	23	8	0	0	0	0	0	0	1	0	0	1
ERR1163667	0	0	91	0	0	6	6	0	15	8	0	0	0	0	0	0	1	0	0	1
ERR1163671	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163673	0	0	91	0	0	6	6	0	15	8	0	0	0	0	0	0	1	0	0	1
ERR1163674	0	0	91	0	0	6	6	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163675	0	0	91	0	0	6	6	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163678	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163681	0	0	91	0	0	6	6	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163682	0	88	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	1	0	1
ERR1163683	0	0	91	0	0	6	6	0	23	8	0	0	0	0	0	0	1	0	0	1
ERR1163686	0	0	82	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163687	0	0	91	0	0	6	0	0	15	8	0	0	0	0	0	0	1	0	0	1
ERR1163688	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163689	0	0	91	0	0	6	6	0	15	8	0	0	0	0	0	0	1	0	0	1
ERR1163690	0	0	91	0	0	6	6	0	15	8	0	0	0	0	0	0	1	0	0	1
ERR1163691	0	0	91	0	0	6	6	0	23	8	0	0	0	0	0	0	1	0	0	1
ERR1163692	0	0	91	0	0	6	0	0	23	8	0	0	0	0	0	0	1	0	0	1
ERR1163694	0	0	91	0	0	6	0	0	23	8	0	0	0	0	0	0	1	0	0	1
ERR1163695	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163696	0	0	91	0	0	6	0	0	23	8	0	0	0	0	0	0	1	0	0	1
ERR1163701	0	0	91	0	0	6	0	0	15	8	0	0	0	0	0	0	1	0	0	1
ERR1163702	0	96	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	1	0	1
ERR1163703	0	0	91	0	0	6	6	0	23	8	0	0	0	0	0	0	1	0	0	1
ERR1163704	0	0	91	0	0	6	6	0	15	8	0	0	0	0	0	0	1	0	0	1
ERR1163705	0	0	91	0	0	6	6	0	15	8	0	0	0	0	0	0	1	0	0	1
ERR1163706	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163709	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	0	1	0	0	1
ERR1163714	0	0	91	0	0	6	6	0	23	8	0	0	0	0	0	0	1	0	0	1
ERR1163715	0	0	91	0	0	6	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163716	0	0	73	0	0	0	0	0	8	8	0	0	0	0	0	0	1	0	0	1
ERR1163719	0	0	91	0	0	6	0	0	15	8	0	0	0	0	0	0	1	0	0	1
ERR1163722	0	100	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	1	0	1
ERR1163729	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	0	1	0	0	1
ERR1163730	0	96	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	1	0	1
ERR1163734	0	0	100	0	0	6	0	0	23	8	0	0	0	0	0	0	1	0	0	1
ERR1163741	0	0	91	0	0	6	6	0	8	8	0	0	0	0	0	0	1	0	0	1

Supplementary Table 5

Gene	Non-unique Gene name	Annotation	Orthologue	coverage	%nt id to <i>C. jejuni</i>
Group 2					
ERR1163614	XIII				
group_7984	pglH_2	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)-diNAcBac-PP-undecapreno	4055	50	76
gspA		General stress protein A	4003	N/A	N/A
group_7985	epsH	Putative glycosyltransferase EpsH	4056	36	82
group_7986	cstIII	GT-42 glycosyltransferase	4057	100	82
legI_1		N,N'-diacetyllegionaminic acid synthase	4155	100	83
group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014	99	87
group_7987		hypothetical protein	4058	98	84
neuA_2		N-acylneuraminate cytidyltransferase	4059	96	77
ERR1163713	XIV				
pglH_1		GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)-diNAcBac-PP-undecapreno	4002	69	73
gspA		General stress protein A	4003	N/A	N/A
group_7985	epsH	Putative glycosyltransferase EpsH	4056	36	82
group_7986	cstIII	GT-42 glycosyltransferase	4057	100	82
legI_1		N,N'-diacetyllegionaminic acid synthase	4155	100	83
group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014	99	87
group_7987		hypothetical protein	4058	98	84
neuA_2		N-acylneuraminate cytidyltransferase	4059	96	77
SRR5152313	incomplete				
group_10079	epsJ_4	putative glycosyltransferase EpsJ	4188	100	100
group_10078		hypothetical protein	4186	100	100
group_10077	epsJ_3	putative glycosyltransferase EpsJ	4187	100	100
?			?	100	100
group_10088	cstIII	GT-42 glycosyltransferase	4159	100	100
legI_2		N,N'-diacetyllegionaminic acid synthase	4181	100	100
siaA		UDP-N-acetylglucosamine 2-epimerase	4182	100	100
neuA		N-acylneuraminate cytidyltransferase	4183	100	100
group_10086		hypothetical protein	4184	100	100

Color code

C. coli clade 1*C. coli* clade 2*C. coli* clade 3

GT42 group 2 cstIII

GT42 group 4 cstII

GT42 group 5 cstV

GT42 group 6 cstVI

GT42 group 7 cstIV

Sialic acid biosynthesis genes

Capsule genes

N/A: No significant similarity found

Gene	Non-unique Gene name	Annotation	Orthologue	coverage	%nt id to <i>C. jejuni</i>
Group 4					
ERR1163434	XV				
group_5652	epsJ_1	putative glycosyltransferase EpsJ	4009	98	81
group_7342		hypothetical protein	4010	98	87
group_7343	epsH	Putative glycosyltransferase EpsH	4011	99	84
group_7344	cstII	GT-42 glycosyltransferase	4012	97	87
legI_1		N,N'-diacetyllegionaminic acid synthase	4155	94	84
group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014	100	87

group_7345	neuA	N-acylneuraminate cytidyltransferase	4015	N/A	N/A
ERR084011		XVI			
group_5814	pglH_1	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)-diNAcBac-PP-undecapreno	4164	100	100
group_5815	epsH	Putative glycosyltransferase EpsH	4163/4172	100	97/99
group_5816	epsJ_2	putative glycosyltransferase EpsJ	4162	100	97
epsH_1		Putative glycosyltransferase EpsH	4161	99	99
group_6953	cstII	GT-42 glycosyltransferase	4149	100	100
group_5817	legI_1	N,N'-diacetyllegionaminic acid synthase	4150	100	100
group_4682	siaA	UDP-N-acetylglucosamine 2-epimerase	4051	100	100
group_4496	neuA	N-acylneuraminate cytidyltransferase	4052	100	100
group_2414	cat	Streptogramin A acetyltransferase	4168	100	100
ERR593723		XVII			
group_5351	epsJ_3	putative glycosyltransferase EpsJ	4172	100	99
group_5816	epsJ_2	putative glycosyltransferase EpsJ	4162	100	97
epsH_1		Putative glycosyltransferase EpsH	4161	99	99
group_4171	cstII	GT-42 glycosyltransferase	4153	100	99
group_5817	legI_1	N,N'-diacetyllegionaminic acid synthase	4150	100	100
group_4682	siaA	UDP-N-acetylglucosamine 2-epimerase	4051	100	100
group_4496	neuA	N-acylneuraminate cytidyltransferase	4052	100	100
vatD		Streptogramin A acetyltransferase	4171	100	99
group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014		
legI_1		N,N'-diacetyllegionaminic acid synthase	4155		
rkpK		UDP-glucose 6-dehydrogenase	4042		
group_4438		UDP-glucose 4-epimerase	4041		
galE_2		UDP-glucose 4-epimerase	4076		
rfbD		UDP-galactopyranose mutase	4075		
neuA_2		N-acylneuraminate cytidyltransferase	4059		
group_2415	epsJ_1	putative glycosyltransferase EpsJ	4169		
SRR4242504		XVIII			
group_5351	epsJ_3	putative glycosyltransferase EpsJ	4172	100	99
group_5816	epsJ_2	putative glycosyltransferase EpsJ	4162	100	97
epsH_1		Putative glycosyltransferase EpsH	4161	99	99
group_4169	cstII	GT-42 glycosyltransferase	4158	100	99
group_5817	legI_1	N,N'-diacetyllegionaminic acid synthase	4150	100	100
group_4682	siaA	UDP-N-acetylglucosamine 2-epimerase	4051	100	100
group_4700		hypothetical protein	4179	100	99
group_4496	neuA	N-acylneuraminate cytidyltransferase	4052	100	100

vatD		Streptogramin A acetyltransferase	4171	100	99
group_2414	cat	Streptogramin A acetyltransferase	4168	100	99
group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014		
legI_1		N,N'-diacetyllegionaminic acid synthase	4155		
rkpK		UDP-glucose 6-dehydrogenase	4042		
group_4438		UDP-glucose 4-epimerase	4041		
galE_2		UDP-glucose 4-epimerase	4076		
rfbD		UDP-galactopyranose mutase	4075		
neuA_2		N-acylneuraminate cytidyltransferase	4059		
SRR4453631 XIX					
group_6437	pglH_1	GalNAC-alpha-(1->4)-GalNAC-alpha-(1->3)-diNACBac-PP-undecapreno	4180	100	85
group_5351	epsJ_3	putative glycosyltransferase EpsJ	4172	100	99
group_5816	epsJ_2	putative glycosyltransferase EpsJ	4162	100	97
epsH_1		Putative glycosyltransferase EpsH	4161	99	99
group_4171	cstII	GT-42 glycosyltransferase	4153	100	99
group_5817	legI_1	N,N'-diacetyllegionaminic acid synthase	4150	100	100
group_4682	siaA	UDP-N-acetylglucosamine 2-epimerase	4051	100	100
group_4496	neuA	N-acylneuraminate cytidyltransferase	4052	100	100
vatD		Streptogramin A acetyltransferase	4171	100	99
ERR593755 XX					
group_5351	epsJ_3	putative glycosyltransferase EpsJ	4172	100	99
group_4168	cstII	GT-42 glycosyltransferase	4158	100	99
group_5817	legI_1	N,N'-diacetyllegionaminic acid synthase	4150	100	100
group_4682	siaA	UDP-N-acetylglucosamine 2-epimerase	4051	100	100
group_4496	neuA	N-acylneuraminate cytidyltransferase	4052	100	100
vatD		Streptogramin A acetyltransferase	4171	100	99
SRR5794756 XXI					
group_10251	epsJ_1	putative glycosyltransferase EpsJ	4197	67	91
group_10252		hypothetical protein	4196	100	100
group_10253	epsJ_2	putative glycosyltransferase EpsJ	4195	100	100
group_4169	cstII	GT-42 glycosyltransferase	4158	100	99
group_5817	legI_1	N,N'-diacetyllegionaminic acid synthase	4150	100	100
group_4682	siaA	UDP-N-acetylglucosamine 2-epimerase	4051	100	100
group_4700		hypothetical protein	4179	100	99
group_4496	neuA	N-acylneuraminate cytidyltransferase	4052	100	100
vatD		Streptogramin A acetyltransferase	4171	100	99

SRR3215130		XXII				
group_5814	pglH_1	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)-diNAcBac-PP-undecapreno	4164	100	100	
group_5351	epsJ_3	putative glycosyltransferase EpsJ	4172	100	99	
group_5816	epsJ_2	putative glycosyltransferase EpsJ	4162	100	97	
epsH_1		Putative glycosyltransferase EpsH	4161	99	99	
group_4169	cstII	GT-42 glycosyltransferase	4158	100	99	
group_5817	legI_1	N,N'-diacetyllegionaminic acid synthase	4150	100	100	
group_4682	siaA	UDP-N-acetylglucosamine 2-epimerase	4051	100	100	
group_4700		hypothetical protein	4179	100	99	
group_4496	neuA	N-acylneuraminate cytidyltransferase	4052	100	100	
group_2414	cat	Streptogramin A acetyltransferase	4168	100	99	
ERR1163833, SRR2920321		XXIII				
group_5920	cstII	GT-42 glycosyltransferase	4132	98	87	
legI_1		N,N'-diacetyllegionaminic acid synthase	4155	94	84	
group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014	100	87	
group_5314	neuA	N-acylneuraminate cytidyltransferase	4131/4156	91/NA	77/NA	
group_8887	vatD	Streptogramin A acetyltransferase	4130/4173	100	85/89	

Group 4 and 5						
Gene	Non-unique Gene name	Annotation	Orthologue	coverage	%nt id to <i>C. jejuni</i>	
SRR1975100		XXIV				
group_9352	pglH_1	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)-diNAcBac-PP-undecapreno	4175	99	78	
group_5160	cstV	GT-42 glycosyltransferase	4139	N/A	N/A	
group_1515	legI_1	N,N'-diacetyllegionaminic acid synthase	4013	N/A	N/A	
group_5920	cstII	GT-42 glycosyltransferase	4132	99	88	
legI_1		N,N'-diacetyllegionaminic acid synthase	4155	100	83	
group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014	100	86	
group_5314	neuA	N-acylneuraminate cytidyltransferase	4131	91	77	
group_8887	vatD	Streptogramin A acetyltransferase	4130	100	85	

Group 5						
Gene	Non-unique Gene name	Annotation	Orthologue	coverage	%nt id to <i>C. jejuni</i>	
ERR1163655		IX				
group_5159	pglH_2	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)-diNAcBac-PP-undecapreno	4138	95	78	
group_5160	cstV	GT-42 glycosyltransferase	4139	N/A	N/A	
group_1515	legI_1	N,N'-diacetyllegionaminic acid synthase	4013	99	84	
group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014	100	87	
group_4585	neuA	N-acylneuraminate cytidyltransferase	4146	N/A	N/A	
group_8111	gspA	General stress protein A	4147	N/A	N/A	

ERR1163885, ERR1163441, ERR1163615, ERR1163639, ERR1163676, ERR1163717, ERR1163831			XXV		
group_5159	pglH_2	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)-diNAcBac-PP-undecapreno	4138	95	78
group_5160	cstV	GT-42 glycosyltransferase	4139	N/A	N/A
group_1515	legI_1	N,N'-diacetyllegionaminic acid synthase	4013	99	84
group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014	100	87
group_4585	neuA	N-acylneuraminate cytidyltransferase	4146/4148	N/A 95	N/A 78

ERR1163416, ERR1163494, ERR1163599, ERR1163863			XXVI		
group_5159	pglH_2	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)-diNAcBac-PP-undecapreno	4138	95	78
group_5160	cstV	GT-42 glycosyltransferase	4139	N/A	N/A
group_1515	legI_1	N,N'-diacetyllegionaminic acid synthase	4013	99	84
group_7292	siaA	UDP-N-acetylglucosamine 2-epimerase	4140	82	88
group_7293		hypothetical protein	4141	N/A	N/A
group_7294		hypothetical protein	4142	97	78
group_7295	neuA	N-acylneuraminate cytidyltransferase	4143	N/A	N/A

Group 6						
Gene	Non-unique Gene name	Annotation	Orthologue	coverage	%nt id to <i>C. jejuni</i>	
SRR5005386	III					
pglH_1		GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)-diNAcBac-PP-undecapreno	4002	69	73	
gspA		General stress protein A	4003	N/A	N/A	
group_5761	epsJ_1	putative glycosyltransferase EpsJ	4004	35	84	
group_2111		hypothetical protein	4211	N/A	N/A	
group_3997	cstVI	GT-42 glycosyltransferase	4127/cst-F	N/A	N/A	
ERR1163830	XXVII					
group_8851	epsJ_3	putative glycosyltransferase EpsJ	4129	66	75	
group_6191		hypothetical protein	4067	N/A	N/A	
group_4660	epsJ_2	putative glycosyltransferase EpsJ	4068	97	80	
group_8850		hypothetical protein	4069	N/A	N/A	
group_8256		hypothetical protein	4070	N/A	N/A	
group_4000	cstVI	GT-42 glycosyltransferase	4200	N/A	N/A	
ERR1163806, ERR1163854	XXVIII					
group_6190	epsJ_1	putative glycosyltransferase EpsJ	4066	91	76	
group_6191		hypothetical protein	4067	N/A	N/A	
group_4660	epsJ_2	putative glycosyltransferase EpsJ	4068	97	80	
tagE		putative poly(glycerol-phosphate) alpha-glycosyltransferase	4124	N/A	N/A	
group_8755	epsJ_3	putative glycosyltransferase EpsJ	4125	40	83	

group_2110		hypothetical protein	4126	N/A	N/A
group_3997	cstVI	GT-42 glycosyltransferase	4127	N/A	N/A
group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014		
legI_1		N,N'-diacetyllegionaminic acid synthase	4155		
group_4658	cstI	hypothetical protein	4157		

Group 7					
Gene	Non-unique Gene name	Annotation	Orthologue	coverage	%nt id to <i>C. jejuni</i>
ERR1163758	II				
group_7713	epsJ_1	putative glycosyltransferase EpsJ	4201	N/A	N/A
group_7712		hypothetical protein	4208	N/A	N/A
group_2667		hypothetical protein	4209	N/A	N/A
group_6160	cstIV	GT-42 glycosyltransferase	4210	N/A	N/A
group_7711		hypothetical protein	4202	N/A	N/A
cptA		Phosphoethanolamine transferase CptA	4203	N/A	N/A
group_7710		hypothetical protein	4204	N/A	N/A
group_4620		hypothetical protein	4205	N/A	N/A
group_6063		hypothetical protein	4206	N/A	N/A
group_3722		hypothetical protein	4207	N/A	N/A
group_4658		hypothetical protein	4157		
group_7183	neuA	N-acylneuraminate cytidyltransferase	4156		
legI_1		N,N'-diacetyllegionaminic acid synthase	4155		
group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014		
ERR1163726	XXIX				
group_5303	epsJ_1	putative glycosyltransferase EpsJ	4060	67	76
group_4659		hypothetical protein	4061	N/A	N/A
group_8389		hypothetical protein	4083	N/A	N/A
group_6184	cstIV	GT-42 glycosyltransferase	4063	N/A	N/A
eptC		Phosphoethanolamine transferase EptC	4084	N/A	N/A
group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014		
legI_1		N,N'-diacetyllegionaminic acid synthase	4155		
group_4585	neuA	N-acylneuraminate cytidyltransferase	4146		
ERR1163788, SRR1975124, ERR1163731, SRR3214492				30	
group_5303	epsJ_1	putative glycosyltransferase EpsJ	4060	67	76
group_4659		hypothetical protein	4061	N/A	N/A
group_2671		hypothetical protein	4062	N/A	N/A

group_6184	cstIV	GT-42 glycosyltransferase	4063	N/A	N/A
eptC		Phosphoethanolamine transferase EptC	4084	N/A	N/A
ERR1163754		XXXI			
group_8560	epsJ_1	putative glycosyltransferase EpsJ	4101	91	77
group_4659		hypothetical protein	4061	N/A	N/A
group_2671		hypothetical protein	4062	N/A	N/A
group_6184	cstIV	GT-42 glycosyltransferase	4063	N/A	N/A
eptC		Phosphoethanolamine transferase EptC	4084	N/A	N/A
group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014		
legI_1		N,N'-diacetyllegionaminic acid synthase	4155		
group_4658		hypothetical protein	4157		
group_5314	neuA	N-acylneuraminate cytidyltransferase	4131		
ERR1163708 and SRR3214482		XXXII			
group_5303	epsJ_1	putative glycosyltransferase EpsJ	4060	67	76
group_4659		hypothetical protein	4061	N/A	N/A
group_2671		hypothetical protein	4062	N/A	N/A
group_6184	cstIV	GT-42 glycosyltransferase	4063	N/A	N/A
pssM		Exopolysaccharide glucosyl ketal-pyruvate-transferase	4064	N/A	N/A
group_8235		hypothetical protein	4065	43	81
ERR1163818, ERR1163835		XXXIII			
group_5303	epsJ_1	putative glycosyltransferase EpsJ	4060	67	76
group_4659		hypothetical protein	4061	N/A	N/A
group_2671		hypothetical protein	4062	N/A	N/A
group_6184	cstIV	GT-42 glycosyltransferase	4063	N/A	N/A
eptC		Phosphoethanolamine transferase EptC	4084	N/A	N/A
pssM		Exopolysaccharide glucosyl ketal-pyruvate-transferase	4064	N/A	N/A
group_8235		hypothetical protein	4065	43	81
SRR1975129		XXXIV			
group_5303	epsJ_1	putative glycosyltransferase EpsJ	4060	67	76
group_4659		hypothetical protein	4061	N/A	N/A
group_2671		hypothetical protein	4062	N/A	N/A
group_6184	cstIV	GT-42 glycosyltransferase	4063	N/A	N/A
eptC		Phosphoethanolamine transferase EptC	4084	N/A	N/A
eptC_2		Phosphoethanolamine transferase EptC	4176	N/A	N/A
group_9399		hypothetical protein	4177	N/A	N/A

group_2770	siaA	UDP-N-acetylglucosamine 2-epimerase	4014		
legI_1		N,N'-diacetyllegionaminic acid synthase	4155		
ERR1163785	XXXV				
group_8590	epsJ_2	putative glycosyltransferase EpsJ	4105	66	76
group_6243		hypothetical protein	4106	N/A	N/A
group_6244		hypothetical protein	4107	N/A	N/A
group_8591	cstIV	GT-42 glycosyltransferase	4108	N/A	N/A
eptC_1		Phosphoethanolamine transferase EptC	4109	N/A	N/A
group_8593	eptC_2	Phosphoethanolamine transferase EptC	4110	N/A	N/A
group_4522	fdtB	dTDP-3-amino-3,6-dideoxy-alpha-D-galactopyranose transaminase	4018	N/A	N/A
group_8594	wfgD	UDP-Glc:alpha-D-GlcNAc-diphosphoundecaprenol beta-1,3-glucosyltr	4111	N/A	N/A
group_8595		hypothetical protein	4112	N/A	N/A
group_8596	epsE	Putative glycosyltransferase EpsE	4113	N/A	N/A
group_6229		hypothetical protein	4097	N/A	N/A
group_8529	lacA_1	Galactoside O-acetyltransferase	4098	N/A	N/A
fdtA		TDP-4-oxo-6-deoxy-alpha-D-glucose-3,4-oxoisomerase	4078	N/A	N/A
oatWY		Polysialic acid O-acetyltransferase	4077	N/A	N/A
rfbB		dTDP-glucose 4,6-dehydratase	4026	N/A	N/A
rmlA		Glucose-1-phosphate thymidyltransferase	4027	N/A	N/A

Supplementary Table 6

Genome (Acc. Number)	MLST Sequence type	ReMatCh results for GT-42
ERR221394	19	<i>cstIV</i>
ERR221300	48	<i>cstIV</i>
SRR5170912	459	<i>cstIV</i>
SRR2976823	459	<i>cstIV</i>
SRR3583303	459	<i>cstIV</i>
SRR4018132	459	<i>cstIV</i>
SRR2727977	459	<i>cstIV</i>
SRR3066579	459	<i>cstIV</i>
SRR5808785	459	<i>cstIV</i>
SRR5821279	459	<i>cstIV</i>
SRR3953110	459	<i>cstIV</i>
SRR3611765	459	<i>cstIV</i>
SRR5297422	459	<i>cstIV</i>
SRR5209908	459	<i>cstIV</i>
SRR3543495	459	<i>cstIV</i>
ERR1906140	1286	<i>cstIV</i>
ERR1906133	1286	<i>cstIV</i>
ERR1906188	1316	<i>cstIV</i>
ERR226168	2568	<i>cstIV</i>
ERR152270	2568	<i>cstIV</i>
SRR3402101	4486	<i>cstIV</i>
ERR349076	6564	<i>cstIV</i>
ERR349092	7278	<i>cstIV</i>
ERR1025561	7546	<i>cstIV</i>
SRR5230154	-	<i>cstIV</i>
SRR5297186	-	<i>cstIV</i>
SRR4175504	-	<i>cstIV</i>
SRR4022328	-	<i>cstIV</i>
SRR5222006	-	<i>cstIV</i>
ERR349072	-	<i>cstIV</i>
SRR4022329	-	<i>cstIV</i>
SRR3196134	-	<i>cstIV</i>
SRR3490037	-	<i>cstIV</i>
ERR222680	-	<i>cstIV</i>
SRR3619952	-	<i>cstIV</i>
ERR349126	-	<i>cstIV</i>
SRR3490036	-	<i>cstIV</i>
SRR3882939	-	<i>cstIV</i>
SRR5925242	-	<i>cstIV</i>
ERR355978	-	<i>cstIV</i>
ERR196590	19	<i>cstVI</i>
ERR211364	21	<i>cstVI</i>
ERR152288	45	<i>cstVI</i>
ERR196398	61	<i>cstVI</i>
ERR196702	1044	<i>cstVI</i>
ERR152190	2568	<i>cstVI</i>
ERR128771	-	<i>cstVI</i>
ERR211196	-	<i>cstVI</i>
ERR211199	-	<i>cstVI</i>
ERR193035	-	<i>cstVI</i>
ERR084058	3029	<i>cstVII</i>
ERR1815190	-	<i>cstVII</i>

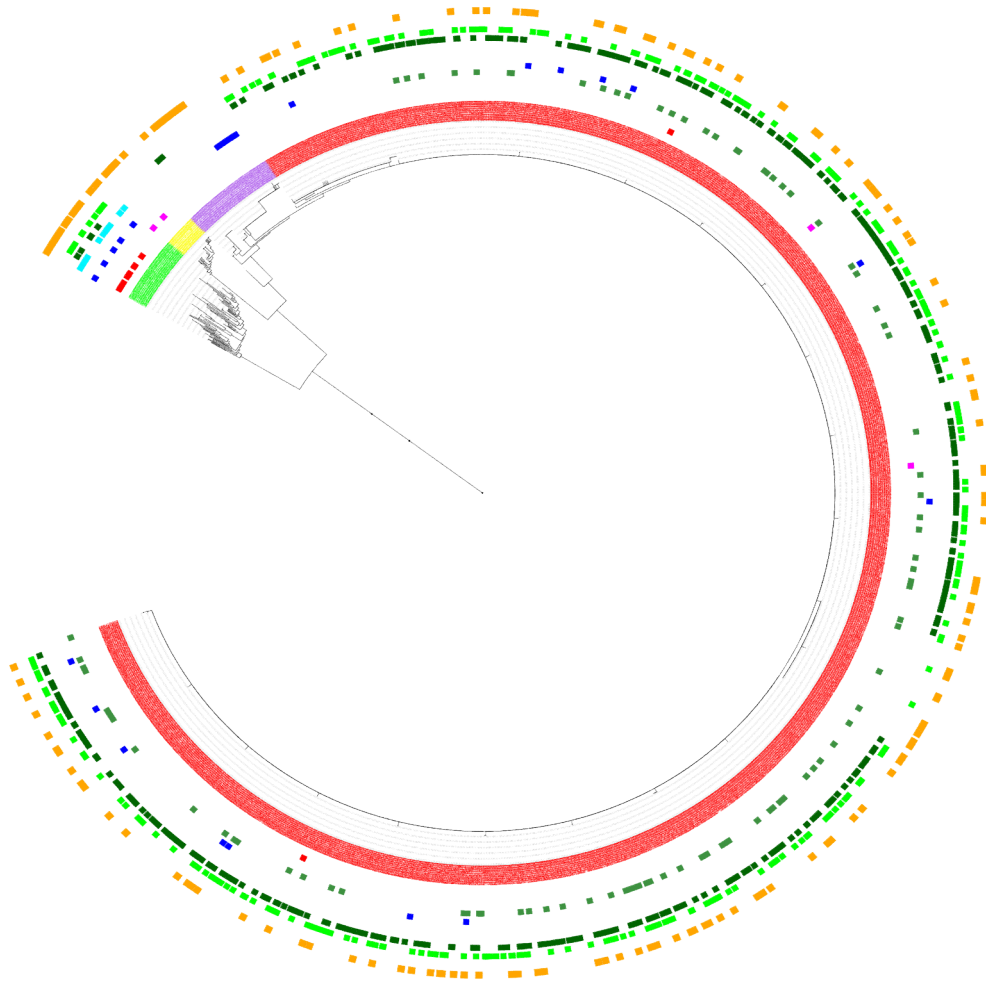


Figure S1. Evolutionary analyses of 2553 *C. coli* strains based on *atpA* conducted in MEGA7 and the evolutionary history was inferred using the Minimum Evolution method calculating the distance using Maximum Composite Likelihood. Colour of the leaf correspond of the *C. coli* clades as defined by Sheppard and colleagues¹, and inferred using hierBAPS2: Clade 1a, red; Clade 1b and 1c, violet; Clade 2, yellow; Clade 3, green. Circles from in to out, positive ReMatCh results for: GT42 Group 1 (cstI), red; GT42 Group 2 (cstIII), pink; GT42 Group 3 (cstVII), turquoise; GT42 Group 4 (cstII), blue; GT42 Group 5 (cstV), cyan; GT42 Group 6 (cstVI), green; GT42 Group 7 (cstIV), light green; neuB, orange.

REFERENCES

- 1 Sheppard, S. K. *et al.* Progressive genome-wide introgression in agricultural *Campylobacter coli*. *Mol. Ecol.* **22**, 1051–1064 (2013).
- 2 Cheng, L., Connor, T. R., Sirén, J., Aanensen, D. M. & Corander, J. Hierarchical and Spatially Explicit Clustering of DNA Sequences with BAPS Software. *Mol. Biol. Evol.* **30**, 1224–1228 (2013).