Original Article

Genetic Characteristics of Lipooligosaccharide and Capsular Polysaccharide of *Campylobacter jejuni* from Different Sources in China*

WANG Jia Qi[1,2], CHEN Xiao Li[1], ZHOU Gui Lan[1], WANG Hai Rui[1], GU Yi Xin[1], ZHANG Jian Zhong[1], SHAO Zhu Jun[1], and ZHANG Mao Jun[1,#]

1. State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China; 2. National Institute of Environmental Health, Chinese Center for Disease Control and Prevention, Beijing 100021, China

Abstract

Objective To determine the distribution of two important virulence factors [lipooligosaccharide (LOS) and capsular polysaccharide (CPS)] in *Campylobacter jejuni* (*C. jejuni*) isolated from different sources in China and to develop a rapid screening method for Guillain–Barré syndrome (GBS)-associated strains.

Methods Whole-genome sequencing was carried out for 494 *C. jejuni* strains. The OrthoMCL software was used to define the LOS/CPS gene clusters. CPS genotyping was performed with serotype-specific sequence alignment using the BLAST software. Real-time Polymerase chain reaction (PCR) was developed with the unique sequences of specific CPS types.

Results Nine novel and 29 previously confirmed LOS classes were identified. LOS classes A, B, and C were the most common (48.2%, 238/494) among the 494 strains. Twenty-six capsular types were identified in 448 strains. HS2, HS4c, HS5/31, HS19, and HS8/17 were the most frequent CPS genotypes (58.7%, 263/448). Strains of 17 CPS genotypes (strain number > 5) had one or two prevalent LOS classes ($P < 0.05$). Multiplex real-time PCR for rapid identification of HS2, HS19, and HS41 was developed and validated with strains of known serotypes.

Conclusion Our results describe the genetic characteristics of the important virulence factors in *C. jejuni* strains in China. The multiplex real-time PCR developed in this study will facilitate enhanced surveillance of GBS-associated strains in China.

Key words: *Campylobacter jejuni*; Lipooligosaccharide; Capsular polysaccharide; Multiplex real-time PCR

Biomed Environ Sci, 2022; 35(12): 1106-1114 *doi: 10.3967/bes2022.140* *ISSN: 0895-3988*

www.besjournal.com (full text) *CN: 11-2816/Q*

INTRODUCTION

Campylobacter jejuni (*C. jejuni*) is a major cause of human gastroenteritis worldwide. It commonly colonizes the intestinal tract of poultry and other livestock. The number of campylobacteriosis cases in China is estimated to be much higher than in Europe and the USA[1].

Lipooligosaccharide (LOS) is an important virulence factor that may play a role in microbial adhesion and invasion. In addition, the sialylated LOS is the determinant structure, which mimics human gangliosides. The LOS loci of classes A, B, and C are the major genetic types related to the Guillain-Barré syndrome (GBS)[2-4].

Capsular polysaccharide (CPS) is another major

*This work was supported by the National Key Research and Development Program of China [2021YFC2301000] and the Sanming Project of Medicine in Shenzhen [SZSM201803081].

#Correspondence should be addressed to ZHANG Mao Jun, PhD, E-mail: zhangmaojun@icdc.cn

Biographical note of the first author: WANG Jia Qi, male, born in 1986, PhD, majoring in prevention and control of infectious diseases.

virulence factor and is the serodeterminant of the Penner serotyping scheme for *C. jejuni*[5]. Some specific serotypes of *C. jejuni* strains, such as HS1/44, HS2, HS4c, HS10, HS19, HS23/36, and HS41[6-8], are associated with GBS. Previous studies in China suggest that HS2, HS19, and HS41 are dominant *C. jejuni* strains that trigger GBS[9-12].

In this study, the genetic characteristics of LOS and CPS of *C. jejuni* strains from different sources in China were analyzed by whole-genome sequencing (WGS), and a multiplex real-time Polymerase chain reaction (PCR) assay was developed for the rapid identification of three crucial GBS-associated CPS types in China.

MATERIALS AND METHODS

Strains and DNA Extraction

A total of 494 *C. jejuni* strains, including 260 strains from diarrhea patients, 13 strains from GBS patients, 70 strains from livestock, and 151 strains from poultry, were collected for this study. All strains were identified as *C. jejuni* using previously described methods[13]. All tested strains were grown on *Campylobacter* selective medium (Karmali; Oxoid, Basingstoke, UK) supplemented with 5% sheep blood at 37 °C under microaerobic conditions (5% O_2, 10% CO_2, and 85% N_2) for 48 h. DNA was extracted from each strain using the QIAamp DNA mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions for WGS.

Genome Sequencing and Genetic Identification for CPS/LOS

The genome of *C. jejuni* isolates was sequenced using an Illumina HiSeq 2000 sequencing platform at the Beijing Genomics Institution with a depth of 450 × coverage. The low-quality reads were removed if the quality scores of ≥ 3 consecutive bases were ≤ Q30. The reads were then assembled into contigs and scaffolds using SOAP*denovo* v2.04 (http://soap.genomics.org.cn/soapdenovo.html). PCR amplification and Sanger sequencing were used to fill the gaps in the region of the CPS/LOS sequences. The CPS/LOS locus sequences were annotated using Prokka V1.13.3[14] with default parameters and the genus-specific database from RefSeq[15].

According to a previous study[16], the integrated sequences between the conserved genes *kpsF* and *kpsC* were considered the CPS biosynthesis gene cluster, and the LOS biosynthesis gene cluster was recognized by the genes *waaC* and *waaF*[17].

OrthoMCL v2.0.9 was used to cluster all CPS/LOS genes into orthologous and paralogous groups[18] (the accession numbers of CPS/LOS reference sequences used in this study are listed in Supplementary Tables S1 and S2 available in www.besjournal.com). A database of translated coding sequences for *C. jejuni* LOS/CPS biosynthesis was assembled according to the orthologue nomenclature described in a previous study[19]. Reciprocal all-versus-all BLASTP was performed (E-value < 1×10^{-5}), and the results were processed by OrthoMCL with default parameters (percent match length ≥ 50%). A database was created containing orthologues and paralogues. Orthologues are homologous sequences derived from speciation events, and paralogues are homologous sequences derived from duplication events[20].

Serotype-specific Sequence Database for CPS Genotyping

A serotype-specific sequence database for 33 CPS types (HS1, HS2, HS4 complex, HS5, HS6/7, HS8/17, HS9, HS10, HS11, HS12, HS15, HS18, HS19, HS21, HS22, HS27, HS29, HS31, HS32, HS33, HS37, HS38, HS40, HS41, HS42, HS44, HS45, HS52, HS53, HS55, HS57, HS58, and HS60) was created. Serotype-specific genes in each serotype were identified by a shell script run on CentOS release 6.9. This shell script requires several preinstalled bioinformatics software programs, including Samtools[21], BWA[22], Wgsim, and Megahit[23]. Briefly, Wgsim was used to simulate sequencing reads from each CPS reference sequence, and the reads were then mapped to other CPS reference sequences with BWA. The unmapped reads were obtained with Samtools and subsequently assembled with Megahit. The assembled contigs were subjected to BLAST in the NCBI database to eliminate the sequences mapped onto the CPS loci of other known serotypes. The serotype-specific sequence database was established through the above process. The CPS genotype for each strain in this study was determined based on searching and BLASTing with serotype-specific sequences in the database.

Multiplex Real-time PCR for Rapid Identification of HS2, HS19, and HS41

The optimized primers and probes for this study were designed using Primer-BLAST (Table 1). The reaction volume was 25 μL in total, containing 12.5 μL 2 × qPCR SuperMix, 2 μmol/L HS2-F and HS2-R, 3 μmol/L HS19-F and HS19-R, 3 μmol/L HS41-F and HS41-R, 1 μmol/L probe HS2, 1.5 μmol/L probe HS19, 1.5 μmol/L probe HS41 and 1 μL of DNA

template. The thermocycling conditions were optimized as follows: initial denaturation at 94 °C for 30 s, followed by 40 cycles of 94 °C for 10 s and 60 °C for 30 s. This assay was further validated with 109 strains of known serotype strains. The DNA template of each strain was prepared after lysing whole bacterial cells by *boiling*. The *Ct* value of ≤ 35 was considered positive, and the result was considered negative when the *Ct* value was more than 40 or no amplification curve was obtained. Sensitivity and specificity were calculated for this multiplex real-time PCR assay. Sensitivity was calculated as the number of true positives divided by the sum of true positives and false negatives (true positive was defined as the Penner phenotypic serotyping results). Specificity was calculated as the number of true negatives divided by the number of true negatives and false positives.

Genetic Variation in the Same CPS Types

Among the HS2, HS19, and HS41 strains identified by multiplex real-time PCR, the integrated CPS loci were selected for further analysis of genetic variation. CPS loci from six HS19 strains and three HS41 strains were selected. The phenotypic serotype was analyzed previously. Eleven CPS loci from HS2 strains were also selected for genetic variation analysis. The sequences of each CPS locus from the strains with the same CPS type were aligned with mafft v7.471. Variant calling was performed using the Find Variations/SNPs program on Geneious v2021.03 with the default settings.

Statistical Analyses

The Chi-square test was used for all of the statistical analyses throughout the whole study. A *P*-value < 0.05 was considered to be statistically significant. All analyses were conducted with a statistical software package (IBM SPSS Statistics 19.0, Chicago, IL, USA).

RESULTS

CPS/LOS Types of C. jejuni Strains from Different Sources

A panel of 26 different capsular types was identified in 448 *C. jejuni* strains. The CPS genotypes of 46 strains (9.31%) could not be determined using serotype-specific sequences. The number of different CPS genotypes and the distribution of different CPS genotypes in strains from different sources are presented in Figure 1. The five most common CPS types, which accounted for 58.71% (263/448) of all the tested strains, were HS2 (90 strains, 20.08%), HS4c (70 strains, 15.63%), HS5/31 (40 strains, 8.93%), HS19 (35 strains, 7.81%), and HS8/17 (28 strains, 6.25%).

The CPS genotypes of 12 strains (4.62%) from diarrhea patients could not be determined. The remaining 248 strains were classified into 22 CPS genotypes. The five most common CPS genotypes of the strains from diarrhea patients were HS2 (53 strains, 21.4%), HS4c (45 strains, 18.1%), HS5/31 (24 strains, 9.7%), HS19 (16 strains, 6.5%), and HS8/17 (16 strains, 6.5%). The CPS genotype of only one of the strains (7.7%) from GBS patients could not be determined. The remaining 12 GBS-associated strains consisted of five CPS types (HS19, HS37, HS2, HS53, and HS41). The CPS genotypes could not be determined for 11.43% (8/70) and 16.56% (25/151) of the strains from livestock and poultry,

Table 1. Primers and probes for real-time PCR detection of HS2, HS19, and HS41

Capsular type	Primers and probes	Nucleotide sequence (5' to 3')
HS2	HS2-F	AACCAACCATCCACTATTTTCATCT
	HS2-R	ACTCTTTGATTTTCCAATGCAATGTT
	HS2-P	FAM-CGCACCTTCCAATGCAACCAAAGAGC-BHQ1
HS19	HS19-F	CTGAAATATCAACACAGGGAAAATG
	HS19-R	TGTTGGAGGTAAAAGACATGGTG
	HS19-P	HEX-TCCAATCCTGTTGCTCATAACGCATTCA-BHQ1
HS41	HS41-F	CACCCTGATTTCTATTAACTCTACCAC
	HS41-R	GAGTTATTCTTGATGCTGAAAATGGA
	HS41-P	CY5-AAACGGGATAAGATCGCTTGGGGGAATA-BHQ2

Note. PCR, Polymerase chain reaction.

respectively. The five most common CPS genotypes for the strains from livestock were HS2 (12 strains, 17.14%), HS19 (10 strains, 14.29%), HS4c (7 strains, 10.0%), HS10 (7 strains, 10.0%), and HS11 (6 strains, 8.57%), and the five most common CPS types for the strains from poultry were HS2 (23 strains, 15.23%), HS4c (18 strains, 11.92%), HS5/31 (13 strains, 8.61%), HS8/17 (10 strains, 6.62%), and HS1 (9 strains, 5.96%). Except for the strains from GBS patients, HS2 and HS4c were the most frequent CPS types among the tested strains from different sources.

In total, 38 LOS classes were identified in 494 *C. jejuni* strains. Twenty-nine were previously confirmed LOS[3,24-26]. Nine novel classes were identified in 11 strains (9 from poultry and 2 from diarrhea patients) and were named CDC11–CDC19 (Figure 2). Four poultry strains (classified as CDC11, CDC15, CDC16, and CDC19) possessed the sialic acid synthesis genes *ORF1* (*cgtA/NeuA*), *ORF37* (*cstII*), *ORF41* (*NeuB*), and *ORF38* (*NeuC*) in the LOS locus. The distribution of the LOS classes in strains from different sources is presented in Table 2. LOS classes A, B, and C accounted for 12.35% (61/494), 28.54% (141/494), and 7.29% (36/494) of the tested strains, respectively. These three classes accounted for the largest proportion (48.18%, 238/494). Classes P, H, CDC5, F, CDC6, R, CDC2, W, CDC1, and G accounted for 38.46% (190/494) of the tested strains.

Twenty-five LOS classes were identified in 260 strains from diarrhea patients. The five most common LOS classes were class B (28.08%, 73/260), A (12.69%, 33/260), H (9.23%, 24/260), P (8.08%, 21/260), and C (7.31%, 19/260). A total of 9 of 13 (69.23%) GBS-associated strains had class A LOS, and another 4 strains had LOS class P (2 strains), F (1 strain), and H (1 strain). Twelve LOS classes were identified in 70 livestock strains, and the five most abundant classes were class B (38.57%, 27/70), A (21.43%, 15/70), C (10%, 7/70), F (8.57%, 6/70), and

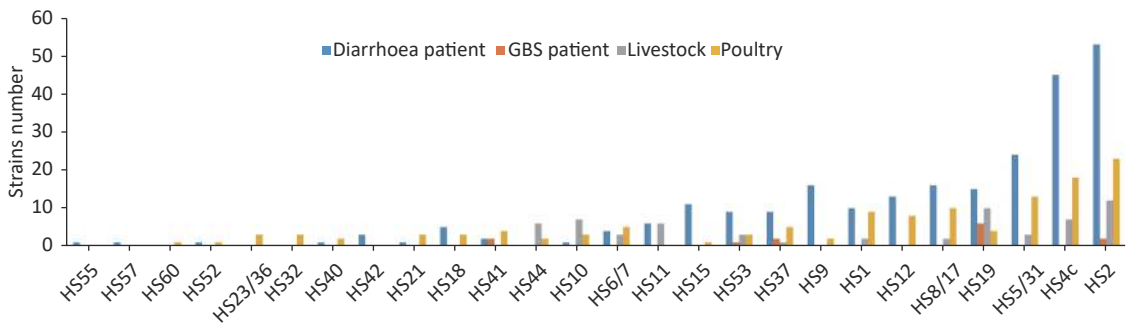

Figure 1. Distribution of different CPS types in strains from different sources: The CPS genotype for each strain was determined based on the different serospecific sequences. The top five CPS types were HS2 (90 strains, 20.1%), HS4c (70 strains, 15.6%), HS5/31 (40 strains, 8.9%), HS19 (35 strains, 7.8%), and HS8/17 (28 strains, 6.3%), which accounted for 58.7% (263/448) of the tested strains. CPS, capsular polysaccharide.

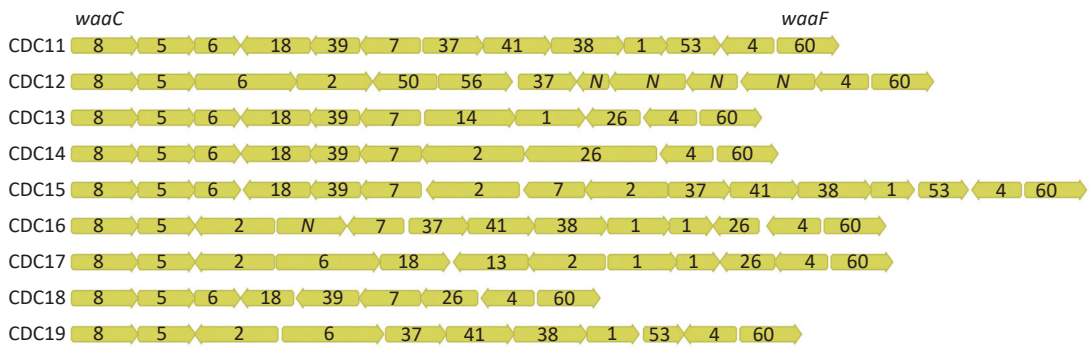

Figure 2. Newly confirmed LOS classes (CDC11–CDC19) OrthoMCL was used to cluster all LOS genes into orthologous and paralogous groups and visualized in Geneious software. Arrows represent orthologues and paralogues. Numbers are orthologue ID numbers, and "N" are paralogues. (Supplementary Table S3 available in www.besjournal.com). LOS, lipooligosaccharide.

Table 2. The distribution of the LOS classes in strains from different sources

LOS class	Diarrhoea patients	GBS patients	Livestock	Poultry	In total (%)
A	33	9	15	4	61 (12.35)
B	73	0	27	41	141 (28.54)
C	19	0	7	10	36 (7.29)
D	0	0	0	1	1 (0.20)
E	3	0	3	1	7 (1.42)
F	9	1	6	2	18 (3.64)
G	10	0	0	2	12 (2.43)
H	24	1	2	4	31 (6.28)
I	0	0	2	0	2 (0.40)
J	2	0	0	0	2 (0.40)
K	5	0	0	3	8 (1.62)
M	0	0	0	1	1 (0.20)
O	2	0	0	0	2 (0.40)
P	21	2	2	8	33 (6.68)
R	1	0	0	14	15 (3.04)
S	5	0	3	1	9 (1.82)
T	1	0	0	1	2 (0.40)
V	1	0	1	2	4 (0.81)
W	4	0	0	10	14 (2.83)
CDC1	1	0	1	11	13 (2.63)
CDC2	12	0	0	2	14 (2.83)
CDC3	3	0	0	1	4 (0.81)
CDC4	5	0	0	0	5 (1.01)
CDC5	18	0	0	5	23 (4.66)
CDC6	2	0	1	14	17 (3.44)
CDC7	0	0	0	1	1 (0.20)
CDC8	3	0	0	0	3 (0.61)
CDC9	0	0	0	1	1 (0.20)
CDC10	1	0	0	2	3 (0.61)
CDC11	0	0	0	1	1 (0.20)
CDC12	2	0	0	1	3 (0.61)
CDC13	0	0	0	1	1 (0.20)
CDC14	0	0	0	1	1 (0.20)
CDC15	0	0	0	1	1 (0.20)
CDC16	0	0	0	1	1 (0.20)
CDC17	0	0	0	1	1 (0.02)
CDC18	0	0	0	1	1 (0.02)
CDC19	0	0	0	1	1 (0.02)

Note. LOS, lipooligosaccharide, GBS, Guillain-Barré syndrome.

S (4.29%, 3/70). Thirty-three LOS classes were identified in 151 strains from poultry. The five most common classes were class B (27.15%, 41/151), CDC6 (9.27%, 14/151), R (9.27%, 14/151), CDC1 (7.28%, 11/151), and W (6.62%, 10/151).

Combination Distribution of CPS/LOS Types

In Table 3, 17 CPS types of strains (strain number > 5) were analyzed. Strains of HS9, HS10, HS11, and HS37 were classified into LOS class CDC5, B, F, and P, respectively. And strains of the other 13 CPS types also had one or two prevalent LOS classes (bold in Table 3, $P < 0.05$). Four GBS-related genes (*NeuA*, *NeuB*, *NseuC,* and *cst*) involved in the sialic acid biosynthesis pathway were detected in 279 strains (56.48%, 279/494). In these strains, 14 CPS genotypes (HS1, HS2, HS4c, HS5/31, HS10, HS12, HS8/17, HS19, HS40, HS41, HS42, HS44, HS52, and HS60) were detected. The prevalence of GBS-related genes was higher (90%–100%) in strains HS1, HS2, HS10, HS19, and HS41 than the average (56.48%, $P < 0.01$). All strains of HS19 and HS41 contained GBS-related genes. Among HS19 strains, LOS class A was dominant (33/35), and the remaining 2 strains were classified into class R (1/35) and CDC11 (1/35). Strains of HS41 were classified into LOS class A (3/8), R (4/8), and CDC2 (1/8). LOS class A, R, CDC2, and CDC11 all possess GBS-related genes.

Assessment for Multiplex Real-time PCR

Among the tested 109 serotype known *C. jejuni* strains, all of the 11 HS2 strains, 9 HS19 strains and 8 HS41 strains were detected as positive, and 81 other serotype strains were all negative. The results were consistent with previously identified serotype results. The sensitivity and specificity of this multiplex real-time PCR assay were 100%, respectively.

Polymorphism Features of the Integrated CPS Loci of H2, HS19, and HS41

Among the six HS19 strains, all had identical phenotypic serotypes. From the alignment, each of the CPS loci in the HS19 strain was 19 kb in length and contained 15 ORFs (Figure 3). The sequence identity of these six CPS loci was 98.2%. Taking the CPS locus sequences from strain HB_CJGB_ZHX (phenotypic serotype HS19) as the reference, the sequences of the CPS locus of strains HB_CJGB_LXC and HB-CJGB-ZB were 100% identical. However, 185 SNPs and one indel (insertion/deletion) were found in the CPS loci in strain NH_A12 (Figure 3 and Table 4), which had the same phenotypic serotype as

strain HB_CJGB_ZHX. One SNP was found at the same location in *ORF30* in each of the CPS loci in strains HBJ_CJGB_96G25 and BJ-CJGB96114. The CPS locus sequences of these two strains were 99.99% identical to the reference.

Two of the three HS41 strains had the same phenotypic serotypes (ICDCCJ07001 and NH_F47). The phenotypic serotypes of strain ICDCCJ07004 could not be determined using the same commercial phenotypic serotyping kit. The CPS locus of HS41 was 36 kb in length and contained 31 ORFs. Taking the CPS locus from ICDCCJ07001 as the reference, the sequence identity of these three CPS loci was 98.9% with the reference. In total, 208 SNPs and 6 indels (insertions/deletions) were found in NH_F47, and 11 SNPs and 8 indels were identified in ICDCCJ07004. Among the 11 HS2 strains, CPS loci were 37 kb in length and contained 30 ORFs. Taking the CPS locus from NCTC11168 as the reference, the sequence identity of 11 CPS gene clusters was 98.5%, and the polymorphisms among these strains included 212 SNPs and 18 indels. The genetic variations and schematics of the ORFs for each CPS type are presented in Figure 3

Table 3. Distribution characteristics of CPS/LOS types for 494 isolates of *C. jejuni* from China

CPS types	Num	LOS classes (Num)	Proportion of strains possessed sialic acid synthesis genes in LOS (%)[a]
HS1	21	**B (8)**, **C (10)**, F (1), V (1), CDC15 (1)	90.5
HS2	90	**A (21)**, **B (54)**, H (1), P (6), R (4), CDC6 (2), CDC8 (2)	90.0
HS4c	70	A (2), **B (35)**, C (2), G (2), **H (20)**, O (2), P (1), R (5), V (1)	64.3
HS5/31	40	A (2), **B (24)**, D (1), F (3), K (6), V (1), CDC10 (3)	67.5
HS6/7	12	**E (6)**, T (1), CDC1 (4), CDC5 (1)	0
HS8/17	28	B (1), **C (17)**, G (7), V (1), CDC13 (1), CDC18 (1)	67.9
HS9	18	**CDC5 (18)**	0
HS10	11	**B (11)**	100
HS11	12	**F (12)**	0
HS12	21	**W (8)**, **CDC2 (13)**	61.9
HS15	12	G (1), H (2), **P (5)**, CDC3 (4)	0
HS18	8	T (1), **CDC4 (5)**, CDC7 (1), CDC17 (1)	0
HS19	35	**A (33)**, R (1), CDC11 (1)	100
HS21	4	W (4)	0
HS23/36	3	P (2), W (1)	0
HS32	3	CDC6 (3)	0
HS37	17	**P (17)**	0
HS40	3	B (1), H (1), M (1)	66.7
HS41	8	A (3), **R (4)**, CDC2 (1)	100
HS42	3	C (1), K (1), W (1)	33.3
HS44	8	**C (4),** F (1), I (2), K (1)	50.0
HS52	2	B (1), CDC9 (1)	/
HS53	16	F (1), J (2), **S (9)**, CDC5 (3), CDC6 (1)	0
HS55	1	E (1)	/
HS57	1	H (1)	/
HS60	1	CDC19 (1)	/
ND	46	B (6), C (2), G (2), H (6), P (2), R (1), CDC1 (9), CDC5 (1), CDC6 (11), CDC8 (1), CDC12 (3), CDC14 (1), CDC16 (1)	21.7
In total	494	/	56.5

Note. [a]LOS classes A, B, C, M, R, V, CDC2, CDC8, CDC11, CDC15, CDC16, or CDC19. LOS: lipooligosaccharide; CPS: capsular polysaccharide.

and Table 4.

DISCUSSION

In this study, we identified five major capsule types (> 5%): HS2, HS4c, HS5/31, HS19, and HS8/17 in 448 *C. jejuni* strains isolated from different sources in China. Forty-six strains could not be genotyped, possibly due to incomplete CPS loci or because their CPS types were not included in the database. Five capsule types (HS2, HS19, HS37, HS41, and HS53) were identified in 13 GBS-associated strains. Capsular-type HS19 was detected in 5.8% (35/259) of the enteritis strains but accounted for 42.9% (6/14) of the GBS-associated strains. Capsular-type HS41 was detected in 0.8% (2/259) of the enteritis strains but accounted for 14.3% (2/14) of the GBS-associated strains. We suggest that capsular types HS19 and HS41 might be associated with GBS.

In this study, we found that certain CPS types and LOS locus classes tended to appear in combination in

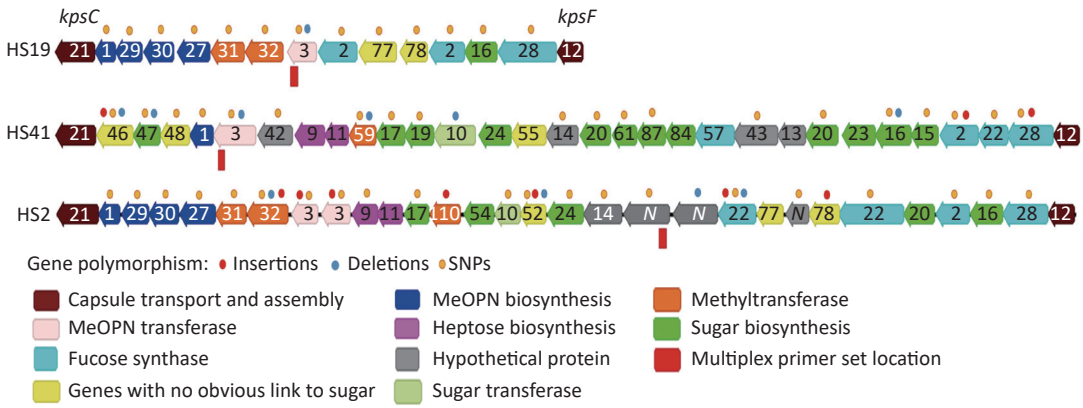

Figure 3. Schematic of CPS loci from HS2/HS19/HS41 and polymorphism distribution: OrthoMCL was used to cluster all CPS genes into orthologous and paralogous groups and visualized in Geneious software. Arrows represent orthologues and paralogues. Numbers are orthologue IDs, and "N" are paralogues. (Supplementary Table S4 available in www.besjournal.com) The different coloured dots represent gene polymorphism types. CPS: capsular polysaccharide.

Table 4. Genetic variation in the same CPS types

Strains	CPS type	Phenotypic serotype	Sequence identity to reference (%)	Polymorphisms in ORFs	Sense mutations in ORFs
HB_CJGB_LXC	HS19	HS19	100	/	/
HB-CJGB-ZB	HS19	HS19	100	/	/
HBJ_CJGB_96G25	HS19	HS19	99.99	1 SNP (30)	/
BJ-CJGB96114	HS19	HS19	99.99	1 SNP (30)	/
NH_A12	HS19	HS19	98.0	185 SNPs (1, 2, 3, 16, 27, 28, 29, 30, 31, 32, 77, 78 and a noncoding region); 1 deletion (3); 1 insertion (in a noncoding region)	56 substitutions (1, 2, 3, 16, 27, 28, 29, 30, 31, 32, 78); 1 deletion (3)
NH_F47	HS41	HS41	98.9	208 SNPs (1, 2, 3, 14, 15, 16, 17, 20, 22, 28, 42, 43, 46, 47, 48, 59, 61, 87 and three noncoding regions); 6 deletions (3, 10, 16, 46, 47, 59); 4 insertions (2, 28, 46, and one noncoding region)	99 substitutions (1, 2, 3, 14, 15, 16, 20, 22, 28, 42, 43, 46, 47, 48, 61, 87); 2 extensions (3, 46, 47); 7 frame shifts (3, 10, 16, 42, 46, 47, 59); 2 insertions (2, 28)
ICDCCJ07004	HS41	NT[a]	99.9	11 SNPs (3, 42, 46, 61, 87 and a noncoding region); 1 insertion (46); 7 deletions (3, 10, 16, 46, 47, 59, 87)	8 substitutions (3, 42, 46, 61, 87); 1 truncation (46); 2 extensions (3, 46); 7 frame shifts (3, 10, 16, 46, 47, 59, 87)

Note. [a]The phenotypic serotype could not be determined using a commercial kit. CPS: capsular polysaccharide.

C. jejuni strains. As described in Table 3, strains of most CPS types involved in this study had one or two prevalent LOS classes (Bold in Table 3); however, some CPS types could not be analyzed because the number of strains was too small (< 5). We speculate that certain combinations of CPS/LOS types might be related to the pathogenicity and environmental adaptation of *C. jejuni*.

The LOS classes A, B, C, M, R, V, CDC2, CDC8, CDC11, CDC15, CDC16, and CDC19 possess GBS-related genes (*NeuA*, *NeuB*, *NeuC,* and *cst*) involved in the sialic acid biosynthesis pathway. Approximately 90%–100% of strains of GBS-associated CPS types in this study (HS1, HS2, HS10, HS19, HS41) have LOS loci containing GBS-related genes, a prevalence that is much higher than the average (56.48%) (Table 3). All strains with capsular types HS19 and HS41 were classified into LOS classes A, R, CDC2, and CDC11. Interestingly, LOS classes A, R, CDC2, and CDC11 possess the sialylation genes necessary to produce ganglioside mimics. This ability may partly explain why HS19 and HS41 are the most prevalent serotypes among GBS-associated strains in China. All strains of capsular-type HS10 were identified as LOS class B, and HS10 was also reported as a Miller-Fisher syndrome-associated CPS type[7]; Miller-Fisher syndrome is a clinical variant of GBS.

A major weakness of this study is that our findings were not comprehensive. HS23/36 and HS37 were reported as GBS-associated serotypes, but strains of these capsular types do not possess sialylation genes. In this study, most strains of HS23/36 and HS37 were identified as LOS class P, which is a nonsialylated LOS with N-acetyl quinovosamine[27]. Further research is required to obtain evidence for the mechanism underlying the specific combination of capsular types and LOS classes resulting in synergistic effects on GBS risk.

The integrated sequence alignment revealed that the strains had identical CPS locus sequences and were classified into the same serotype (strains HB_CJGB_ZHX, HB_CJGB_LXC, HB-CJGB-ZB). Some strains had only a few SNPs and almost 99.99% similarity with the reference (strains HBJ_CJGB_96G25 and BJ-CJGB96114) and had the same phenotypic serotypes as the reference. Some other ORFs in the CPS locus had sense mutations, and the phenotypic serotype remained the same as the reference in strain NH_A12. We speculated that these mutations might not have affected the antigen structure of the CPS for the specific serotyping antibodies. For the same reason, the CPS locus in strain NH_F47 also had many SNPs, indels, extensions and frameshifts in the ORF regions, and none of these variations affected the phenotypic serotype. Eight nonsynonymous SNPs in *ORFs* 3/42/46/61/87, one truncation in *ORF46*, 7 frameshifts in *ORFs* 3/10/16/46/47/59/87, and two extensions in *ORFs* 3/46 were identified in the CPS locus from strain ICDCCJ07004. These mutations in the CPS locus caused a phenotypic variation in the serotype. Further investigation is needed to determine which mutation is crucial or causes the phase change in different serotypes.

The phenotypic serotyping method is time-consuming and labor-intensive. In this study, multiplex real-time PCR for the CPS types HS19, HS41, and HS2 was developed based on serotype-specific sequences. The multiplex PCR results were consistent with the phenotypic serotype results, which indicated the DNA variations in the CPS types HS19, HS41, and HS2 were not in the locations of the primers and probes. This approach offered a rapid screening strategy, which could not be affected by the expression of the CPS and focused on the rapid identification of the highly GBS-associated strains. However, the specific phenotypic serotypes of these strains need to be confirmed with specific serotypic antibodies.

CONCLUSION

The analysis of LOS and CPS typing characteristics contributes to recognizing the prevalence of virulence gene clusters of *C. jejuni* in China. The LOS class CDC11–CDC19 was first identified in this study. CDC11, CDC15, CDC16, and CDC19 possess genes for sialic acid synthesis and translocation, which are worthy of attention in strain monitoring. Strains of most capsular types had one or two prevalent LOS classes, supporting the hypothesis that capsular types and LOS locus classes tend to appear as a combination. The incidence of GBS is rare, and it is difficult to isolate *Campylobacter* strains once the neurological sign occurs; more GBS-related strains are needed for further confirmation study in the future. The multiplex real-time PCR developed in this study will be helpful for the identification of GBS-associated CPS types in China.

AUTHORS' CONTRIBUTIONS

All authors contributed substantially to the study. ZHANG Mao Jun, ZHANG Jian Zhong, and SHAO Zhu Jun designed, supervised and oversaw the study; WANG Jia Qi, CHEN Xiao Li, ZHOU Gui Lan, and

WANG Hai Rui analyzed the data and wrote the manuscript; GU Yi Xin isolated and identified the strains; ZHANG Mao Jun revised the manuscript and supported this study. All authors read and approved the final manuscript.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

Availability: Shell scripts for identifying specific serotype sequences are available from the authors upon request.

Received: May 5, 2022;
Accepted: September 20, 2022

REFERENCES

1. Kaakoush NO, Castaño-Rodríguez N, Mitchell HM, et al. Global epidemiology of *Campylobacter* infection. Clin Microbiol Rev, 2015; 28, 687−720.
2. Koga M, Gilbert M, Takahashi M, et al. Comprehensive analysis of bacterial risk factors for the development of Guillain-Barre syndrome after *Campylobacter jejuni* enteritis. J Infect Dis, 2006; 193, 547−55.
3. Godschalk PCR, Heikema AP, Gilbert M, et al. The crucial role of *Campylobacter jejuni* genes in anti-ganglioside antibody induction in Guillain-Barré syndrome. J Clin Invest, 2004; 114, 1659−65.
4. Yuki N. Campylobacter sialyltransferase gene polymorphism directs clinical features of Guillain-Barré syndrome. J Neurochem, 2007; 103 Suppl 1, 150−8.
5. Karlyshev AV, Linton D, Gregson NA, et al. Genetic and biochemical evidence of a *Campylobacter jejuni* capsular polysaccharide that accounts for Penner serotype specificity. Mol Microbiol, 2000; 35, 529−41.
6. Heikema AP, Islam Z, Horst-Kreft D, et al. *Campylobacter jejuni* capsular genotypes are related to Guillain-Barré syndrome. Clin Microbiol Infect, 2015; 21, 852. e1−9.
7. Salloway S, Mermel LA, Seamans M, et al. Miller-Fisher syndrome associated with *Campylobacter jejuni* bearing lipopolysaccharide molecules that mimic human ganglioside GD3. Infect Immun, 1996; 64, 2945−9.
8. Takahashi M, Koga M, Yokoyama K, et al. Epidemiology of *Campylobacter jejuni* isolated from patients with Guillain-Barre and Fisher syndromes in Japan. J Clin Microbiol, 2005; 43, 335−9.
9. Zhang MJ, Li Q, He LH, et al. Association study between an outbreak of Guillain-Barre syndrome in Jilin, China, and preceding *Campylobacter jejuni* infection. Foodborne Pathog Dis, 2010; 7, 913−9.
10. Zhang MJ, He LH, Li Q, et al. Genomic characterization of the Guillain-Barre syndrome-associated *Campylobacter jejuni* ICDCCJ07001 Isolate. PLoS One, 2010; 5, e15060.
11. Yuki N. Molecular mimicry between gangliosides and lipopolysaccharides of *Campylobacter jejuni* isolated from patients with Guillain-Barré syndrome and Miller Fisher syndrome. J Infect Dis, 1997; 176 Suppl 2, S150−3.
12. Yuki N, Takahashi M, Tagawa Y, et al. Association of Campylobacter jejuni serotype with antiganglioside antibody in Guillain-Barré syndrome and Fisher's syndrome. Ann Neurol, 1997; 42, 28−33.
13. Liang H, Wen ZY, Li Y, et al. Comparison of the filtration culture and multiple real-time PCR examination for *Campylobacter* spp. from stool specimens in diarrheal patients. Front Microbiol, 2018; 9, 2995.
14. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics, 2014; 30, 2068−9.
15. Pruitt KD, Tatusova T, Brown GR, et al. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res, 2012; 40, D130−5.
16. Karlyshev AV, Champion OL, Churcher C, et al. Analysis of *Campylobacter jejuni* capsular loci reveals multiple mechanisms for the generation of structural diversity and the ability to form complex heptoses. Mol Microbiol, 2005; 55, 90−103.
17. Klena JD, Gray SA, Konkel ME. Cloning, sequencing, and characterization of the lipopolysaccharide biosynthetic enzyme heptosyltransferase I gene (*waaC*) from *Campylobacter jejuni* and *Campylobacter coli*. Gene, 1998; 222, 177−85.
18. Li L, Stoeckert CJ Jr, Roos DS. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res, 2003; 13, 2178−89.
19. Richards VP, Lefébure T, Pavinski Bitar PD, et al. Comparative characterization of the virulence gene clusters (lipooligosaccharide and capsular polysaccharide) for *Campylobacter coli, Campylobacter jejuni* subsp. *jejuni* and related *Campylobacter* species. Infect Genet Evol, 2013; 14, 200−13.
20. Kuzniar A, van Ham RCHJ, Pongor S, et al. The quest for orthologs: finding the corresponding gene across genomes. Trends Genet, 2008; 24, 539−51.
21. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. Bioinformatics, 2009; 25, 2078−9.
22. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv https://arxiv.org/abs/1303.3997, 2013. [2022-3-12].
23. Li DH, Liu CM, Luo RB, et al. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de *Bruijn graph*. Bioinformatics, 2015; 31, 1674−6.
24. Wang JQ, Gu YX, Zhou GL, et al. Genetic diversity of lipooligosaccharide core biosynthesis gene clusters in *Campylobacter jejuni*. Dis Surveill, 2020; 35, 11−5. (In Chinese)
25. Gilbert M, Karwaski MF, Bernatchez S, et al. The genetic bases for the variation in the lipo-oligosaccharide of the mucosal pathogen, *Campylobacter jejuni*: biosynthesis of sialylated ganglioside mimics in the core oligosaccharide. J Biol Chem, 2002; 277, 327−37.
26. Parker CT, Gilbert M, Yuki N, et al. Characterization of lipooligosaccharide-biosynthetic loci of *Campylobacter jejuni* reveals new lipooligosaccharide classes: evidence of mosaic organizations. J Bacteriol, 2008; 190, 5681−9.
27. Poly F, Read TD, Chen YH, et al. Characterization of two *Campylobacter jejuni* strains for use in volunteer experimental-infection studies. Infect Immun, 2008; 76, 5655−67.

Supplementary Table S1. The accession numbers for CPS gene clusters of *C. jejuni*

CPS types	Accession number
HS1	BX545859
HS2	AL111168.1
HS3	HQ343268
HS4c	HQ343269
HS5/31	KT868847
HS6/S7	NC_009839
HS8/17	HQ343270
HS9	KT868844
HS10	HQ343271
HS11	KT868845
HS12	KT868848
HS15/31/58	HQ343272
HS18	KT932997
HS19	BX545860
HS21	KT868849
HS22	KT893439
HS23/36	BX545858
HS27	KT893437
HS29	KT868846
HS32	KT893435
HS33/35	KT893436
HS37	KT893431
HS38	KT893430
HS40	KT893434
HS41	BX545857
HS42	HQ343274
HS44	JF496678
HS45/5/32/60	KT893432
HS52	KT893429
HS53	CP000025.1
HS55	KT893433
HS57	KT893428
HS58/32	KT893427
HS60	KT893426
HS63	KT893438

Supplementary Table S2. The accession numbers for LOS gene clusters of *C. jejuni*

LOS class	Accession number
A	AF215659
B	AF401528
C	AF400047
D	AF400669
E	AJ131360
F	AY423554
G	AY436358
H	AY800272
I	EU404107
J	EU404104
K	AY573819
L	EU404111
M	EF140720
N	AY816330
O	EF143352
P	AY943308
Q	EU404112
R	AY962325
S	EU404110
T	AIOC00000000
U	AIOU00000000
V	AIOP00000000
W	CP001900
CDC1-10	reference[24]
CDC11-19	In this study

Supplementary Table S3. The annotations of ORFs in novel LOS classes of *C. jejuni*

LOS class	ORF	Products (annotated by prokka)	Length (bp)
CDC11	8	Lipopolysaccharide heptosyltransferase 1	1,029
	5	Lipid A biosynthesis lauroyltransferase	879
	6	hypothetical protein	702
	18	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)- diNAcBac-PP-undecaprenol alpha-1,4-N-acetyl-D-galactosaminyltransferase	1,071
	39	hypothetical protein	780
	7	hypothetical protein	933
	37	hypothetical protein	929
	41	N,N'-diacetyllegionaminic acid synthase	1,041
	38	UDP-N-acetylglucosamine 2-epimerase	1,119
	1	N-acylneuraminate cytidylyltransferase	666
	53	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase	831
	4	putative glycosyltransferase EpsJ	813
	60	ADP-heptose--LPS heptosyltransferase 2	944
CDC12	8	Lipopolysaccharide heptosyltransferase 1	1,029
	5	Lipid A biosynthesis lauroyltransferase	888
	6	hypothetical protein	1,551
	2	hypothetical protein	1,158
	50	hypothetical protein	981
	56	hypothetical protein	1,121
	37	hypothetical protein	891
	N	hypothetical protein	489
	N	hypothetical protein	1,168
	N	hypothetical protein	782
	N	hypothetical protein	1,128
	4	Putative glycosyltransferase EpsH	819
	60	ADP-heptose--LPS heptosyltransferase 2	944
CDC13	8	Lipopolysaccharide heptosyltransferase 1	1,029
	5	Lipid A biosynthesis lauroyltransferase	879
	6	hypothetical protein	702
	18	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)- diNAcBac-PP-undecaprenol alpha-1,4-N-acetyl-D-galactosaminyltransferase	1,071
	39	hypothetical protein	780
	7	hypothetical protein	933
	14	hypothetical protein	1,386
	1	hypothetical protein	1,074
	26	hypothetical protein	822
	4	putative glycosyltransferase EpsJ	810
	60	ADP-heptose--LPS heptosyltransferase 2	939

Continued

LOS class	ORF	Products (annotated by prokka)	Length (bp)
CDC14	8	Lipopolysaccharide heptosyltransferase 1	1,029
	5	Lipid A biosynthesis lauroyltransferase	879
	6	hypothetical protein	702
	18	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)- diNAcBac-PP-undecaprenol alpha-1,4-N-acetyl-D-galactosaminyltransferase	1,071
	39	hypothetical protein	780
	7	hypothetical protein	936
	2	Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase	1,584
	26	hypothetical protein	2,022
	4	putative glycosyltransferase EpsJ	810
	60	ADP-heptose--LPS heptosyltransferase 2	939
CDC15	8	Lipopolysaccharide heptosyltransferase 1	1,029
	5	Lipid A biosynthesis lauroyltransferase	879
	6	hypothetical protein	702
	18	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)- diNAcBac-PP-undecaprenol alpha-1,4-N-acetyl-D-galactosaminyltransferase	1,027
	39	hypothetical protein	780
	7	hypothetical protein	936
	2	Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase	1,420
	7	hypothetical protein	936
	2	hypothetical protein	1,281
	37	hypothetical protein	944
	41	N,N'-diacetyllegionaminic acid synthase	1,041
	38	UDP-N-acetylglucosamine 2-epimerase	1,119
	1	N-acylneuraminate cytidylyltransferase	666
	53	2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-acetyltransferase	755
	4	Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase	758
	60	ADP-heptose--LPS heptosyltransferase 2	944
CDC16	8	Lipopolysaccharide heptosyltransferase 1	1,029
	5	Lipid A biosynthesis lauroyltransferase	879
	2	hypothetical protein	1,224
	N	hypothetical protein	1,066
	7	hypothetical protein	872
	37	hypothetical protein	918
	41	N,N'-diacetyllegionaminic acid synthase	1,032
	38	UDP-N-acetylglucosamine 2-epimerase	1,116
	1	hypothetical protein	954
	1	N-acylneuraminate cytidylyltransferase	663
	26	hypothetical protein	705
	4	putative glycosyltransferase EpsJ	825
	60	ADP-heptose--LPS heptosyltransferase 2	944

Continued

LOS class	ORF	Products (annotated by prokka)	Length (bp)
CDC17	8	Lipopolysaccharide heptosyltransferase 1	1,029
	5	Lipid A biosynthesis lauroyltransferase	879
	2	hypothetical protein	1,224
	6	hypothetical protein	1,601
	18	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)- diNAcBac-PP-undecaprenol alpha-1,4-N-acetyl-D-galactosaminyltransferase	1,068
	13	hypothetical protein	1,148
	2	hypothetical protein	1,203
	1	hypothetical protein	1,058
	1	N-acylneuraminate cytidylyltransferase	663
	26	hypothetical protein	843
	4	putative glycosyltransferase EpsJ	825
	60	ADP-heptose--LPS heptosyltransferase 2	944
CDC18	8	Lipopolysaccharide heptosyltransferase 1	1,029
	5	Lipid A biosynthesis lauroyltransferase	879
	6	hypothetical protein	702
	18	GalNAc-alpha-(1->4)-GalNAc-alpha-(1->3)- diNAcBac-PP-undecaprenol alpha-1,4-N-acetyl-D-galactosaminyltransferase	811
	39	hypothetical protein	966
	7	hypothetical protein	936
	26	hypothetical protein	891
	4	putative glycosyltransferase EpsJ	810
	60	ADP-heptose--LPS heptosyltransferase 2	944
CDC19	8	Lipopolysaccharide heptosyltransferase 1	1,029
	5	Lipid A biosynthesis lauroyltransferase	879
	2	Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase	1,275
	6	hypothetical protein	1,560
	37	hypothetical protein	935
	41	N,N'-diacetyllegionaminic acid synthase	1,041
	38	UDP-N-acetylglucosamine 2-epimerase	1,119
	1	N-acylneuraminate cytidylyltransferase	798
	53	Streptogramin A acetyltransferase	615
	4	putative glycosyltransferase EpsJ	813
	60	ADP-heptose--LPS heptosyltransferase 2	944

Supplementary Table S4. The annotations of ORFs in CPS gene clusters of HS2, HS19, and HS41 of *C. jejuni*

CPS types	ORF	Product (annotated by prokka)	Length (bp)
HS2	21	Capsule polysaccharide modification protein	2,067
	1	adenylyl-sulfate kinase	513
	29	sugar nucleotidyltransferase	762
	30	amidotransferase	603
	27	hypothetical protein	2,340
	31	methyltransferase	762
	32	methyltransferase	774
	3	sugar transferase	1,839
	3	sugar transferase	1,878
	9	D-glycero-D-manno-heptose 1-phosphate guanosyltransferase	666
	11	phosphoheptose isomerase	606
	17	D-glycero-D-manno-heptose 7-phosphate kinase	1,020
	110	methyltransferase family protein	852
	54	sugar-nucleotide epimerase/dehydratease	942
	10	GDP-L-fucose synthetase	1,041
	52	hypothetical protein	927
	24	dTDP-4-dehydrorhamnose 3,5-epimerase	546
	14	capsular polysaccharide heptosyltransferase	1,749
	N	sugar transferase	3,096
	N	hypothetical protein	1,107
	22	sugar transferase	1,335
	77	phosphatase	636
	N	aminotransferase	1,230
	78	aminotransferase	1,104
	22	sugar transferase	2,328
	20	UDP-galactopyranose mutase	1,107
	2	sugar transferase	1,224
	16	UDP-glucose 6-dehydrogenase	1,182
	28	sugar transferase	1,635
	12	D-arabinose 5-phosphate isomerase	495
HS19	21	Capsule polysaccharide modification protein	2,067
	1	adenylyl-sulfate kinase	513
	29	sugar nucleotidyltransferase	762
	30	amidotransferase	603
	27	hypothetical protein	2,340
	31	methyltransferase	762
	32	methyltransferase	774
	3	sugar transferase	1,833
	2	hypothetical protein	1,509
	77	phosphatase	633
	78	aminotransferase	1,077
	2	hypothetical protein	2,499
	16	UDP-glucose 6-dehydrogenase	1,176
	28	hypothetical protein	2,124
	12	D-arabinose 5-phosphate isomerase	495

Continued

CPS types	ORF	Product (annotated by prokka)	Length (bp)
	21	Capsule polysaccharide modification protein	2,067
	46	Na(+)/H(+)-K(+) antiporter GerN	1,254
	47	pyruvate kinase	960
	48	sulfate adenylyltransferase	1,047
	1	adenylyl-sulfate kinase	534
	3	sugar transferase	1,854
	42	hypothetical protein	1,749
	9	D-glycero-D-manno-heptose 1-phosphate guanosyltransferase	666
	11	phosphoheptose isomerase	606
	59	hypothetical protein	1,236
	17	D-glycero-D-manno-heptose 7-phosphate kinase	1,020
	19	NAD-dependent 4,6-dehydratase	1,032
	10	GDP-L-fucose synthase	2,214
	24	dTDP-4-dehydrorhamnose 3,5-epimerase	546
	55	hypothetical protein	1,251
HS41	14	capsular polysaccharide heptosyltransferase	1,608
	20	UDP-galactopyranose mutase	1,131
	61	NAD-dependent 4,6-dehydratase	885
	87	dTDP-4-dehydrorhamnose 3,5-epimerase	630
	84	NAD-dependent 4,6-dehydratase	1,071
	57	Glucose-1-phosphate cytidylyltransferase	801
	43	hypothetical protein	2,157
	13	hypothetical protein	1,380
	20	UDP-galactopyranose mutase	1,170
	23	UDP-GlcNAc/Glc 4-epimerase	975
	16	UDP-glucose 6-dehydrogenase	1,299
	15	NAD-dependent 4,6-dehydratase	957
	2	hypothetical protein	1,371
	22	sugar transferase	1,050
	28	sugar transferase	1,632
	12	D-arabinose 5-phosphate isomerase	495