

Novel Capsular Polysaccharide Loci and New Diagnostic Tools for High-Throughput Capsular Gene Typing in *Streptococcus suis*

Xiaotong Qiu,^a Xuemei Bai,^a Ruiting Lan,^b Han Zheng,^a Jianguo Xu^a

State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Changping, Beijing, China^a; School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, New South Wales, Australia^b

ABSTRACT

Streptococcus suis is an important pathogen of pigs and may cause serious disease in humans. Serotyping is an important tool for detection and epidemiological studies of *S. suis*. Thirty-three reference serotypes and nine novel *cps* loci (NCLs) are recognized in *S. suis*. To gain a better understanding of the prevalence and genetic characteristics of NCLs, we investigated the serotype identity of 486 isolates isolated between 2013 and 2015 in China by capsular gene typing methods. Two hundred seventy-six isolates carried NCLs belonging to 16 groups, 8 of which appear to have not been reported previously. These isolates showed autoagglutination, polyagglutination, or nonagglutination with reference antisera and thus were nonserotypeable. Almost all isolates carrying the unknown NCLs were encapsulated, with various capsular thicknesses, indicating that they are most likely novel serotypes. To simultaneously identify the currently recognized 17 NCLs, an 18-plex detection system using the Luminex xTAG universal array technology was developed. Our data also provide valuable genetic information for monitoring the variations within NCLs by investigating the genetic characteristics of different subtypes within NCLs.

IMPORTANCE

Nonserotypeable *Streptococcus suis* isolates have been reported in many studies, and 9 novel *cps* loci (NCLs) have already been identified in nonserotypeable isolates. Moreover, novel *cps* loci are continually being found. The main purpose of this study was to investigate the prevalence and characteristics of NCLs in *S. suis* isolates recovered between 2013 and 2015 in China. This study provides valuable genetic information for monitoring the variations within NCLs. Meanwhile, a fast and cost-effective 18-plex detection system that can simultaneously identify the currently recognized 17 NCLs was developed in this study. This system will serve as a valuable tool for detecting known and identifying additional novel *cps* loci among nonserotypeable *S. suis* isolates.

Streptococcus suis is an important pathogen of pigs and may cause serious disease in humans (1–3). Clinically healthy pigs can carry *S. suis* in their nasopharynx, contributing to the dissemination of this pathogen (4, 5). Moreover, a potential role of isolates from healthy pigs is as donors to transfer the new adaptive phenotypes to pathogenic isolates (6). The capsular polysaccharide (CPS) shields *S. suis* from host phagocytes and is a major virulence factor (7). Antisera to CPSs are used to distinguish antigenic differences among them. Serotyping is also an important epidemiological method for *S. suis* surveillance. This reflects the importance of including isolates from healthy pigs to better understand the diversity and evolution of CPSs in *S. suis* populations.

A total of 35 serotypes (types 1 through 34 and type 1/2) of *S. suis* have been identified by different studies in the 1980s and 1990s (8–11). In 2005, strains of serotypes 32 and 34 were reclassified as *Streptococcus orisratti* (12). The CPS synthesis genes are known to be clustered at the *cps* locus. CPSs of all serotypes are thought to be synthesized by the Wzx/Wzy pathway. Based on serotype-specific *wzy* genes, 3 multiplex capsular gene typing systems have been developed to identify *S. suis* serotypes (13–15). These systems allow the simultaneous detection of multiple nucleic acid sequences in a single reaction and can greatly reduce the time, cost, and work associated with conventional serotyping technologies. They have become attractive alternatives to existing serological tests, especially for nonserotypeable isolates, which can be more conveniently identified.

Recently, 9 novel *cps* loci (NCLs) were identified in nonsero-

typeable isolates and were designated NCL1 to -8 (16) and Chz (17). These loci possessed the specific polysaccharide polymerase gene *wzy*, the flippase gene *wzx*, and glycosyltransferase (GT) and acetyltransferase genes, which were obviously different from those of reference serotypes. All evidence from genetic analyses and phenotypes strongly supported that they represented novel serotypes of *S. suis*. Based on the variable presence of 13 genes and 4 transposase genes, multiple subtypes of NCL1, -2, -3, -7, and -8 were found (16). Those studies showed the high level of diversity within the same NCL and underscore the importance of investigating the subtypes of different NCLs.

In this study, 486 isolates recovered from pigs in China between 2013 and 2015 were typed by capsular gene typing systems. Two hundred seventy-six isolates carried NCLs belonging to 16 groups, 8 of which appear to have not been reported previously

Received 17 July 2016 Accepted 24 September 2016

Accepted manuscript posted online 30 September 2016

Citation Qiu X, Bai X, Lan R, Zheng H, Xu J. 2016. Novel capsular polysaccharide loci and new diagnostic tools for high-throughput capsular gene typing in *Streptococcus suis*. *Appl Environ Microbiol* 82:7102–7112. doi:10.1128/AEM.02102-16.

Editor: M. A. Elliot, McMaster University

Address correspondence to Han Zheng, zhenghan@icdc.cn.

Supplemental material for this article may be found at <http://dx.doi.org/10.1128/AEM.02102-16>.

Copyright © 2016, American Society for Microbiology. All Rights Reserved.

TABLE 1 Subtype-specific HG arrangement of NCL1, NCL2, NCL3, NCL7, and NCL8^a

Subtype-specific gene type	Presence of HG in subtype																						
	NCL1										NCL2					NCL3		NCL7		NCL8			
	1-1	1-2	1-3	1-4	1-5 ^b	1-6	1-7	1-8	1-9	1-10	1-11	2-1	2-2	2-3	2-4	2-5	3-1	3-2	7-1	7-2	8-1	8-2	8-3
HG55	-	+	+	-	-	+	-	-	+	+	-	+	-	-	+	-	/	/	/	/	/	/	/
HG292	+	+	+	+	+	+	+	+	+	+	+	+	+	-	+	+	/	/	/	/	/	/	/
HG293	-	+	-	-	-	+	-	+	-	-	-	-	-	-	+	+	/	/	/	/	/	/	/
HG294	-	+	-	-	-	+	-	+	-	-	-	-	-	-	+	+	/	/	/	/	/	/	/
HG332	-	-	-	-	-	-	-	-	-	+	-	+	-	-	-	-	/	/	/	/	/	/	/
HG312	+	-	+	+	+	-	+	-	+	-	+	/	/	/	/	/	/	/	/	/	/	/	/
HG313	+	-	+	+	+	-	+	-	+	-	+	/	/	/	/	/	/	/	/	/	/	/	/
HG314	+	-	+	+	+	-	+	-	-	-	-	/	/	/	/	/	/	/	/	/	/	/	/
HG315	+	-	+	+	+	-	+	-	-	-	-	/	/	/	/	/	/	/	/	/	/	/	/
HG329	+	-	-	-	-	-	+	-	-	-	-	/	/	/	/	/	/	/	/	/	/	/	/
HG354	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	-	+	/	/	/
HG355	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	-	+	/	/	/
HG23	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	/	-	+	-	-
Transposase genes ^c	-	-	-	-	-	+	+	-	-	-	-	/	/	/	/	/	-	+	/	/	-	-	+

^a +, positive; -, negative; /, no need for detection.

^b The PCR product size of HG292 in NCL1-5 is 724 bp.

^c Including 4 subtype-specific transposase genes for NCL1-6, NCL1-7, NCL3-2, and NCL8-3, which are listed in Table 2.

and are named NCL9 to -16. All 276 isolates showed autoagglutination, polyagglutination, or nonagglutination with reference antisera and thus were nonserotypeable. The capsular phenotypes of 39 isolates carrying NCL9 to -16 and 10 isolates carrying Chz were investigated by transmission electron microscopy. Based on NCL-specific *wzy* gene sequences, we developed an 18-plex detection system by using multiplex PCR (mPCR) and the Luminex xTAG universal array technology in combination. This system can simultaneously identify the 17 currently recognized NCLs in a single reaction. The genetic variations within NCLs were also investigated. In order to determine the phylogenetic position of isolates carrying NCLs in the *S. suis* population, minimum core genome (MCG) typing was used to analyze these isolates.

MATERIALS AND METHODS

Bacterial strains and chromosomal DNA preparation. A total of 454 field isolates from the nasopharynx of healthy pigs and 32 isolates from lungs of diseased pigs were used in this study and are listed in Table S1 in the supplemental material. All isolates were serotyped by using the agglutination test (serum provided by Statens Serum Institute, Copenhagen, Denmark). Chromosomal DNA was prepared by using a method described previously (15). The species identity of the 486 isolates was determined to be *S. suis* by amplification of the 16S rRNA, *recN*, and *thrA* genes (13, 18, 19).

Capsular gene typing and determination of novel subtypes of NCL1, NCL2, NCL3, NCL7, and NCL8. Capsular gene typing methods developed in our laboratory were used to assign *cps* loci of 486 isolates (13, 16, 17, 20). Known subtypes of NCL1, NCL2, NCL3, NCL7, and NCL8 were determined based on the variable presence of 13 genes and 4 transposase genes found in our previous study by PCR amplification (16) (Tables 1 and 2).

Sequencing of *cps* loci and bioinformatics analysis. Isolates ($n = 39$) that could not be assigned to NCL1 to -8, isolates ($n = 10$) that were assigned to Chz, and isolates ($n = 7$) of novel subtypes of NCL1 and NCL2 were sequenced by Illumina sequencing as previously described (21). Each *cps* locus sequence was extracted from the draft genome sequence, and open reading frames (ORFs) were identified and annotated based on methods reported previously (22). *cps* loci were analyzed by using the same bioinformatics methods as those described in previous studies (16,

22). The *cps* genes were assigned to the 361 known homology groups (HGs) of 35 serotype reference strains (22) and NCL1 to -8 (16) if the length of the match was >50% of the sequence and the identity was also >50%. Genes that did not match these criteria were assigned to novel HGs and numbered from HG362 onwards. Novel HGs that were present in all isolates of a NCL are referred to as NCL-specific HGs.

Detection of the capsule by transmission electron microscopy. The presence or absence of the capsule was examined by electron microscopy using the same method as the one described in our previous study (16).

The capsular thickness was measured between the inner and external edges of the capsular layer. Each value is based on 25 to 30 measurements per experiment. Each experiment was done twice independently.

MCG typing. MCG typing was performed by using PCR amplification and DNA sequencing as described previously (23).

Development of a high-throughput detection system for the 17 NCLs. The 18-plex detection system was set up based on a protocol described in a previous study (13), and the primers used to amplify *S. suis* NCL-specific *wzy* target sequences are listed in Table 3. In brief, the unique "TAG" sequence and biotin label were specifically incorporated into *wzy*-specific amplification products in an mPCR reaction using cycling parameters of 94°C for 5 min and 30 cycles of 94°C for 30 s, 58°C for 30 s, and 72°C for 30 s, followed by a final elongation step at 72°C for 10 min. After amplification, a volume containing 5 µl of each PCR product or H₂O, 20 µl of MagPlex-TAG microsphere mixture (125 of each microsphere set per µl), and 75 µl of reporter solution (10 µg/ml streptavidin-R-phycoerythrin conjugate [SAPE] in 1× Tm hybridization buffer [Luminex]) was hybridized in a thermocycler for 30 min at 40°C. The hybridization solutions were read by using the Bioplex 200 instrument (Bio-Rad), and the median fluorescence intensity (MFI) data were analyzed by using Luminex xPONENT 3.1 software. The arbitrary MFI of a positive signal was defined as ≥2 times the MFI value of the background (consisting of all reaction mixture components except DNA). The threshold of the detection limit was determined by using serially diluted DNA from a representative isolate of each NCL type. Non-*S. suis* isolates used in our previous study (13) were also used to determine the specificity of the system in the study. Two independent experiments were performed to establish the sensitivity and specificity of the system.

Accession number(s). DNA sequences obtained in this study were deposited in GenBank under accession numbers KT163361 to KT163377, KT802744, KU665257 to KU665288, and KU983471 to KU983476.

TABLE 2 Primers used to determine the subtypes of NCL1, -2, -3, -7, and -8

Target gene type	Primer direction, sequence (5'–3')	Annealing temp (°C)	PCR product size (bp)	Description
HG23	Forward, TTTTCAGGGTATGATTCT Reverse, CATGTACTTCAGCTATTG	45.5	550	NAD-dependent epimerase/dehydratase
HG55	Forward, TGCTTCATCATCTTCGGT Reverse, TGTTGCTTATTAGCCTTG	48.2	846	Hypothetical protein
HG292	Forward, TGACTTTATTGATACCTT Reverse, TATTTCTGTTTAGCAGAT	45.3	572	Hypothetical protein
HG293	Forward, TCTTGGTTTCGTTTCTAC Reverse, GTTTTCTTATTTTTGTT	43.5	352	Hypothetical protein
HG294	Forward, AATATAGCTGGGAGTGTG Reverse, TGCGGGTATGTTAAAAAG	45.4	300	Hypothetical protein
HG312	Forward, TAGAAGTGCTTATTTTGT Reverse, CTTGAAGTCCATTGCCTG	45	384	Integral membrane protein
HG313	Forward, TTGATTAAGTTGGAGGT Reverse, CCCAGATAGAATAGAGCC	44.3	396	Nucleotidyltransferase family protein
HG314	Forward, AAGAGGAGCTTATTGAGG Reverse, AAATGTCTGCTGTTGTGG	46	402	Phosphorylcholine transferase
HG315	Forward, AGGTAGACGAACTATCCC Reverse, CCCCATCTGTAACCAAAA	47	506	Choline kinase
HG329	Forward, ATGGCTGGCTACTATGCT Reverse, AATGTCGTTCCGTTCTCT	48.6	206	Hypothetical protein
HG332	Forward, CCACTACAATAACCCTTC Reverse, TAAAACCTCCATCTCCTCA	46.2	968	Hypothetical protein
HG354	Forward, TAATGAATCAACGCAAAC Reverse, CAAAAGCAGAATAAACCC	46.1	212	Hypothetical protein
HG355	Forward, TCTTGTGGATTTATTCG Reverse, TTAACCTTTGGCCTGCTTT	45.3	328	Hypothetical protein
Transposase (NCL1-6)	Forward, GCTTTACCCAATCTATGC Reverse, TGACCTTCTACTGGACCT	49.2	1,792	Transposase
Transposase (NCL1-7)	Forward, CAGAAGAGAAGTGGAAGT Reverse, TAGAACAAAGTGTGCGATA	47.9	1,446	Transposase
Transposase (NCL3-2)	Forward, CATTTAGGATTGACGAAG Reverse, TTTTACCACCCAACACAG	48.8	950	Transposase
Transposase (NCL8-3)	Forward, GATAGAGCGAATAGGGTA Reverse, ATTTCAAAAAAGTGGAC	48.1	1,020	Transposase

RESULTS

Serotype identification of isolates. The 486 isolates were typed by both agglutination tests using reference antisera and capsular gene typing systems. By agglutination tests, 195 isolates were typed as being of a known serotype, and 291 isolates showed autoagglutination, polyagglutination, or nonagglutination with the reference antisera and thus were nonserotypeable. All 486 isolates were also typed by using a previously developed 32-plex Luminex assay that can type isolates of traditional serotypes (13). Two hundred ten isolates were typed as the reference serotypes, 195 of which were

the same serotypes as those determined by the agglutination test. However, 15 isolates were typed into serotypes 5, 12, 29, and 30 by the 32-plex Luminex assay, which were nonserotypeable by the agglutination test. The 276 isolates that were not typeable by the 32-plex Luminex assay were further typed by using multiplex PCR for NCL1 to -7 (20) and simplex PCR for NCL8 (16) and Chz (17), 237 isolates of which were typed as known NCLs. However, 39 isolates remained untypeable, suggesting that these isolates carry unknown NCLs (Table 4; see also Table S1 in the supplemental material).

TABLE 3 Primers used with the 18-plex detection system^a

<i>cps</i> locus type	No. of beads coupled with anti-TAG sequences	Forward and reverse primer sequences (5'-3')	Working concn (μM)	Sensitivity (pg)	PCR product size (bp)
NCL1	54	<u>TTAATACAATTCTCTCTTTCTCTA</u> -C ₁₂ -ATTCACATAGTAACATTGCGGA *CAACATTGCGCAGGAAATAATA	0.2	10	254
NCL2	55	<u>ACATCAAATTCTTTCAATATCTTC</u> -C ₁₂ -TTCTTGATTATGCTGTTCTCGT *AAACACAACATCCTGTACTIONCA	0.2	0.5	333
NCL3	56	<u>CTTAAACTCTACTTACTTCTAATT</u> -C ₁₂ -ATTCAGGAGGTATCAACCAAG *AATTCACTAGCATCAACAAACG	0.2	10	370
NCL4	57	<u>ACTTACAATAACTACTAATACTCT</u> -C ₁₂ -TGCTTATTATGACTGTTGCCTT *ATCAGTTGATAAGGTTGCTGTT	0.2	10	293
NCL5	61	<u>AATCTCTACAATTTCTCTCTAATA</u> -C ₁₂ -CAGATGAGTCAGCAAGTAATCA *AGGGAAGAGTAAGATTCAAGGT	0.2	1	262
NCL6	62	<u>CTAAACATACAAATACACATTTCA</u> -C ₁₂ -TGATACGGGTACTGTTGAGTAT *ATTACTACTTCTGGTTGGGTCA	0.2	0.5	220
NCL7	63	<u>CTAAATCACATACTTAAACAACAAA</u> -C ₁₂ -GTTGATTTATTTGCGGGACTAC *CAGAAAAACAATAGCAGTGACC	0.2	1	447
NCL8	64	<u>TTCAATTCAAATCAAACACATCAT</u> -C ₁₂ -AAAATTTCACTTCACCTCGAC *AATCTTCCAATCAATGCTACGA	0.2	20	390
NCL9	65	<u>TACTTAAACATACAAACTTACTCA</u> -C ₁₂ -GGGTAGATACGTTCTATTGGG *GCGACGGTATATAGAGCTATTC	0.2	0.5	309
NCL10	66	<u>TCTTACTAATTTCAATACTCTTAC</u> -C ₁₂ -ACAAAAATAGTAGACGGGCTTT *TTGCACGCCAAGTATAAAAATTC	0.2	0.5	231
NCL11	67	<u>ATCTCAATTACAATAACACACAAA</u> -C ₁₂ -GCCTTAATAATGGTGGGTTTTG *TACCTAAAACATTTTGCCAGA	0.2	1	454
NCL12	72	<u>CTATCATTATCTCTTTCTCAATT</u> -C ₁₂ -GTGAGAGATTTCCGGTGTAGTTT *GCATCAGCATACATCTTCTCA	0.2	10	380
NCL13	73	<u>CTTTATCAAATTCTAATTCTCAAC</u> -C ₁₂ -TCTTGCTAGGTCTAATCGTAGT *CTCGCTTCCAATTAATAAACCG	0.2	0.5	179
NCL14	74	<u>ACACTCATTTAACACTATTTTCATT</u> -C ₁₂ -AAAAGAAATGGAAAGCAGTGTG *TCTTTGCTCAGCTATTGAGTTT	0.2	0.5	178
NCL15	75	<u>CATAAATCTTCTCATTCTAACAAA</u> -C ₁₂ -TATGCTATTGTTACGATGTGGG *CTTGGAGAGTAAAACGATAGGG	0.2	0.5	169
NCL16	76	<u>TCTCATCTATCATACTAATTCTTT</u> -C ₁₂ -AGGGTTATTCTTTTTGGTGGAT *TCTTGTAAAGCAGAAAATCGTGA	0.2	0.5	200
Chz	77	<u>AATAACAACACTACTATATCATAAC</u> -C ₁₂ -ATTTTGGGCAGTGAGTGATATT *TCAATACTTTTCTTGAACCCGA	0.2	0.5	374
All (<i>thrA</i>)	33	<u>ACTACTTATTTCTCAAACCTCAATA</u> -C ₁₂ -GAAAATATGAAGAGCCATGTCG *GACAACGAACATAACGAAACTTC	0.3	0.5	120

^a xTAG sequences and the 12-carbon amine-containing group are indicated by underlining and boldface type, respectively. * indicates that the reverse primer is biotinylated at the 5' terminus (i.e., it indicates the beginning of the reverse primer sequence).

Of the 237 isolates carrying a known NCL, NCL1 ($n = 82$), NCL2 ($n = 43$), NCL3 ($n = 42$), NCL7 ($n = 35$), and NCL8 ($n = 19$) constituted 93.2% (221/237) of them. Chz ($n = 10$), NCL4 ($n = 4$), and NCL5 ($n = 2$) were also found.

Identification of eight new NCLs. The 39 isolates that could not be assigned to the 33 reference serotypes and 9 known NCLs were sequenced by Illumina sequencing. The *cps* gene clusters were extracted from the genome sequence and were divided into 8

TABLE 4 Serotype identification of 486 isolates using antisera to reference serotypes and capsular gene typing assays^a

Serotype(s) determined by capsular gene typing assays	No. of isolates	No. of isolates determined by using antiserum to serotype:																			
		2	3	4	5	8	9	11	12	13	16	17	19	21	23	24	26	28	29	30	31
32-plex Luminex assay																					
1/2 and 2	7	7																			
3	6		6																		
4	5			5																	
5	17				14																3
8	4					4															
9	1						1														
11	4							4													
12	19								13												6
13	3									3											
16	7										7										
17	5											5									
19	10												10								
21	10													10							
23	6														6						
24	12															12					
26	1																1				
28	1																	1			
29	61																		57		4
30	25																			23	2
31	6																				6
18-plex Luminex assay																					
NCL1	82																				82
NCL2	43																				43
NCL3	42																				42
NCL4	4																				4
NCL5	2																				2
NCL7	35																				35
NCL8	19																				19
NCL9	1																				1
NCL10	1																				1
NCL11	26																				26
NCL12	3																				3
NCL13	2																				2
NCL14	1																				1
NCL15	2																				2
NCL16	3																				3
Chz	10																				10

^a NT, nontypeable (i.e., nonserotypeable strains).

different NCLs based on the sequence of the *wzy* gene (see Fig. 2A). The NCLs were named NCL9 to -16. NCL11 ($n = 26$) was the most prevalent, followed by NCL12 ($n = 3$), NCL16 ($n = 3$), NCL13 ($n = 2$), and NCL15 ($n = 2$). The remaining 3 NCLs contained only a single isolate.

All 8 NCLs were flanked by the *orfZ-orfX* region and the *glf* gene (UDP-galactopyranose mutase) and were classified into pattern I-b (16, 22). The G+C content of the NCLs ranged from 33.18% to 36.76%. The sizes ranged from 21.12 kb to 32.87 kb. The sialic acid synthesis genes were not found in any of these NCLs.

The *cps* genes from the 8 NCLs were grouped into 92 HGs, as done previously (16, 22). Thirty-six HGs contained known genes of serotype reference strains, NCL1 to -8 and Chz, including GT, acetyltransferase, aminotransferase, nucleotidyltransferase, phosphotransferase, cytidyltransferase, glucuronate epimerase, and glucose dehydrogenase. The *cpsA*, *cpsB*, *cpsC*, and *cpsD* genes are

present and located on the 5' side of all 8 new NCLs. An initial sugar transferase gene was located in the 5'-side region and was classified into 3 HGs: HG8 (NCL10 and NCL15), HG21 (NCL11, NCL12, and NCL16), and HG295 (NCL9, NCL13, and NCL14).

The 3' side of 7 NCLs (excluding NCL15) is relatively conserved and contained predominantly HG7, HG34, HG55, HG292, HG293, and HG294, while central regions of the 8 NCLs are highly variable (see Fig. 2A).

Fifty-three HGs (HG363 to HG415) were NCL specific, and nearly all NCL-specific genes are present in the central region of novel *cps* clusters. Each NCL contained 5 to 11 NCL-specific genes, with 11 HGs for NCL9, 6 HGs for NCL10, 6 HGs for NCL11, 5 HGs for NCL12, 8 HGs for NCL13, 5 HGs for NCL14, 9 HGs for NCL15, and 4 HGs for NCL16. Among them, 21 HGs encode putative glycosyltransferases, and 5 encode acetyltransferases. All Wzy polymerases were NCL specific, as expected. Seven of the eight Wzx flippases were NCL specific, with the *wzx*

gene of NCL12 being homologous to that of NCL1 (HG301). Interestingly, NCL16 contained two types of *wzx* genes (HG301 and HG417).

Only 3 novel HGs were not composed of NCL-specific genes. HG362 was present in both NCL12 and Chz-2. Two HGs (HG416 and HG417) were present only in NCL16 but not in all isolates of NCL16 (see Table S2 in the supplemental material).

Development and evaluation of a Luminex-based high-throughput detection system for the 17 NCLs. Detection was based on the unique sequence of *wzy* for each NCL. The *wzy* gene was amplified in a multiplex PCR format. The detection limit for the 17 NCLs varied from 0.5 pg of purified DNA/reaction (equivalent to $\sim 2 \times 10^2$ CFU/reaction) to 20 pg of purified DNA/reaction (equivalent to $\sim 1 \times 10^4$ CFU/reaction).

The performance of the system was tested on 486 isolates used in this study. Cross-hybridization and nonspecific hybridization between sequences were not observed. Two hundred seventy-six isolates carrying novel *cps* loci were correctly assigned to the corresponding type. All 210 isolates were typed as the reference serotypes by the 32-plex Luminex assay, and 62 non-*S. suis* isolates did not give a positive signal.

Determination of subtypes of NCLs. NCL3 and NCL8 isolates were found to belong to a single subtype, NCL3-1 and NCL8-1, respectively. Genetic heterogeneity was not found within NCL12 and NCL15 isolates.

(i) **NCL1.** Of 82 isolates, 76 were assigned to 6 known subtypes of NCL1 (NCL1-1, NCL1-2, NCL1-3, NCL1-4, NCL1-5, and NCL1-7). Six isolates could not be assigned to known subtypes. Five of them were sequenced by Illumina sequencing, and four novel subtypes were found and named NCL1-8, NCL1-9, NCL1-10, and NCL1-11. The subtypes differed at the 3' ends with indels of HG55, HG293, HG294, HG312, HG313, HG314, HG315, or HG332 (Table 1 and Fig. 1A). NCL1-1 ($n = 24$), NCL1-2 ($n = 23$), and NCL1-4 ($n = 23$) were dominant subtypes.

(ii) **NCL2.** Of 43 isolates, only 8 were assigned to 2 known subtypes of NCL2 (NCL2-1 and NCL2-2). Thirty-five isolates could not be assigned to known subtypes. Two novel subtypes were found among them by Illumina sequencing and named NCL2-4 and NCL2-5. The novel subtypes varied due to the variable presence of HG55, HG293, and HG294 (Table 1 and Fig. 1B). NCL2-4 ($n = 32$) was the dominant subtype.

(iii) **NCL7.** For NCL7, two types of genetic organization were found in 35 isolates, with NCL7-1 containing 32 isolates and NCL7-2 containing 3 isolates.

(iv) **Chz.** Sequencing of 10 isolates showed that they differ from Chz reference strain CZ130302 (Chz-1) and were named Chz-2. One gene (HG55) was inserted, and three genes (HG293, HG294, and HG362) were deleted in the 3'-side region of Chz-2 (Fig. 1C).

(v) **NCL11.** Five types of genetic organization were found in NCL11 because of insertions and deletions (Fig. 2B): NCL11-1 ($n = 17$), NCL11-2 ($n = 5$), NCL11-3 ($n = 2$), NCL11-4 ($n = 1$), and NCL11-5 ($n = 1$). Compared to NCL1-1, two different transposase genes were inserted in NCL11-2 and NCL11-3, whereas deletions of HG55, HG292, HG293, and HG294 in NCL11-4 and a deletion of HG55 in NCL11-5 were found.

(vi) **NCL16.** Two types of genetic organizations were found in NCL16 (Fig. 2C): NCL16-1 ($n = 2$) and NCL16-2 ($n = 1$). NCL16-1 differed from NCL16-2 in the 3' side, with HG416,

HG417, HG293, and HG294 in NCL16-1 and HG301, HG314, HG315, HG312, HG313, and HG329 in NCL16-2.

Determination of the presence of CPS in the NCL9 to NCL16 and Chz isolates. The thickness of the capsules of 39 isolates of NCL9 to NCL16 and 10 isolates of Chz-2 was measured by transmission electron microscopy (representative isolates are shown in Fig. 3; see also Table S1 in the supplemental material). Strain SC84 (serotype 2) was used as a control and was well encapsulated, with a capsular thickness of 110 nm to 130 nm. Forty-seven isolates were encapsulated with various capsular thicknesses (Table S1). Capsular thicknesses also showed wide variation within an NCL or NCL subtypes, but capsular thicknesses from the same isolate were consistent based on two independent experiments. Only two isolates (YS196 and YS351) distributed in NCL12 and Chz-2 were likely to be nonencapsulated (see Table S1 in the supplemental material).

MCG typing. The 276 isolates carrying NCLs were typed by MCG typing. The isolates were in MCG group 6 (83.7%; 231/276) and group 7 (12%; 33/276 isolates) and were ungroupable (4.3%; 12/276 isolates) (see Table S1 in the supplemental material).

DISCUSSION

Serotyping is an important tool for detection and epidemiological studies of *S. suis*. Although the prevalence of nonserotypeable *S. suis* isolates has been reported in many studies (4, 14, 24–27), little was known about the distribution and diversity of NCLs of nonserotypeable isolates. The main purpose of this study was to investigate the prevalence, characteristics, and evolution of NCLs in the *S. suis* population.

In this study, the 486 *S. suis* isolates from pigs were analyzed for their serotype identity by capsular gene typing methods, and the performance of the methods was evaluated by the seroagglutination test. Of the 210 isolates typed as the reference serotypes by the 32-plex Luminex assay, the agglutination results for 195 isolates were completely consistent with their 32-plex Luminex assay results. Only 15 isolates identified as belonging to serotype 5, 12, 29, or 30 by the 32-plex Luminex assay could not be identified in the agglutination test. Similar results were also found in isolates from the United Kingdom, the genomes of which have been sequenced (6). Serotypes of 46 isolates were not identified in agglutination tests using antisera to the reference serotypes. We analyzed the *cps* sequences of these 46 isolates from the genome sequences available in GenBank (see Table S3 in the supplemental material). We found that except for 8 isolates that carried a NCL, the *cps* types of the remaining 38 isolates were assigned to reference serotypes 6, 8, 9, 10, 15, 16, 19, 21, 24, and 31 by *in silico* typing using the *cps* gene sequence (see Table S3 in the supplemental material). Antigenic differences or a deficiency in CPS may attribute to one or more of the defects in the *cps* gene cluster or genes outside the *cps* locus involved in antigenic modification, which remain to be identified. Therefore, for a small proportion of isolates, discrepancies between the results of agglutination test-based serotyping (phenotyping) and capsular gene-based serotyping (genotyping) are expected.

In the present study, 94.8% of the nonserotypeable isolates (276/291) carried 1 of the 17 NCLs. It is noteworthy that 8 lung isolates carried 4 NCLs: NCL1, -3, -7, and -11. Using discriminant analysis of principal components, many lung isolates were genetically similar to systemic isolates (6) and also shared identical pulsed-field gel electrophoresis (PFGE) types with invasive iso-

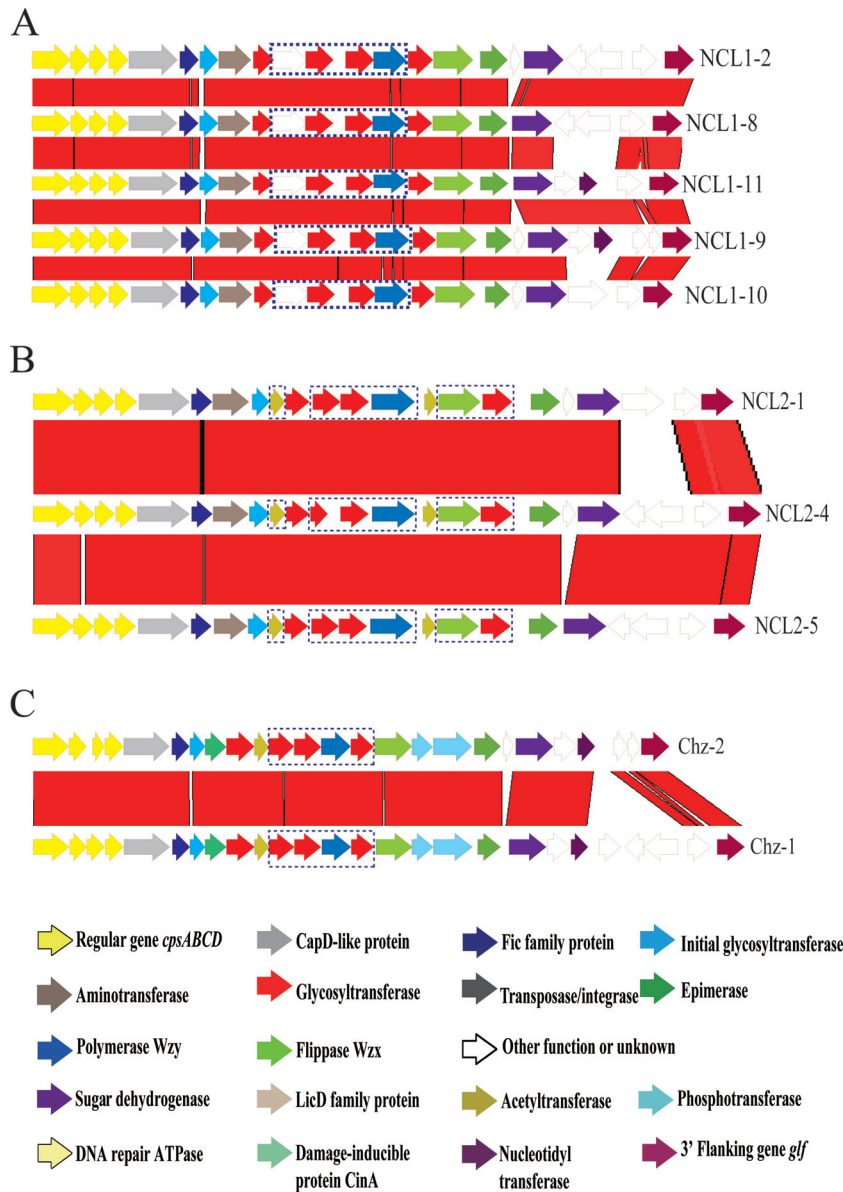


FIG 1 Comparison of the *cps* loci within NCL1 (A), NCL2 (B), and Chz (C). Each colored arrow represents a gene whose predicted function is shown at the bottom. NCL-specific genes are indicated by dotted blue lines. The *glf* gene is located on the 3' side of each locus.

lates (28). These findings indicated that many lung isolates possessed the potential to cause systemic infection, suggesting that isolates carrying these 4 NCLs are potentially pathogenic. In 2013, serotype Chz isolates caused an outbreak of streptococcosis in piglets at multiple large-scale pig farms in Jiangsu Province, China (17). The high proportion of isolates carrying NCL1, -3, -7 and -11 and Chz in the present study indicates that they are prevalent in the swine population and highlights the need for increased surveillance of *S. suis* isolates carrying these 5 NCLs in China.

In this study, we also developed an 18-plex detection system based on specific *wzy* genes of the NCLs using the Luminex xTAG universal array technology, which can simultaneously identify the 17 currently recognized NCLs. The Luminex xMAP system is a multiplexed microsphere-based suspension array

platform. mPCR coupled with Luminex xTAG technology-based detection provides an open and attractive approach for multiplexed analysis. The detection system was validated with a panel of 486 tested isolates and 62 non-*S. suis* isolates.

The 18-plex detection system, as a high-throughput, low-time-consuming assay, can be completed 40 min after PCR amplification. Moreover, this system has great potential to increase multiplicity in a single reaction in which new NCLs are continuously found. For example, one nonserotypeable *S. suis* isolate from pigs in the United Kingdom (see Table S3 in the supplemental material) did not belong to any of the 17 NCLs (data not shown).

In the present study, NCL9 to -16, distributed in 39 isolates, appear to have not been reported previously. Most of these isolates expressed a capsule. In addition, 53 NCL-specific HGs were identified among these 8 NCLs. The functions of these type-specific

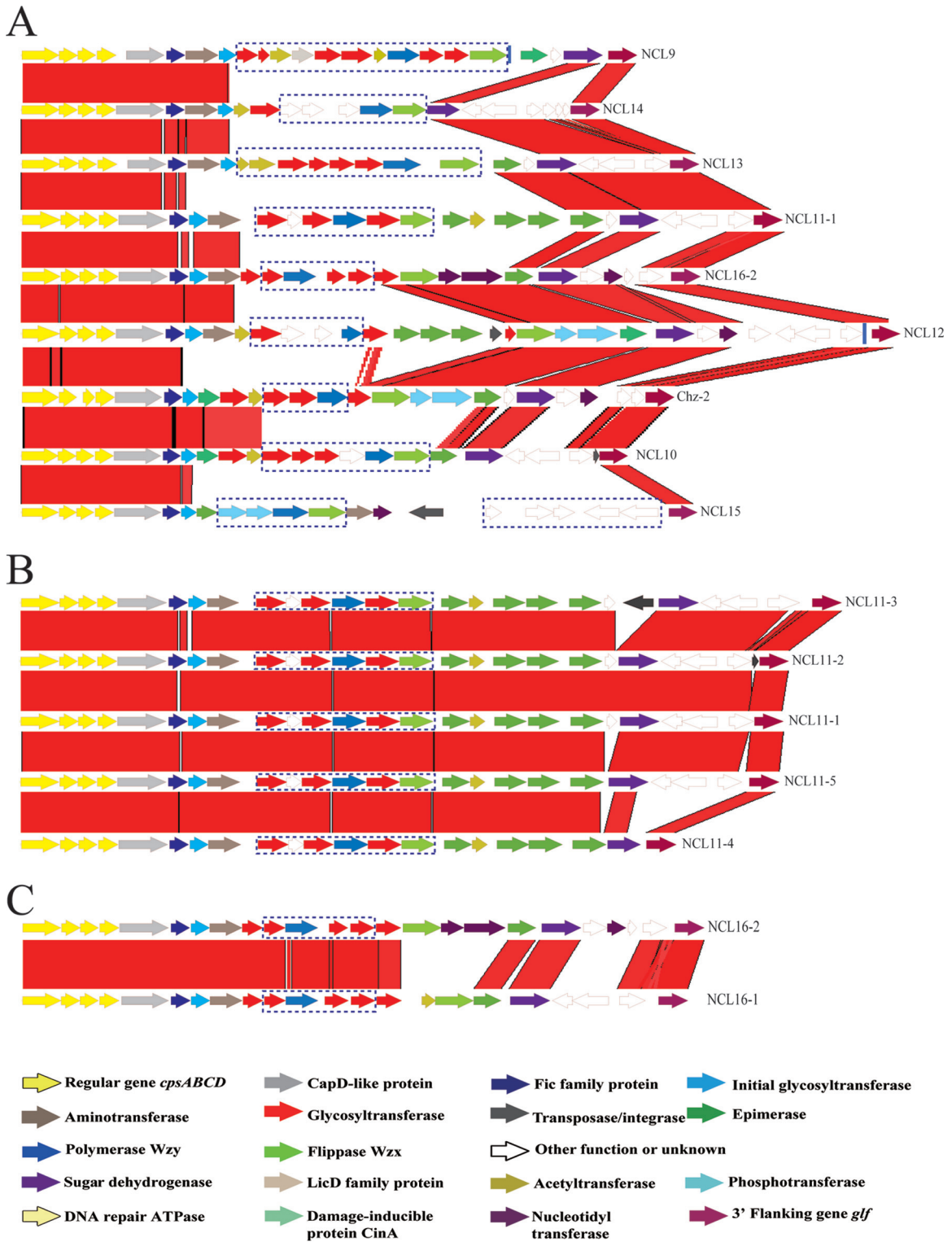


FIG 2 Comparison of the *cps* loci among 9 NCLs (A) and within NCL11 (B) and NCL16 (C). Each colored arrow represents a gene whose predicted function is shown at the bottom. NCL-specific genes are indicated by dotted blue lines. The *glf* gene is located on the 3' side of each locus.

HGs with striking heterogeneity in amino acid sequences were Wzx, Wzy, and various transferases. The antigenic identity of the capsule depends on the repeat unit of single or multiple monosaccharides, which in turn is synthesized, exported, and polymerized

by these enzymes (29). The high proportions of Wzx, Wzy, and various transferases in NCL-specific HGs indicated that their oligosaccharide structures are unique. Serotype Chz has already been shown to be a novel serotype of *S. suis* (17). Similarly to serotype

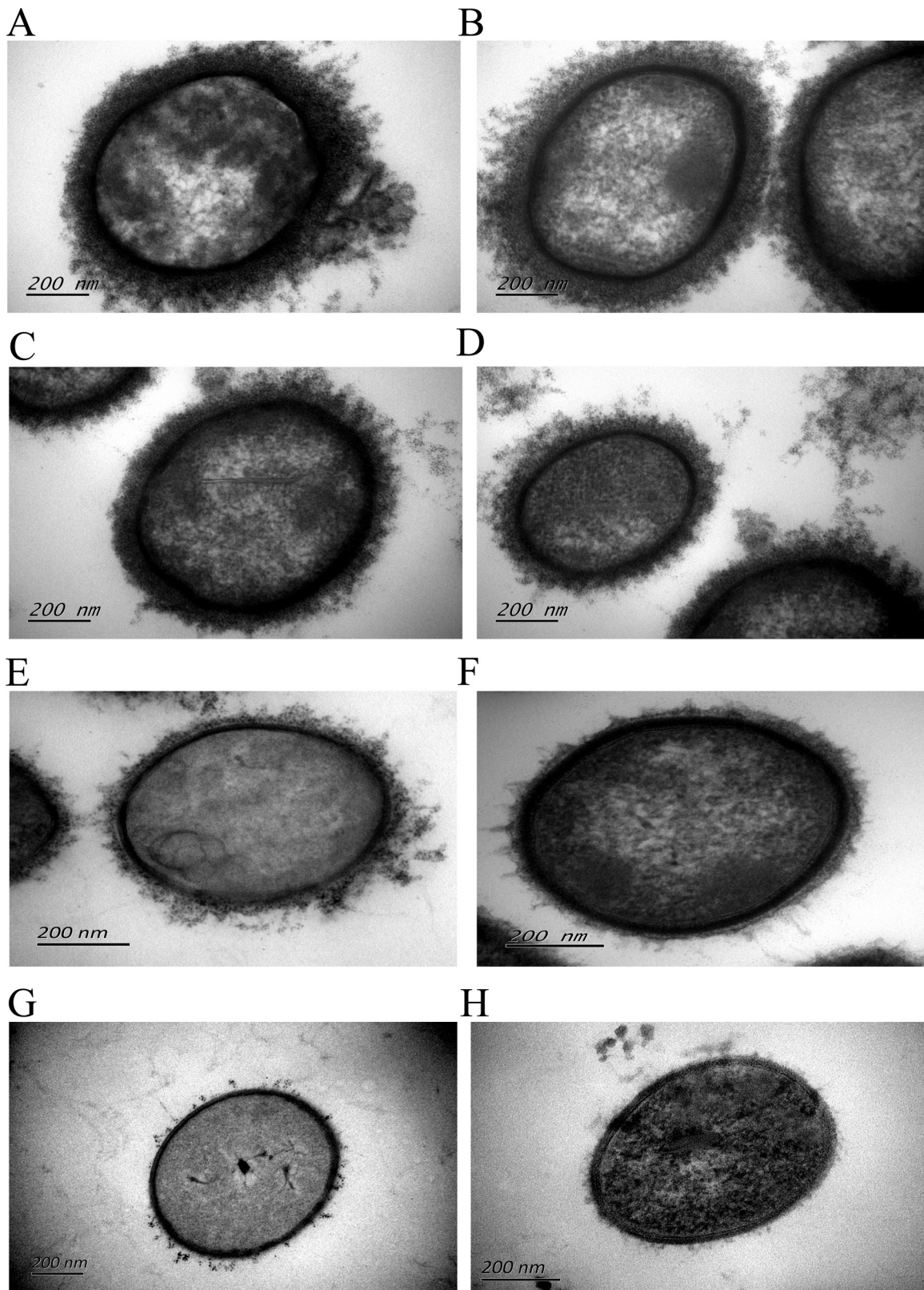


FIG 3 Variation of capsular thicknesses of isolates with different NCLs and NCL subtypes. (A) Capsulated control strain SC84 (*S. suis* serotype 2; capsular thickness, 110 to 130 nm); (B) ND96 (NCL11-3; capsular thickness, 70 to 90 nm); (C) YS108 (NCL9; capsular thickness, 60 to 80 nm); (D) YS77 (Chz-2; capsular thickness, 40 to 60 nm); (E) YS495 (NCL11-1; capsular thickness, 20 to 40 nm); (F) YS443 (NCL15; capsular thickness, 10 to 20 nm); (G) YS196 (NCL12; no capsule); (H) YS351 (Chz-2; no capsule).

Chz, we tentatively concluded that the 8 NCLs found in this study represent novel serotypes.

The capsular thicknesses varied widely between NCLs and between subtypes of an NCL. Two isolates showed an absence of a capsule in spite of no defective genes in their *cps* loci. Isolates with identical *cps* sequences can also have different capsular thicknesses. Examples are the pairs of isolates YS196 and YS205, YS576 and YS577, YS408 and YS444, and YS241 and YS601. A lack of capsule production or a varied capsular thickness may be attributed to a defect or difference in genes outside the *cps* locus that are involved in the modification or transcriptional regulation of capsular polysaccharides.

Capsule expression is essential for *S. suis* to establish asymptomatic colonization of the nasopharynx. The continuous and ongoing evolution of *cps* gene clusters also contributes to *S. suis* evasion of the host immune system. It is important to continuously survey the diversity of *S. suis* serotypes. Horizontal gene transfer within and between species may have been involved in the evolution of *cps* loci. Previous studies showed that gene replacement could easily result in the emergence of novel agglutination phenotypes (15, 30, 31). The sequence differences between Chz-2 and NCL1-3 were caused mainly by the replacement of 5 NCL-specific HGs in the center of Chz-2 by 4 NCL-specific HGs in NCL1-3. A similar transfer between NCL1-1 and NCL16-2 was also found.

The discovery and characterization of subtypes within these NCLs are also essential for understanding the mosaic structure of the capsule. Of the 17 NCLs, 8 showed two or more subtypes. The subtypes varied mainly due to the variable presence of HG55, HG293, HG294, HG312, HG313, HG314, HG315, and the transposase genes in the 3' region. This indicated that they were transferred to *cps* loci by complex recombination events between *S. suis* *cps* loci. This may have been facilitated by the transposase genes.

Compared to the above-mentioned multiple mobile HGs, NCL-specific genes were highly conserved and fixed in each NCL. The G+C contents of NCL-specific genes were obviously lower than that of whole *cps* locus. These results indicated that these NCL-specific genes were integrated into their *cps* locus from an unknown source on only one occasion.

Almost all isolates carrying NCLs were from the earliest ancestral MCG group 6 and group 7 (21), indicating that they appeared to have been acquired a long time ago and achieved wide distribution geographically. The NCLs constitute a mosaic structure originating from diverse donors. It is possible that other nasopharynx streptococci, such as *Streptococcus salivarius*, *Streptococcus mitis*, and *Streptococcus oralis*, are the leading donor candidates.

The key limitation of capsular gene typing methods based on the *wzy* genes and conventional serotyping antisera is that they are unable to distinguish the genetic variations within the same *cps* locus. In the new era of high-throughput sequencing and online bioinformatics, the unprecedented level of discrimination provided by whole-genome sequencing (WGS) will provide a novel serotyping strategy to achieve this. Recently, the SerotypeFinder tool based on WGS data was established to serotype *Escherichia coli* strains, providing a typing method that is faster and cheaper than the currently used routine procedures (32). Our 18-plex detection system will serve as a valuable tool for detecting known and additional NCLs among nonserotypeable *S. suis* isolates. Moreover, detection based on a combination of *wzy* gene variation and the presence or absence of subtype-specific HGs would be

an optimal strategy to monitor and evaluate the serotype diversity of *S. suis*.

In conclusion, our study offers an expanded view of the *cps* genetic diversity of *S. suis* with the identification of 8 unknown NCLs and 6 new subtypes of known NCLs. The clinical and public health significance of these new NCLs warrants further investigation. This study provides a high-throughput detection tool for detecting novel serotypes of *S. suis* and valuable genetic information for monitoring the diversity of these NCLs. Our data also contribute to a better understanding of capsular biosynthesis in *S. suis*.

ACKNOWLEDGMENTS

This work was supported by grant 2013ZX10004216-001-002 from the Ministry of Science and Technology, People's Republic of China; grants 81290340, 81290345, 81572044, and 81261120559 from the National Natural Science Foundation of China; and Fundamental Research Funds for the Central Universities from the Key Laboratory of Molecular Microbiology and Technology, Ministry of Education, TEDA Institute of Biological Sciences and Biotechnology, Nankai University.

REFERENCES

- Gottschalk M, Segura M, Xu J. 2007. Streptococcus suis infections in humans: the Chinese experience and the situation in North America. *Anim Health Res Rev* 8:29–45. <http://dx.doi.org/10.1017/S1466252307001247>.
- Yu H, Jing H, Chen Z, Zheng H, Zhu X, Wang H, Wang S, Liu L, Zu R, Luo L, Xiang N, Liu H, Liu X, Shu Y, Lee SS, Chuang SK, Wang Y, Xu J, Yang W, Streptococcus suis Study Groups. 2006. Human Streptococcus suis outbreak, Sichuan, China. *Emerg Infect Dis* 12:914–920. <http://dx.doi.org/10.3201/eid1206.051194>.
- Ye C, Zhu X, Jing H, Du H, Segura M, Zheng H, Kan B, Wang L, Bai X, Zhou Y, Cui Z, Zhang S, Jin D, Sun N, Luo X, Zhang J, Gong Z, Wang X, Wang L, Sun H, Li Z, Sun Q, Liu H, Dong B, Ke C, Yuan H, Wang H, Tian K, Wang Y, Gottschalk M, Xu J. 2006. Streptococcus suis sequence type 7 outbreak, Sichuan, China. *Emerg Infect Dis* 12:1203–1208. <http://dx.doi.org/10.3201/eid1208.060232>.
- Marois C, Le Devendec L, Gottschalk M, Kobisch M. 2007. Detection and molecular typing of Streptococcus suis in tonsils from live pigs in France. *Can J Vet Res* 71:14–22.
- Mwaniki CG, Robertson ID, Trott DJ, Atyeo RF, Lee BJ, Hampson DJ. 1994. Clonal analysis and virulence of Australian isolates of Streptococcus suis type 2. *Epidemiol Infect* 113:321–334. <http://dx.doi.org/10.1017/S095026880005175X>.
- Weinert LA, Chaudhuri RR, Wang J, Peters SE, Corander J, Jombart T, Baig A, Howell KJ, Vehkala M, Valimaki N, Harris D, Chieu TT, Van Vinh Chau N, Campbell J, Schultsz C, Parkhill J, Bentley SD, Langford PR, Ycroft AN, Wren BW, Farrar J, Baker S, Hoa NT, Holden MT, Tucker AW, Maskell DJ, BRaDPIT Consortium. 2015. Genomic signatures of human and animal disease in the zoonotic pathogen Streptococcus suis. *Nat Commun* 6:6740. <http://dx.doi.org/10.1038/ncomms7740>.
- Segura M, Gottschalk M, Olivier M. 2004. Encapsulated Streptococcus suis inhibits activation of signaling pathways involved in phagocytosis. *Infect Immun* 72:5322–5330. <http://dx.doi.org/10.1128/IAI.72.9.5322-5330.2004>.
- Gottschalk M, Higgins R, Jacques M, Beaudoin M, Henrichsen J. 1991. Characterization of six new capsular types (23 through 28) of Streptococcus suis. *J Clin Microbiol* 29:2590–2594.
- Gottschalk M, Higgins R, Jacques M, Mittal KR, Henrichsen J. 1989. Description of 14 new capsular types of Streptococcus suis. *J Clin Microbiol* 27:2633–2636.
- Perch B, Pedersen KB, Henrichsen J. 1983. Serology of capsulated streptococci pathogenic for pigs: six new serotypes of Streptococcus suis. *J Clin Microbiol* 17:993–996.
- Higgins R, Gottschalk M, Boudreau M, Lebrun A, Henrichsen J. 1995. Description of six new capsular types (29–34) of Streptococcus suis. *J Vet Diagn Invest* 7:405–406. <http://dx.doi.org/10.1177/104063879500700322>.
- Hill JE, Gottschalk M, Brousseau R, Harel J, Hemmingsen SM, Goh SH. 2005. Biochemical analysis, cpn60 and 16S rDNA sequence data indicate that Streptococcus suis serotypes 32 and 34, isolated from pigs, are Strep-

- tococcus orisratti. *Vet Microbiol* 107:63–69. <http://dx.doi.org/10.1016/j.vetmic.2005.01.003>.
13. Bai X, Liu Z, Ji S, Gottschalk M, Zheng H, Xu J. 2015. Simultaneous detection of 33 *Streptococcus suis* serotypes using the luminex xTAG assay. *J Microbiol Methods* 117:95–99. <http://dx.doi.org/10.1016/j.mimet.2015.07.018>.
 14. Okura M, Lachance C, Osaki M, Sekizaki T, Maruyama F, Nozawa T, Nakagawa I, Hamada S, Rossignol C, Gottschalk M, Takamatsu D. 2014. Development of a two-step multiplex PCR assay for typing of capsular polysaccharide synthesis gene clusters of *Streptococcus suis*. *J Clin Microbiol* 52:1714–1719. <http://dx.doi.org/10.1128/JCM.03411-13>.
 15. Liu Z, Zheng H, Gottschalk M, Bai X, Lan R, Ji S, Liu H, Xu J. 2013. Development of multiplex PCR assays for the identification of the 33 serotypes of *Streptococcus suis*. *PLoS One* 8:e72070. <http://dx.doi.org/10.1371/journal.pone.0072070>.
 16. Zheng H, Ji S, Liu Z, Lan R, Huang Y, Bai X, Gottschalk M, Xu J. 2015. Eight novel capsular polysaccharide synthesis gene loci identified in non-typeable *Streptococcus suis* isolates. *Appl Environ Microbiol* 81:4111–4119. <http://dx.doi.org/10.1128/AEM.00315-15>.
 17. Pan Z, Ma J, Dong W, Song W, Wang K, Lu C, Yao H. 2015. Novel variant serotype of *Streptococcus suis* isolated from piglets with meningitis. *Appl Environ Microbiol* 81:976–985. <http://dx.doi.org/10.1128/AEM.02962-14>.
 18. Ishida S, Tien LHT, Osawa R, Tohya M, Nomoto R, Kawamura Y, Takahashi T, Kikuchi N, Kikuchi K, Sekizaki T. 2014. Development of an appropriate PCR system for the reclassification of *Streptococcus suis*. *J Microbiol Methods* 107:66–70. <http://dx.doi.org/10.1016/j.mimet.2014.09.003>.
 19. Baker GC, Smith JJ, Cowan DA. 2003. Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* 55:541–555. <http://dx.doi.org/10.1016/j.mimet.2003.08.009>.
 20. Liu Z, Bai X, Ji S, Xu J, Zheng H. 2014. Development of a multiplex PCR assay to identify seven new capsular gene loci of *Streptococcus suis*. *Chin J Zoonoses* 30:337–346.
 21. Chen C, Zhang W, Zheng H, Lan R, Wang H, Du P, Bai X, Ji S, Meng Q, Jin D, Liu K, Jing H, Ye C, Gao GF, Wang L, Gottschalk M, Xu J. 2013. Minimum core genome sequence typing of bacterial pathogens: a unified approach for clinical and public health microbiology. *J Clin Microbiol* 51:2582–2591. <http://dx.doi.org/10.1128/JCM.00535-13>.
 22. Okura M, Takamatsu D, Maruyama F, Nozawa T, Nakagawa I, Osaki M, Sekizaki T, Gottschalk M, Kumagai Y, Hamada S. 2013. Genetic analysis of capsular polysaccharide synthesis gene clusters from all serotypes of *Streptococcus suis*: potential mechanisms for generation of capsular variation. *Appl Environ Microbiol* 79:2796–2806. <http://dx.doi.org/10.1128/AEM.03742-12>.
 23. Zheng H, Ji S, Lan R, Liu Z, Bai X, Zhang W, Gottschalk M, Xu J. 2014. Population analysis of *Streptococcus suis* isolates from slaughtered swine by use of minimum core genome sequence typing. *J Clin Microbiol* 52:3568–3572. <http://dx.doi.org/10.1128/JCM.00536-14>.
 24. Sanchez del Rey V, Fernandez-Garayzabal JF, Briones V, Iriso A, Dominguez L, Gottschalk M, Vela AI. 2013. Genetic analysis of *Streptococcus suis* isolates from wild rabbits. *Vet Microbiol* 165:483–486. <http://dx.doi.org/10.1016/j.vetmic.2013.04.025>.
 25. Gottschalk M, Lacouture S, Bonifait L, Roy D, Fittipaldi N, Grenier D. 2013. Characterization of *Streptococcus suis* isolates recovered between 2008 and 2011 from diseased pigs in Quebec, Canada. *Vet Microbiol* 162:819–825. <http://dx.doi.org/10.1016/j.vetmic.2012.10.028>.
 26. Wang S, Liu P, Li C, Tan Y, Cai X, Zhou D, Jiang Y. 2012. Isolation and characterization of 89K pathogenicity island-positive ST-7 strains of *Streptococcus suis* serotype 2 from healthy pigs, Northeast China. *ScientificWorldJournal* 2012:302386. <http://dx.doi.org/10.1100/2012/302386>.
 27. Han DU, Choi C, Ham HJ, Jung JH, Cho WS, Kim J, Higgins R, Chae C. 2001. Prevalence, capsular type and antimicrobial susceptibility of *Streptococcus suis* isolated from slaughter pigs in Korea. *Can J Vet Res* 65:151–155.
 28. Princivalli MS, Palmieri C, Magi G, Vignaroli C, Manzin A, Camporese A, Barocci S, Magistrali C, Facinelli B. 2009. Genetic diversity of *Streptococcus suis* clinical isolates from pigs and humans in Italy (2003–2007). *Euro Surveill* 14(33):pii=19310. <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=19310>.
 29. Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabinowitz E, Collins M, Donohoe K, Harris D, Murphy L, Quail MA, Samuel G, Skovsted IC, Kalltoft MS, Barrell B, Reeves PR, Parkhill J, Spratt BG. 2006. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet* 2:e31. <http://dx.doi.org/10.1371/journal.pgen.0020031>.
 30. Yun KW, Cho EY, Choi EH, Lee HJ. 2014. Capsular polysaccharide gene diversity of pneumococcal serotypes 6A, 6B, 6C, and 6D. *Int J Med Microbiol* 304:1109–1117. <http://dx.doi.org/10.1016/j.ijmm.2014.08.004>.
 31. Morona JK, Morona R, Paton JC. 1999. Comparative genetics of capsular polysaccharide biosynthesis in *Streptococcus pneumoniae* types belonging to serogroup 19. *J Bacteriol* 181:5355–5364.
 32. Iguchi A, Iyoda S, Kikuchi T, Ogura Y, Katsura K, Ohnishi M, Hayashi T, Thomson NR. 2015. A complete view of the genetic diversity of the *Escherichia coli* O-antigen biosynthesis gene cluster. *DNA Res* 22:101–107. <http://dx.doi.org/10.1093/dnares/dsu043>.