AEM
Journals.ASM.org

Genetic Analysis of Capsular Polysaccharide Synthesis Gene Clusters from All Serotypes of *Streptococcus suis*: Potential Mechanisms for Generation of Capsular Variation

Masatoshi Okura,[a] Daisuke Takamatsu,[a,b] Fumito Maruyama,[c,d] Takashi Nozawa,[c] Ichiro Nakagawa,[c] Makoto Osaki,[a]* Tsutomu Sekizaki,[e] Marcelo Gottschalk,[f] Yumi Kumagai,[g,h]* Shigeyuki Hamada[g,h]

Bacterial and Parasitic Disease Research Division, National Institute of Animal Health, National Agriculture and Food Research Organization, Tsukuba, Japan[a]; The United Graduate School of Veterinary Sciences, Gifu University, Gifu, Japan[b]; Section of Bacterial Pathogenesis, Tokyo Medical and Dental University, Graduate School of Medical and Dental Sciences, Tokyo, Japan[c]; Section of Microbial Genomics and Ecology, Tokyo Medical and Dental University, Graduate School of Medical and Dental Sciences, Tokyo, Japan[d]; Research Center for Food Safety, Graduate School of Agricultural and Life Sciences, The University of Tokyo, Tokyo, Japan[e]; Groupe de Recherche sur les Maladies Infectieuses du Porc and Centre de Recherche en Infectiologie Porcine, Faculté de Médecine Vétérinaire, Université de Montréal, St-Hyacinthe, Quebec, Canada[f]; Research Institute for Microbial Diseases, Osaka University, Osaka, Japan[g]; Thailand-Japan Research Collaboration Center on Emerging and Re-emerging Infections (RCCERI), Nonthaburi, Thailand[h]

***Streptococcus suis* strains are classified into 35 serotypes on the basis of the antigenicity of their capsular polysaccharides (CPs). CP synthesis genes are known to be clustered on the chromosome (*cps* gene cluster). The entire *cps* gene clusters of *S. suis* have so far been sequenced in 15 serotypes and found to be located between *orfZ* and *aroA*. In this study, to provide comprehensive information about *S. suis* CPs, we sequenced the entire *cps* gene clusters of the remaining serotypes and analyzed the complete set of *S. suis cps* gene clusters. Among the 35 *cps* gene clusters, 22 were located between *orfZ* and *aroA*, whereas the other 13 were flanked by other gene(s) on the chromosomes, and the chromosomal locus was classified into five patterns. By clustering analysis, the predicted products of *cps* genes found in the 35 serotypes were assigned into 291 homology groups, and all serotypes possessed a serotype-specific gene, except for serotypes 1, 2, 1/2, and 14. Because of the presence of genes encoding flippase (*wzx*) and polymerase (*wzy*), CPs of all serotypes were thought to be synthesized by the Wzx/Wzy pathway. Our data also implied the possibility of the transfer of the entire or partial *cps* gene clusters among *S. suis* strains, as well as the influence of spontaneous mutations in a single gene or a few genes on the antigenicity of some serotypes. Accumulation of these gene transfers and small-scale mutations may have generated the antigenic diversity of *S. suis* CPs.**

Streptococcus suis, a Gram-positive coccus, can cause a wide range of diseases, including meningitis, septicemia, and endocarditis in pigs, and is recognized as an important pathogen responsible for severe economic losses to the swine industry worldwide (1, 2). *S. suis* can also affect humans in close contact with infected pigs or pork and is increasingly recognized as an emerging zoonotic agent in Asia (3–5). Most *S. suis* strains possess a capsular polysaccharide (CP) and, on the basis of the antigenic differences of the CPs, the encapsulated strains have so far been divided into 35 serotypes (serotypes 1 to 34 and serotype 1/2, which reacts with both serotype 1 and 2 typing sera) (6–9). Although the CP of serotype 2 strains has been shown to be an important antiphagocytic factor of *S. suis* (10, 11), the current understanding of *S. suis* CP, including its role, diversity, and evolution, remains limited. For elucidation of such issues, further studies of *S. suis* CP, including the genes responsible for synthesis, are required.

Bacterial CPs are usually linked to their cell surface via covalent attachments and are composed of repeating units of a single or multiple monosaccharides joined by glycosidic linkages (12). In Gram-positive bacteria, CPs of *Streptococcus pneumoniae* have been well studied and are known to be generally synthesized by the Wzx/Wzy-dependent pathway (13, 14). In this pathway, first, an initial monosaccharide is linked as a sugar phosphate to a membrane-associated lipid carrier by an initial sugar transferase. Second, further monosaccharides are added sequentially by specific glycosyltransferases to produce repeat units. Then, Wzx flippase transports the repeat units to the outer surface of the cytoplasmic membrane, and each repeat unit is polymerized to form the lipid-linked CP by Wzy polymerase. Finally, mature CP is translocated to the peptidoglycan by the membrane protein complex (Wzd/Wze complex in *S. pneumoniae*). The genes involved in this pathway comprise a gene cluster (*cps* gene cluster) and are usually located at the same chromosomal locus (13). The *cps* gene cluster includes the genes encoding the initial sugar transferase, additional glycosyltransferases, Wzy polymerase (*wzy*), Wzx flippase (*wzx*), and enzymes to modify the repeat units or to add other moieties on CP (12, 13).

In *S. suis*, the entire *cps* gene cluster was identified in a serotype 2 strain (15, 16), and 14 additional entire *cps* gene clusters have so far been sequenced in serotype 1, 1/2, 3, 4, 5, 7, 8, 9, 10, 14, 16, 19, 23, and 25 reference strains (17, 18). In serotypes 6, 11, 12, 13, 15, 17, 18, 20, 21, 22, 24, 27, 28, 29, 30, and 31, partial sequences of the

Received 4 December 2012 **Accepted** 7 February 2013
Published ahead of print 15 February 2013
Address correspondence to Daisuke Takamatsu, p1013dt@affrc.go.jp, or Fumito Maruyama, fumito-m.bac@tmd.ac.jp.
* Present address: Makoto Osaki, Agriculture, Forestry and Fisheries Research Council Secretariat, Ministry of Agriculture, Forestry and Fisheries, Tokyo, Japan; Yumi Kumagai, Department of Entomology, University of Minnesota, Minneapolis, Minnesota, USA.
Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.03742-12.

doi:10.1128/AEM.03742-12

TABLE 1 Properties of the *cps* gene clusters in 35 serotypes of *S. suis*

Strain	Serotype	No. of *cps* genes[a]	HG of[b]: IT	Wzy	Wzx	Location pattern[c]	Accession no.
NCTC 10237	1	26	HG6	HG50	HG13	I-a	AB737817
2651	1/2	23	HG6	HG54	HG13	I-a	AB737816
P1/7	2	22	HG6	HG54	HG13	I-a	BR001000[d]
4961	3	16	HG42	HG90	HG89	I-a	BR001001[d,e]
6407	4	17	HG21	HG96	HG98	I-a	BR001002[d]
11538	5	20	HG21	HG103	HG105	I-a	BR001003[d,e]
2524	6	21	HG6	HG109	HG13, HG58	I-a	AB737818
8074	7	18	HG21	HG113	HG114	I-a	BR001004[d]
14636	8	16	HG6	HG120	HG43	I-a	BR001005[d]
22083	9	19	HG8	HG123	HG43	I-b	BR001006[d,e]
4417	10	22	HG8	HG130	HG132	I-a	BR001007[d]
12814	11	18	HG8	HG138	HG71	I-a	AB737819
8830	12	18	HG8	HG140	HG141	I-a	AB737820
10581	13	18	HG42	HG148	HG149	I-b	AB737821
13730	14	22	HG6	HG50	HG13	I-a	AB737822
NCTC 10446	15	13	HG6	HG153	HG155	I-a	AB737823
2726	16	22	HG6	HG157	HG13, HG58	I-a	BR001008[d]
93A	17	26	HG21	HG164	HG163	I-a	AB737824
NT77	18	22	HG42	HG170	HG171	I-a	AB737825
42A	19	22	HG21	HG174	HG176	I-a	BR001009[d]
86–5192	20	18	HG6	HG182	HG186	III	AB737826
14A	21	19	HG8	HG194	HG193	I-b	AB737827
88–1861	22	26	HG6	HG200	HG203	III	AB737828
89–2479	23	17	HG21	HG213	HG216	I-a	BR001010[d]
88-5299A	24	17	HG8	HG220	HG221	I-b	AB737829
89-3576-3	25	15	HG6	HG229	HG231	I-a	BR001011[d]
89-4109-1	26	22	HG8	HG240	HG238	III	AB737830
89–5259	27	20	HG6	HG246	HG13	II	AB737831
89–590	28	17	HG8	HG254	HG256	I-a	AB737832
92–1191	29	18	HG8	HG259	HG262	I-b	AB737833
92–1400	30	17	HG8	HG264	HG265	I-a	AB737834
92–4172	31	15	HG269	HG274	HG275	I-b	AB737835
EA1172.91	32	16	HG278	HG281	HG285	IV	AB737836
EA1832.92	33	15	HG8	HG286	HG43	I-b	AB737837
92–2742	34	19	HG287	HG290	HG71	IV	AB737838

[a] Excluding putative transposase and integrase genes or their remnants.
[b] HG, homology group; IT, initial sugar transferase; Wzy, repeat unit polymerase; Wzx, flippase.
[c] Pattern of chromosomal locus of the *cps* gene cluster.
[d] Reannotation performed in this study.
[e] The sequences determined in this study were merged into the previous study (17) and used for reannotation in this study (see Table S2 in the supplemental material).

cps gene clusters (*orfY-cpsD* genes) have been determined and deposited in the GenBank database (accession numbers JF791152 to JF791167). Furthermore, complete genome sequences of 13 *S. suis* nonreference strains (seven of serotype 2 and one of serotypes 1, 1/2, 3, 7, 9, and 14) have been determined to date (19–24). A previous study analyzing 15 *S. suis cps* gene clusters indicated that all the *cps* gene clusters are located between the *orfZ* (conserved hypothetical protein gene) and *aroA* (3-phosphoshikimate 1-carboxyvinyltransferase gene) genes on the chromosome. In addition, because of the possession of putative *wzx* and *wzy* genes, *S. suis* CPs were suggested to be synthesized by the Wzx/Wzy-dependent pathway (17). Although the previous studies provided some insights into the diversity and evolution of *S. suis* CP, additional information is necessary for comprehensive understanding of this important virulence factor. In this study, to investigate the biosynthetic mechanism of *S. suis* CP, identify conserved and serotype-specific *cps* genes, and suggest possible mechanisms by which the diversity arises, we sequenced the 23 *cps* gene clusters, including those of the remaining 20 serotypes, and analyzed the complete set of the *cps* gene clusters in *S. suis*.

MATERIALS AND METHODS

Bacterial strains and genomic DNA isolation. The 35 *S. suis* strains used in this study are listed in Table 1. Serotype 1, 3 to 34, and 1/2 strains were reference strains for each serotype (6–9). Strain P1/7 (20) was used as the serotype 2 strain. Genomic DNAs were isolated as described previously (25).

Sequencing of *cps* gene clusters. The *cps* gene clusters of 23 serotypes (1, 1/2, 6, 11 to 15, 17, 18, 20 to 22, 24, and 26 to 34) were sequenced in this study.

(i) Serotypes 1, 1/2, 6, 11, 14, 15, 17, and 18. The entire *cps* gene clusters of serotype 1/2, 11, 14, 15, and 18 reference strains were amplified with primers cpsA-F and aroA-R (see Table S1 in the supplemental material) and sequenced by primer walking. Sequences of the entire *cps* gene clusters of serotypes 1, 6, and 17 were determined by amplifying two or three portions of the *cps* gene clusters with the primer pairs listed in Table S1, sequencing the PCR products, and assembling the data. Primers were

designed on the basis of the reported sequences of *S. suis cps* gene clusters (see Table S1), and PCR was carried out using LA-*Taq* or *Ex Taq* polymerase (TaKaRa, Ohtsu, Japan). Sequencing was performed with a BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems, Foster City, CA) using a 3130xl genetic analyzer (Applied Biosystems). The sequence assembly was carried out using SEQUENCHER 4.8 (Hitachi Software, Tokyo, Japan) on the basis of a minimum overlap of 50 bp with a 95% minimal mismatch percentage, and the determined sequences were confirmed by sequencing both strands of multiple PCR products amplified independently.

(ii) Serotype 12, 13, 20 to 22, 24, and 26 to 34. The *cps* gene clusters of the 15 serotypes were identified from draft genome sequences of the respective serotype reference strains. The genomes were sequenced using an Illumina genome analyzer (GA) IIx. For each sample, the reads were 101 bp in length from a paired insert of approximately 400 bp mean separation. Short reads were filtered and trimmed with the qtrim program (Genaris, Inc., Yokohama, Japan) (unpublished method) to maximize the differences in average quality between trimmed sequences and trimmings. The trimmed short reads that were longer than 75 bp and had an average quality of more than 28 were assembled with Velvet (26) at Genaris, Inc. The command line version of the National Center for Biotechnology Information Basic Local Alignment Search Tool (NCBI BLAST) (27) was used to search contigs that contain sequences similar to those of the *cps2* gene cluster and its flanking regions of strain P1/7 (SSU0512 to SSU0563; GenBank accession number AM946016). Gaps were closed through directed PCR and primer-walking approaches.

(iii) *cps3E*, *cps5E*, and *cps9J* and the 3′-side boundary region of *cps9* locus. To verify the reported sequences, the genes designated *cps3E*, *cps5EF* (renamed as *cps5E* in this study), and *cps9J* (17) and the 3′-side boundary region of *cps9* locus (see below for the definition of 3′ side) were amplified by primer pairs cpsD-F/cps3F-R, cpsD-F/cps5F-R, cps9I-F/cps9K-R, and cps9N-F/cps9Rj-R (see Table S1 in the supplemental material), respectively, and sequenced by primer walking.

The genetic organization of each *cps* gene cluster obtained from the assembled data was confirmed by PCRs with specific primers designed on the basis of the resulting sequence data (data not shown).

Bioinformatic methods and cluster analysis. The already-reported sequences of *cps* gene clusters of serotypes 2, 3, 4, 5, 7, 8, 9, 10, 16, 19, 23, and 25 (GenBank accession numbers AM946016 [SSU0512 to SSU0557], JF273646 to JF273652, HQ694980, JF273654 to JF273656, respectively) (17, 18, 20) were also analyzed together with those determined in this study to compare the results obtained by the same method. Because some discrepancies were found between our already-determined sequences and the reported sequences of serotype 1, 1/2, and 14 reference strains (17) (see Table S2 in the supplemental material), we used our sequences for the analysis. Gene prediction and annotation were performed using the RAST and MiGAP servers (28, 29). Insertion sequences (ISs) were identified using ISfinder (30). Except for the genes and remnants associated with putative mobile elements, all predicted genes in the 35 *cps* gene clusters were designated *cps* genes in alphabetical order (*cpsA*-*Z*) according to the nomenclature previously used in *S. suis* (11, 15–18). Of note, according to the resequenced and reannotation results obtained in this study, we renamed some *cps* genes, and thus the names of *cps* genes are not necessarily identical to those in previous studies (17, 18, 20). We defined the 5′ and 3′ sides of the *cps* gene clusters on the basis of the direction of transcription of the four conserved *cps* genes (*cpsA*-*cpsD*; as described below). Predicted proteins of the *cps* genes were clustered into homology groups (HGs) using the gene family method implemented in PGAP-1.01 (31). For each gene pair of the same cluster, the global match region was no less than 50% of the longer gene protein sequence and the identity was also no less than 50% by MCL algorithm (32). The minimum bit score value and E value in BLASTP were 50 and 1e−8, respectively.

The presence or absence of genes encoding proteins of every HG was coded as binary data, with presence as 1 and absence as 0 in the respective *cps* gene cluster. Hierarchical clustering of the binary data was conducted using Cluster 3.0 (33) with the average linkage method and Euclidean distance matrix. Trees were represented by Java TreeView (34). The nucleic acid or translated protein sequences were analyzed by the command line version of NCBI BLAST or the BLAST network service (http://blast.ncbi.nlm.nih.gov/) (35), and the Artemis Comparison Tool (ACT) (36) was used to visualize the analyzed data (bit scores above 50 and E values lower than 1e−8).

Phylogenetic analysis. Multiple DNA alignments were obtained by using the CLUSTALW program (37). Phylogeny calculations and construction of neighbor joining trees were performed by using MEGA software version 5 (38) on the basis of the Kimura 2 parameter model. The confidence of nodes within trees was assessed with 1,000 bootstrap replicates.

Nucleotide sequence accession numbers. DNA sequences or reannotation data obtained in this study were deposited in the DNA Data Bank of Japan (DDBJ) under the accession numbers AB737816 to AB737838, AB738319 to AB738322, and BR001000 to BR001011 (Table 1; see also Table S2 in the supplemental material).

RESULTS AND DISCUSSION

Chromosomal loci of 35 *cps* gene clusters. The draft genome sequence of 15 serotype reference strains determined in this study (overview of these sequencing results is shown in Table S3 in the supplemental material) revealed that the *cps* gene clusters of *S. suis* are not necessarily located at the same chromosomal locus. By comparing the sequences of the 35 *cps* gene clusters with the complete genome sequence of *S. suis* P1/7 (accession number AM946016), the chromosomal loci of the *cps* gene clusters were classified into five patterns (I-a, I-b, II, III, and IV) (Table 1 and Fig. 1A).

In 22 serotypes, the *cps* gene clusters are located between the *orfZ*-*orfX* region (*orfX*; conserved hypothetical protein gene) and the *aroA* gene (pattern I-a) and, except for serotype 9, the results agree with those in the previous study (17). Although the *cps* gene cluster of the serotype 9 reference strain was previously reported to be located between *orfZ* and *aroA* (17), the gene clusters of serotype 9, 13, 21, 24, 29, 31, and 33 reference strains are flanked by the *orfZ*-*orfX* region and the *glf* gene (UDP-galactopyranose mutase gene; corresponding to SSU0563 in P1/7) in our data (pattern I-b). In the seven serotypes, the *aroA* and downstream genes (corresponding to SSU0557 to SSU0562 in P1/7) are located upstream of the *aroF1* gene (corresponding to SSU1086 in P1/7; serotypes 9, 13, 21, 24, 29, and 31) or downstream of the gene corresponding to SSU1090 in P1/7 (serotype 33) (Fig. 1B).

In serotype 27, the *cps* gene cluster is located between genes corresponding to SSU1265 and SSU1264 of P1/7 (pattern II), whereas the *cps* gene clusters of serotypes 20, 22, and 26 are located between genes corresponding to SSU1210 and SSU1209 of P1/7 (pattern III). Interestingly, SSU1265 and SSU1264 and SSU1210 and SSU1209 are very closely and directly linked to each other, respectively, on the chromosome of P1/7 (Fig. 1B; see also Fig. S1 in the supplemental material). In addition, *orfZ*-*orfX* and *glf*, lying in the 5′- and 3′-side flanking regions of the *cps* gene clusters, respectively, in the pattern I-a and I-b strains, are closely located on the chromosomes in the pattern II and III strains (Fig. 1B). On the other hand, the predicted *cps* gene clusters of serotypes 32 and 34 are flanked by genes corresponding to SSU0869 and SSU1828 of P1/7 (pattern IV); however, the two genes are located at different chromosomal loci in P1/7.

Although *S. suis* reference strains were identified as *S. suis* on the basis of their physiological and biochemical characteristics

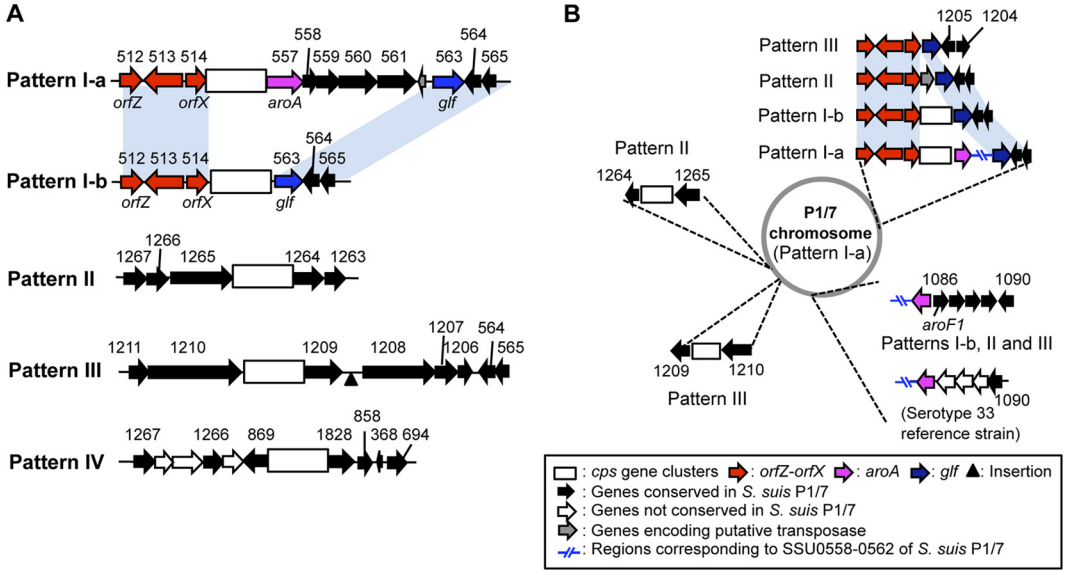

FIG 1 Five patterns of chromosomal loci of *S. suis cps* gene clusters. (A) Schematic representations of the *cps* gene clusters and their flanking genes of each chromosomal locus. In pattern III, approximately 3-kbp and 3.8-kbp sequences are inserted in the genomes of the serotype 22 and 26 reference strains, respectively, at the position indicated by the arrowhead (see Fig. S1 in the supplemental material). (B) Corresponding chromosomal loci on P1/7 genome (pattern I-a) of the *cps* gene clusters of pattern I-a, I-b, II, and III. Corresponding loci of both flanking genes of the gene cluster in pattern I-a are also indicated even if they are separate from the *cps* gene cluster. Pattern IV does not appear in this figure, as the *cps* locus of this pattern had very little relation to those of other patterns (see Fig. S2 in the supplemental material). The locus tag numbers and assigned gene names of *S. suis* P1/7 are appended to the corresponding genes. Regions conserved among strains of different chromosomal loci are indicated as light blue blocks.

(6–9), reference strains of serotypes 20, 22, 26, 32, 33, and 34 have been shown to be more distantly related to the others on the basis of the 16S rRNA gene and *cpn60* sequencing (39, 40). In addition, serotype 32 and 34 reference strains were indicated to be *Streptococcus orisratti* by Hill et al. (40), and a recent work by Tien et al. (41) showed that reference strains of serotypes 20, 22, 26, and 33 should be also removed from taxon of *S. suis*. Intriguingly, in our data, only serotype 20, 22, and 26 reference strains had the *cps* gene clusters of pattern III, and only serotype 32 and 34 reference strains had the clusters of pattern IV. In addition, the genetic organization of the 5′-side flanking regions of the *cps32* and *cps34* gene clusters is relatively similar to that of several other *Streptococcus* species (see Fig. S2 in the supplemental material). In the serotype 33 reference strain, although the chromosomal locus of the *cps* gene cluster was grouped into pattern I-b, genetic organization of the *aroA* region is different from those of the other strains with *cps* gene clusters of pattern I-b (Fig. 1B). The phylogenetic tree on the basis of the nucleotide sequence alignment of the *cpsA-cpsD* regions, which are conserved in all serotypes (as described below), indicates that all of the strains with *cps* gene clusters of pattern III (serotypes 20, 22, and 26) and pattern IV (serotypes 32 and 34) are apparently distant from other serotype reference strains (see Fig. S3 in the supplemental material). Furthermore, the *cpsA-cpsD* region of the serotype 33 reference strain (pattern I-b) has only approximately 70% sequence identities with those of other pattern I-a or I-b strains and is not clustered into any groups including other strains with *cps* locus of pattern I-a or I-b in this tree (see Fig. S3). Because chromosomal loci of the *cps* gene clusters were defined on the basis of the draft genome sequence data, further analysis including the complete genome sequencing of the reference strains will be needed to confirm the precise locus of each *cps* gene cluster. However, different chromosomal loci of the *cps* gene clusters found in this study and the phylogenetic analysis of the *cpsA-cpsD* regions further support the phylogenetically distant relation of these serotype reference strains in the *S. suis* population, as suggested previously (39–41).

General features of the 35 *cps* gene clusters. The length of the *cps* gene clusters ranges from 15,274 bp to 40,198 bp (see Table S2 in the supplemental material), and 672 predicted coding sequences in the 35 *cps* gene clusters were designated *cps* genes (see Table S4 in the supplemental material). Most of the *cps* genes are oriented in the same direction (Fig. 2). The genes involved in the regulation and processing of CP, *cpsA*, *cpsB*, *cpsC*, and *cpsD* (designated *wzg*, *wzd*, *wze*, and *wzh*, respectively, in the bacterial polysaccharide gene database [BPGD]) (42), are present in all serotypes and located on the 5′-side of the clusters, although the gene order is *wzg-wzh-wzd-wze* in the *cps32* and *cps34* gene clusters. All *cps* gene clusters also contain genes encoding putative flippase (*wzx*) and polymerase (*wzy*), as well as various sets of glycosyltransferase genes, including an initial sugar transferase gene (Fig. 2; see also Table S4 in the supplemental material), strongly suggesting that CPs of all *S. suis* serotypes are synthesized by the Wzx/Wzy pathway. Some of the *cps* genes were predicted to encode modifying enzymes (such as acetyltransferase, aminotransferase, phosphotransferase, pyruvyl transferase, and nucleotidyl transferase), nucleotide sugar phosphate biosynthesis enzymes, and other enzymes, which are involved in the biosynthesis and addition of other components on CP (such as glycerol and choline) (Fig. 2; see also Table S4). The percentage G+C content of all the *cps* gene clusters (32.5 to 36.7%) is lower than those of several reported *S. suis* genomes (41.0 to 41.4%) (see Table S2), and most of the *cps* gene clusters have more than one intact or disrupted gene encoding transposase and/or integrase family proteins in the 3′-side regions (Fig. 2). Twenty-four *cps* genes (*cps1G-1H*, *cps1X-*

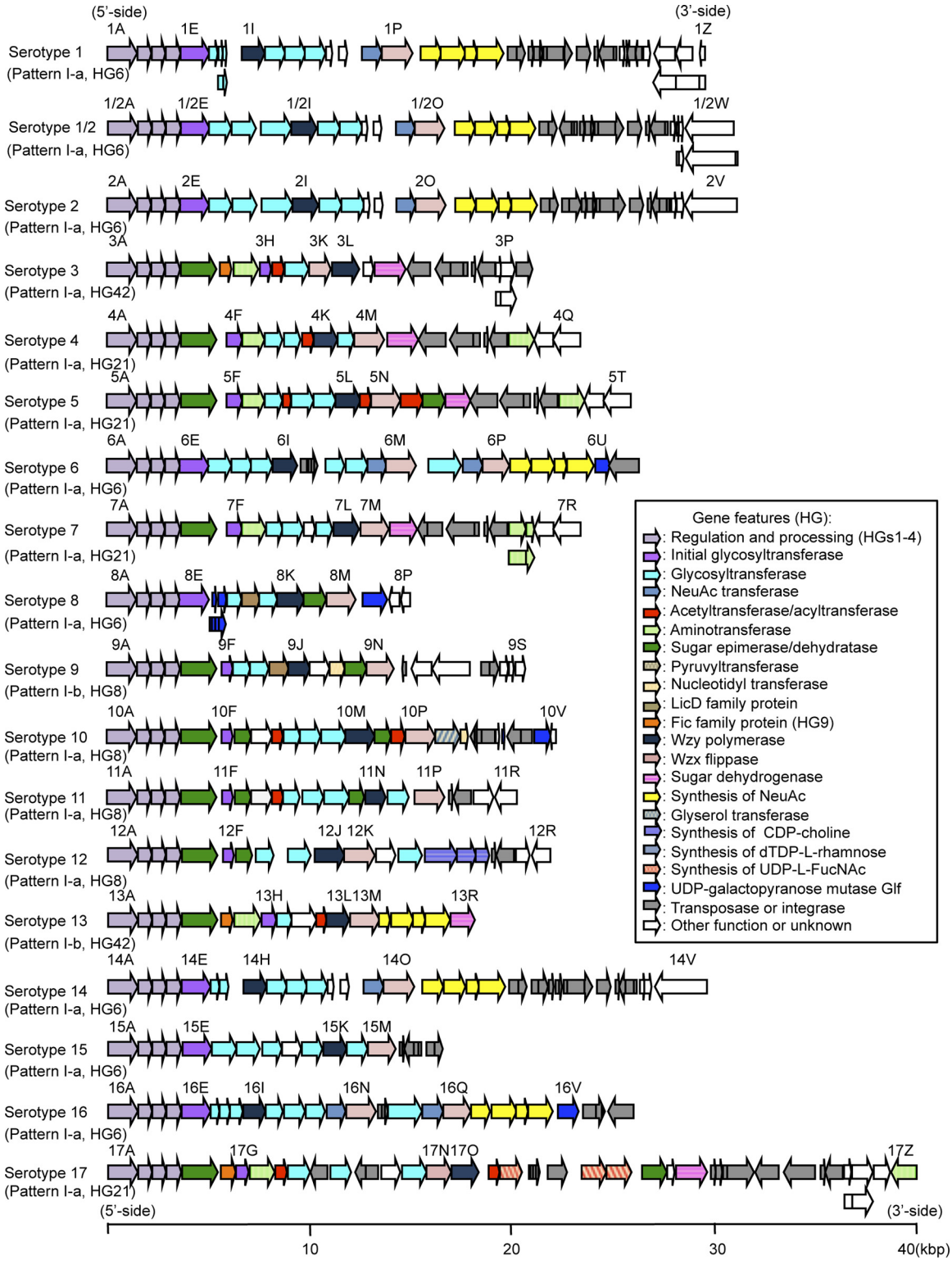

FIG 2 Schematic representations of *cps* gene clusters in *S. suis* 35 serotypes. The pattern of chromosomal locus and homology group (HG) of initial transferase in each *cps* gene cluster are shown in parentheses. Each colored arrow represents the gene whose predicted function is shown in the right panel. Gene names ("*cps*" omitted) of the first and last *cps* genes, initial transferase genes, and *wzx* and *wzy* genes in each *cps* gene cluster are appended to the corresponding arrows. Of note, the names of *cps* genes in this study are not necessarily identical to those in a previous study (17). Scale bar indicates the size of 40 kbp in units of 10-kbp increments.

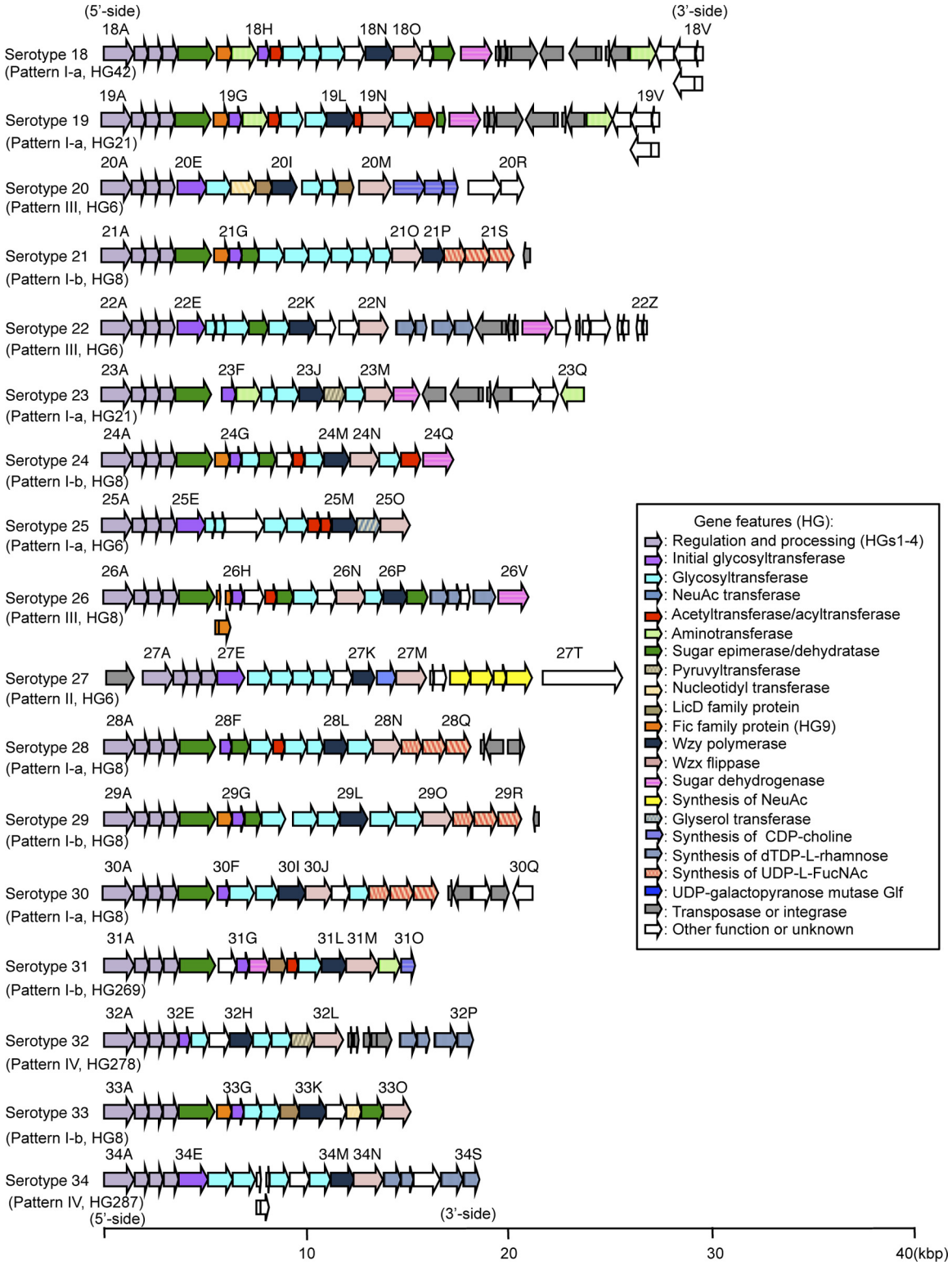

FIG 2 continued

1Z, *cps1/2U-1/2V*, *cps1/2W*, *cps3O-3P*, *cps7O-7P*, *cps8F-8G*, *cps17W-17X*, *cps18U-18V*, *cps19U-19V*, *cps26F-26G*, and *cps34H-34I*) were presumed to be affected by nonsense or frameshift mutation(s), and the results were verified by resequencing.

In spite of using the same strains, several discrepancies were found between our sequence data and those determined previously by Wang et al. (accession numbers JF273644 to JF273656 and JF791152 to JF791167) (see Table S2 in the supplemental

material). For example, in 11 reference strains (serotypes 11, 12, 15, 20, 21, 22, 27, 28, 29, 30, and 31), there were more than 100 nucleotide sequence differences between our data and the previously sequenced data, although some have been registered as "unverified sequence" in the GenBank database. In the previous data, the *orfZ-orfX* genes were located on the 5′ flanking region of the *cps20*, *cps22*, and *cps27* gene clusters. However, in our data, the three genes were located far from the *cps* gene clusters (pattern II or III; their locations were also verified by PCR in our strains) (data not shown). Moreover, our sequence data of *cps3E*, *cps5E*, and *cps9J* had three base deletions, one base substitution, and one base deletion, respectively, compared to the previous study (17), and although these three genes are likely to be pseudogenes according to the previous data (accession numbers JF273646, JF273648, and JF273651), they were indicated to be intact in this study. On the other hand, in our data, *cps1G* has one base deletion compared to the previous data and is divided into two open reading frames (*cps1G-1H* in our data) (as described in detail below). Furthermore, as described above, the 3′-side flanking gene of the *cps* gene cluster of the serotype 9 reference strain is not *aroA* but *glf* in our data. Some of these discrepancies might have occurred during repeated passages of the strains; however, at present, it is unknown why such large numbers of discrepancies were observed between the two studies.

Assignment of HGs. To make a more specific assignment of the *cps* genes, we performed the MCL program to assemble their products into homology groups (HGs). Sixty-nine percent of the products assembled into 82 HGs containing 2 to 35 members, and the remainder formed 209 single-member HGs (see Table S5 in the supplemental material). By hierarchical clustering analysis on the basis of the presence or absence of genes encoding proteins of every HG, a dendrogram of the 35 *cps* gene clusters was constructed (see Fig. S4 in the supplemental material). Of note, this does not accurately present the phylogeny of the *cps* gene clusters and its robustness is not assessed.

(i) 5′-side regions of *cps* gene clusters. Proteins of HG1 to HG4 cover every serotype and are encoded by the first four *cps* genes (*cpsA-cpsD*). In all serotypes, an initial sugar transferase gene is also located in the 5′-side regions, and the products were classified into seven HGs (HG6, HG8, HG21, HG42, HG269, HG278, and HG287) (Table 1; see also Fig. S4 in the supplemental material).

The initial transferases of HG6 (serotypes 1, 1/2, 2, 6, 8, 14, 15, 16, 20, 22, 25, and 27), HG278 (serotype 32), and HG287 (serotype 34) are encoded by the fifth *cps* gene (*cpsE*), whereas those of HG8 (serotypes 9, 10, 11, 12, 21, 24, 26, 28, 29, 30, and 33), HG21 (serotypes 4, 5, 7, 17, 19, and 23), HG42 (serotypes 3, 13, and 18), and HG269 (serotype 31) are encoded by the sixth (*cpsF*), seventh (*cpsG*), or eighth (*cpsH*) *cps* genes in the clusters (see Tables S4 and S5 in the supplemental material). In the *cps* gene clusters encoding the initial transferases of HG8, HG21, HG42, and HG269, the nucleoside-diphosphate sugar epimerase of HG5 is always encoded between *cpsD* and the initial transferase genes. In addition, genes encoding Fic family proteins (HG9, *cpsF* in serotypes 3, 13, 17, 18, 19, 21, 24, 29, and 33; HG232 and HG233, *cpsF-G* [probable pseudogene] in serotype 26), HpcH/HpaI aldolase/citrate lyase family protein of HG268 (*cpsF* in serotype 31), and/or aminotransferases of HG41 (*cpsG* in serotypes 3, 13, and 18) are also present between *cpsD* and the initial transferase genes in several serotypes (see Fig. S4 and Table S4 in the supplemental material).

(ii) Biosynthesis genes of CP components conserved among several serotypes. In addition to five serotypes (serotypes 1, 1/2, 2, 14, and 16), which were reported to have four sialic acid synthesis genes in the *cps* gene clusters (17), the sialic acid synthesis genes were also found in the clusters of three serotypes (serotypes 6, 13, and 27) (Fig. 2; see also Fig. S4 and Tables S4 and S5 in the supplemental material). Except for the *cps13* gene cluster, the four products were classified into HG10, HG14, HG15, and HG16. In the *cps13* gene cluster, one of the products was assigned to HG10, but the other three were classified into different HGs (HG150 to HG152) (see Table S4 in the supplemental material). In addition, the order of the four genes was different between serotype 13 and the other serotypes; therefore, the sialic acid synthesis genes of serotype 13 seem to be phylogenetically different from those of the other seven serotypes.

The proteins predicted to be involved in *N*-acetyl fucosamine synthesis, CDP-choline synthesis, and dTDP-L-rhamnose synthesis were also found to be encoded in the newly sequenced *cps* gene clusters (Fig. 2; see also Fig. S4 and Tables S4 and S5 in the supplemental material). The *N*-acetyl fucosamine synthesis-related proteins (HG25 to HG27) and CDP-choline synthesis-related proteins (HG74 and HG76) are encoded in the gene clusters of serotypes 17, 21, 28, 29, and 30 and serotypes 12 and 20, respectively. The *cps* gene clusters of serotypes 22, 26, 32, and 34 encode proteins involved in dTDP-L-rhamnose synthesis (HG36 to HG38 and HG49), although the gene encoding dTDP-4-dehydrorhamnose reductase (HG49) is absent from the *cps26* gene cluster. Although rhamnose has been shown to be a component of serotype 2 CP (43, 44), these genes are absent in the *cps* gene cluster. However, dTDP-L-rhamnose synthesis genes are present on the chromosome but located apart from the *cps* loci in serotype 2 (SSU1129, SSU1130, SSU1132, and SSU1133 in P1/7; identity ≥79% and coverage ≥99% in blastp compared to the *cps22*, *cps26*, *cps32*, and/or *cps34* gene clusters). Interestingly, further BLAST analysis using our draft genome sequence data of serotype 22, 26, 32, and 34 reference strains indicated the presence of one more set of dTDP-L-rhamnose synthesis genes on their chromosomes (identity ≥90% and coverage ≥90% in TBLASTX compared to the *cps22*, *cps26*, *cps32*, and/or *cps34* gene clusters) (data not shown). Moreover, tiling microarray analysis (45) and further BLAST analysis using our sequence data suggested that, at least in 28 serotypes (serotypes 1/2, 2 to 14, 16 to 19, 21, 23 to 25, 27 to 31, and 33), these genes are also present apart from their *cps* gene clusters (data not shown). Although rhamnosyltransferase (HG20) is predicted to be encoded only in *cps1/2*, *cps2*, *cps6*, *cps15*, *cps20*, and *cps34* gene clusters, rhamnose might be incorporated into CPs in many serotypes.

(iii) Serotype-specific genes in the *cps* gene clusters. Except for the proteins of 13 HGs (HG83 to HG87, HG91 and HG92, HG115 and HG116, HG136, HG179, and HG232 and HG233), which are less than 50% in size compared with the proteins of other HGs and have more than 50% amino acid sequence identity with the partial regions of the other Cps proteins, the proteins of single-member HGs can be considered to be serotype-specific gene products (see Table S5 in the supplemental material). No serotype-specific gene is present in the *cps* gene clusters of serotypes 1, 2, 1/2, and 14, although several putative transposase genes conserved in serotypes 1 and 14 and in serotypes 2 and 1/2 are inverted and/or translocated in the *cps* gene clusters (see Fig. S5 and S6 in the supplemental material). With the exception of these

four serotypes, every serotype has at least one serotype-specific gene in the cluster. The genes encoding Wzy polymerase, Wzx flippase, glycosyltransferase, and enzymes to modify the repeat units or to add other moieties on CP are included in the serotype-specific genes. Serotype-specific glycosyltransferases were found in 25 serotypes (serotypes 3 to 8, 12, and 15 to 32), suggesting that oligosaccharide structures of the repeat units are different from one another, at least in these serotypes. Although there are no serotype-specific glycosyltransferase genes in *cps10* and *cps13* gene clusters, these clusters possess specific genes involved in the glycerol and sialic acid moieties, respectively. On the other hand, except for a gene encoding a hypothetical protein (*cps11H*), only the Wzy polymerase gene was specific in serotypes 11 and 33. According to the results of our clustering analysis, *cps9J* (*wzy*) and *cps9O-9S* were specific genes of the serotype 9 reference strain. However, because the *cps9O-9S* genes were absent in the genome of another serotype 9 strain sequenced (strain D12, accession number CP002644), the Wzy polymerase gene (*cps9J*) was the only serotype-specific gene for serotype 9 strains. Therefore, to verify the serotype specificity, further studies examining a large number of isolates within each serotype will be required.

(iv) Wzy polymerase and Wzx flippase. As described above and shown in Table 1, all Wzy polymerases, except for those of serotypes 1 and 14 (HG50) and serotypes 2 and 1/2 (HG54), were allocated to serotype-specific HGs. The *cps* gene clusters of serotypes 1 and 14 and those of serotypes 2 and 1/2 were shown to be almost identical in each pair (17, 24), suggesting that the repeat unit structures of CPs are also very similar in each pair. Therefore, Wzy polymerases may differ from each other according to the repeat unit structures in the respective serotypes.

Wzx flippases of 23 serotypes also belong to serotype-specific HGs, and those of the other 12 serotypes were assigned to four HGs (HG13 [serotypes 1, 1/2, 2, 6, 14, 16, and 27], HG43 [serotypes 8, 9, and 33], HG58 [serotypes 6 and 16], and HG71 [serotypes 11 and 34]) (Table 1). Except for serotype 8, Wzx flippases of HG13, HG43, HG58, and HG71 are always encoded together with the glycosyltransferases of HG11 and HG12, HG63, HG57 and HG59, and HG47, respectively (see Tables S4 and S5 in the supplemental material). Therefore, Wzx flippases of these serotypes may recognize common oligosaccharide structures conserved in the different repeat units. Interestingly, the *cps6* and *cps16* gene clusters encode two Wzx flippases; however, it is unknown whether both products are necessary for CP synthesis.

Relationship between similarity of *cps* gene clusters and serological cross-reactivity. (i) Serotypes 1 and 14. As reported previously (17, 24) and shown in this study (see Fig. S5A in the supplemental material), the *cps* gene clusters of serotypes 1 and 14 are almost identical. As expected from the sequence similarity, the serotype 1 reference strain is known to cross-react with anti-serotype 14 serum (7). However, no cross-reaction was reported between the serotype 14 reference strain and anti-serotype 1 serum (7), indicating that some differences in the CP structures certainly exist between the two serotypes. Indeed, Elliott and Tai showed that the serotype 1 CP contains *N*-acetyl galactosamine (43), but the sugar was not present in the serotype 14 CP structure determined by Van Calsteren et al. (46). It is noteworthy that, in serotype 1 reference strain NCTC 10237 and in another strain (strain ST1, accession number CP002651), the genes corresponding to a glycosyltransferase gene (*cps14G*) of the serotype 14 reference strain seem to be disrupted by a single base deletion and the resultant frameshift mutation (see Fig. S5A). According to the reported sequence (accession number JF273644), the gene is not disrupted in serotype 1 strain 5428, but the deduced amino acid sequence showed some differences from that of Cps14G (Fig. S5A). In addition, sequence differences were also observed among several *cps* gene products of the serotype 1 and 14 strains (one example observed in CpsE is shown in Fig. S5A). Therefore, the antigenic differences between serotype 1 and 14 strains might be attributed to such point mutations, which may lead to gene inactivation and/or altered specificity of the enzyme, although further study is needed to investigate this possibility.

(ii) Serotypes 1/2 and 2. Serotype 1/2 strains react with both anti-serotype 1 and 2 sera. In accordance with previous reports (17, 24), our results also showed that the *cps* gene clusters of serotypes 1/2 and 2 are almost identical (see Fig. S5B in the supplemental material). In addition to the *cps1/2* and *cps2* gene clusters analyzed in this study, *cps* gene clusters of serotype 1/2 strain SS12 and several serotype 2 strains have been sequenced so far (SS12, accession number CP002640; serotype 2 strains, accession numbers FM252031, FM252032, CP003736, CP000407, CP000408, CP002570, and CP000837). Although C-terminal amino acid sequences of a glycosyltransferase of strain SS12 (corresponding to Cps1/2G of the serotype 1/2 reference strain 2651) are different from those of Cps2G of serotype 2 strain P1/7, such differences were not observed between strains 2651 and P1/7 (see Fig. S5B) and, as far as we analyzed, we could not find any sequence differences that are likely to contribute to the antigenic differences between the two serotypes. Because the serotype 2 CP structure has already been determined (44), structural determination of the serotype 1 and 1/2 CPs will provide some insights into the epitopes recognized by the anti-serotype 1 serum.

(iii) Other serotypes. Our clustering analysis using the *cps* gene products of all serotypes showed that amino acid sequences of many *cps* gene products are relatively well conserved between serotypes 9 and 33, serotypes 10 and 11, and serotypes 15 and 34 (Fig. 3). As expected from the results, two-way cross-reactions have been observed between serotype 9 and 33 and serotype 15 and 34 strains (8). However, no cross-reaction was reported between serotype 10 and 11 reference strains (7). Serotype 10-specific enzymes that are predicted to be related to the glycerol moieties on CP (Cps10Q of HG133 and Cps10R of HG134) and/or a glycosyltransferase of serotype 11, which is not encoded in the *cps10* gene cluster (Cps11O of HG47) (Fig. 3A), may contribute to the antigenic differences between the two capsular types.

Except for the above cases, every serotype possesses more than five specific *cps* genes compared to the other serotypes; nevertheless, cross-reactivity has been observed in several cases (6–9). Therefore, some serotype-specific sera may recognize common structures present in different repeat units (e.g., serotypes 6 and 16 both possess sialic acid synthesis genes and two-way cross-reactions have been observed) (7). Alternatively, these sera may react not only with the CP but also with other cell surface components, which are common to several serotypes.

Potential serotype conversion and evolution of the *cps* gene clusters in *S. suis*. (i) Horizontal transfer and replacement of *cps* gene clusters. Multilocus sequence typing (MLST) analysis using 294 *S. suis* isolates by King et al. (47) suggested the potential horizontal transfer of *cps* genes due to the presence of a sequence type (ST) containing isolates of the multiple serotypes and the presence of isolates of the same serotype in different STs. Our present data

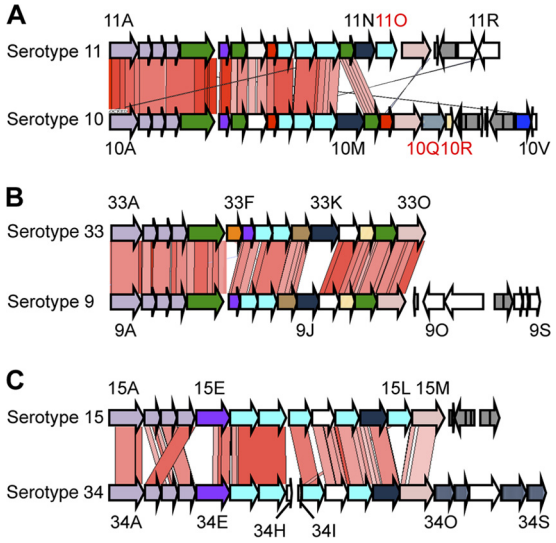

FIG 3 Comparison of the *cps* gene clusters between serotypes 11 and 10 (A), serotypes 33 and 9 (B), and serotypes 15 and 34 (C). Pairwise TBLASTX comparison data are displayed. Conserved regions are indicated by red (same direction) or blue (opposite direction) blocks. Some gene names ("*cps*" omitted) are indicated beside the arrows. Each colored arrow represents the gene whose predicted function is shown in Fig. 2.

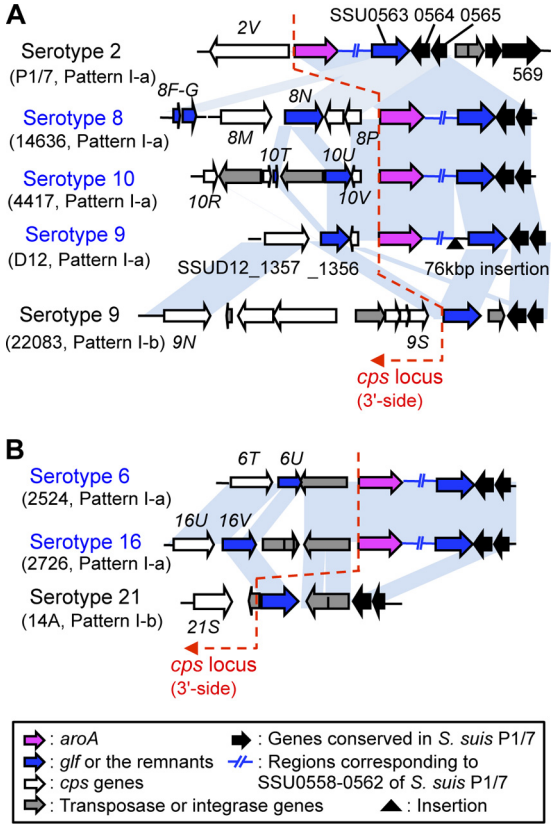

FIG 4 Genetic organizations of 3′-side boundary regions of the *cps* loci. (A) The 3′-side boundary regions of *cps2*, *cps8*, *cps9*, and *cps10* loci. The serotype 8 reference strain 14363 possesses genes corresponding to SSU0563 (*glf*) to SSU0565 of *S. suis* serotype 2 strain P1/7 (*cps8N-P*) and the putative remnants of a *glf* homologue (*cps8F-G*) in the *cps* gene cluster. The serotype 10 reference strain 4417 and serotype 9 strain D12 also harbor *glf* and SSU0565 homologues in the *cps* gene clusters. On the other hand, serotype 9 reference strain 22083 does not have these homologues in the *cps* gene cluster. (B) The 3′-side boundary regions of *cps6*, *cps16*, and *cps21* loci. The *glf* homologues are present in the *cps* gene clusters of serotype 6 and 16 reference strains, but the genes corresponding to SSU0564 and SSU0565 are absent from the gene clusters. Homologues of a putative transposase gene located downstream of *glf* in the serotype 21 reference strain are present upstream of *aroA* in serotype 6 and 16 reference strains. Conserved regions by BLASTN analysis are represented by light blue blocks. The locus tag numbers or gene names ("*cps*" omitted) are appended to the corresponding genes.

further support the possibility of horizontal *cps* gene transfer among *S. suis* strains. For example, by the MLST analysis, reference strains that possess different *cps* gene clusters (serotypes 17 and 19) (Fig. 2) have been shown to have similar genetic backgrounds (both are ST76 strains) (47). This strongly suggests the presence of intraspecies horizontal transfer of serotype-specific *cps* genes. The chromosomal loci of both *cps17* and *cps19* gene clusters are pattern I-a, and the flanking sequences of the *cps* gene clusters are conserved (Fig. 1). Therefore, although the mechanism of the horizontal gene transfer is still unknown, once *cps* regions are horizontally acquired, replacement of the entire *cps* gene clusters by homologous recombination via the conserved flanking regions may occur among strains with the *cps* gene clusters of the same chromosomal location pattern. In addition, as shown in Fig. 1, the *cps* gene clusters of chromosomal location patterns II and III are inserted into intergenic regions apart from the *orfX-glf* locus of P1/7 (pattern I-a strain). Therefore, loss of an entire *cps* gene cluster from a chromosomal locus and acquisition of a novel cluster at a different locus may also occur by independent events among strains of different patterns. Alternatively, in some *S. suis* strains, large-scale reorganization events such as inversions, translocations, and segmental duplications may occur in their genomes and yield different chromosomal location patterns.

The first four *cps* genes (*cpsA-cpsD*) are conserved in all *cps* gene clusters, and some of them shared more than 95% similarity in the nucleotide sequences (see Fig. S3 in the supplemental material). In addition, the 3′-side regions of several *cps* gene clusters whose *cpsA-cpsD* genes share high sequence similarity show similar genetic organizations (e.g., serotypes 1, 1/2, 2, and 14, serotypes 4, 5, 7, and 23, serotypes 17, 18, and 19, and serotypes 11, 12, 28, and 30) (see Fig. S3 and S6 in the supplemental material). Therefore, it is possible to speculate that replacement of partial *cps* gene clusters with horizontally acquired *cps* genes may occur via the conserved regions in *cps* gene clusters. Because some 3′-side genes are conserved even among the *cps* gene clusters with different chromosomal location patterns (e.g., *cps21Q-21S* of serotype 21 [pattern I-b], *cps28O-28Q* of serotype 28 [pattern I-a], *cps29P-29R* of serotype 29 [pattern I-b], and *cps30M-30O* of serotype 30 [pattern I-a]) (see Fig. S7 in the supplemental material), such gene transfer events may occur between *cps* gene clusters with different chromosomal location patterns.

As shown in Fig. 1, genes corresponding to SSU0563 (*glf*) to SSU0565 of *S. suis* P1/7 are located outside the *cps* loci of patterns I-a, I-b, II, and III. Interestingly, homologues of these genes are also present in the *cps* gene clusters of the serotype 6, 8, 10, and 16 reference strains (*cps6U*, *cps8F-8G* and *8N-8P*, *cps10T-10V*, and *cps16V*, respectively) (Fig. 2 and 4). In addition, although the *glf* homologue is not present in the *cps* gene cluster of the serotype 9 reference strain 22083, serotype 9 strain D12 possesses a *glf* homologue (SSUD12_1356) in the *cps* gene cluster (Fig. 4A). Moreover,

several mobile genetic elements are also conserved both inside and outside several *cps* loci. For example, a putative transposase gene located downstream of *glf* in the serotype 21 reference strain (i.e., located outside the *cps21* locus of pattern I-b) is highly homologous to those located upstream of *aroA* in the serotype 6 and 16 reference strains (i.e., located inside the *cps6* and *cps16* loci of pattern I-a) (Fig. 4B). Although precise boundaries of the *cps* gene clusters are still controversial, homologous recombination events between sequences conserved both inside and outside the *cps* loci may also contribute to the generation of the diversity of the *cps* gene clusters.

(ii) Small-scale mutations and deletion, insertion, and inversion events. In *S. pneumoniae*, mutations of a few *cps* genes affect the repeat unit structures of their CPs, resulting in the generation of serotype variation (e.g., serotypes 6A and 6B, serotypes 9L and 9N, serotypes 12A and 12F, and serotypes 15B and 15C) (13, 48). As discussed above, a glycosyltransferase gene of *S. suis* serotype 1 (*cps1G*) seems to be disrupted by a frameshift mutation (see Fig. S5 in the supplemental material), and this mutation might contribute to the different antigenicity between serotypes 1 and 14. In addition to *cps1G*, more than 20 *cps* genes in serotypes 1, 1/2, 3, 7, 8, 17, 18, 19, 26, and 34 are disrupted by nonsense or frameshift mutations (Fig. 2). In the *cps1* and *cps10* gene clusters, truncation by the insertion of an IS element is found (*cps1V-1W* and *cps10T-10U*) (Fig. 2). Moreover, inversion events (e.g., 3′-side of three genes in *cps* gene clusters of serotypes 4, 5, 7, 17, 18, 19, and 23) and insertion or deletion events (e.g., *cps3O-3P* and *cps30P-30Q*) appear to have occurred (see Fig. S6 in the supplemental material). Some of these mutations and genetic events may also contribute to the generation of serotype variation in *S. suis*.

In summary, the variation of *S. suis* CP is assumed to be generated, at least in part, by the accumulation of intraspecies *cps* gene transfers and spontaneous mutations. The natural habitat of *S. suis* is the upper respiratory tract (particularly the tonsils and nasal cavities) and the genital and alimentary tracts of pigs, and it is known that many pigs harbor a variety of *S. suis* strains or serotypes in their tonsils (2, 49, 50). Therefore, intraspecies *cps* gene transfer in the *S. suis* communities inhabiting pigs may play an important role in generating the variation of *S. suis* CP.

Future research directions in *S. suis* CP. In this study, we analyzed the complete set of *S. suis cps* gene clusters and provided insight into the antigenic differences of *S. suis* CP and how CP variations could arise in this bacterium. However, the precise function is still unknown in most of the *cps* genes. In addition, CP structures have so far been determined only in serotypes 2 and 14 (44, 46). Therefore, a more exhaustive analysis of each *cps* gene and the CP structure determination in other serotypes will be necessary for complete elucidation of the role, biochemistry, genetics, and evolution of *S. suis* CPs. In this regard, our results will serve as an important base for further studies.

Our results may also contribute to the development of a novel molecular serotyping method. Although several serotype-specific PCR assays are available at present (16, 18, 51–53), it is still not possible to distinguish one from another in all serotypes. For serotyping, preparation of typing sera for all serotypes is necessary, but this is not easy in all the diagnostic laboratories. Moreover, cross-reactions make serotyping difficult in some cases. Development of a novel PCR assay using the serotype-specific genes found in this study will therefore facilitate the determination of *S. suis* infection in diagnostic laboratories.

ACKNOWLEDGMENTS

This work was supported by a Grant-in-Aid for Young Scientists (B) (23480310) and a Grant-in-Aid for Exploratory Research (24659197) from the Japan Society for the Promotion of Science, a Grant-in-Aid for Scientific Research (C) (22592032, 23580420, and 24590525), a Grant-in-Aid for Scientific Research on Innovative Areas (24117508), and by the Program of Japan Initiative for Global Research Network on Infectious Diseases from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

REFERENCES

1. **Gottschalk M, Xu J, Calzas C, Segura M.** 2010. *Streptococcus suis*: a new emerging or an old neglected zoonotic pathogen? Future Microbiol. **5**:371–391.
2. **Staats JJ, Feder I, Okwumabua O, Chengappa MM.** 1997. *Streptococcus suis*: past and present. Vet. Res. Commun. **21**:381–407.
3. **Mai NT, Hoa NT, Nga TV, Linh LD, Chau DTT, Sinh DX, Phu NH, Chuong LV, Diep TS, Campbell J, Nghia HD, Minh TN, Chau NV, de Jong MD, Chinh NT, Hien TT, Farrar J, Schultsz C.** 2008. *Streptococcus suis* meningitis in adults in Vietnam. Clin. Infect. Dis. **46**:659–667.
4. **Takamatsu D, Wongsawan K, Osaki M, Nishino H, Ishiji T, Tharavichitkul P, Khantawa B, Fongcom A, Takai S, Sekizaki T.** 2008. *Streptococcus suis* in humans, Thailand. Emerg. Infect. Dis. **14**:181–183.
5. **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, 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.
6. **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.
7. **Gottschalk M, Higgins R, Jacques M, Mittal K, Henrichsen J.** 1989. Description of 14 new capsular types of *Streptococcus suis.* J. Clin. Microbiol. **27**:2633–2636.
8. **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.
9. **Perch B, Pedersen K, Henrichsen J.** 1983. Serology of capsulated streptococci pathogenic for pigs: six new serotypes of *Streptococcus suis.* J. Clin. Microbiol. **17**:993–996.
10. **Lecours MP, Segura M, Lachance C, Mussa T, Surprenant C, Montoya M, Gottschalk M.** 2011. Characterization of porcine dendritic cell response to *Streptococcus suis.* Vet. Res. **42**:72.
11. **Smith HE, Damman M, van der Velde J, Wagenaar F, Wisselink HJ, Stockhofe-Zurwieden N, Smits MA.** 1999. Identification and characterization of the *cps* locus of *Streptococcus suis* serotype 2: the capsule protects against phagocytosis and is an important virulence factor. Infect. Immun. **67**:1750–1756.
12. **Roberts IS.** 1996. The biochemistry and genetics of capsular polysaccharide production in bacteria. Annu. Rev. Microbiol. **50**:285–315.
13. **Bentley SD, Aanensen DM, Mavroidi A, Saunders D, Rabbinowitsch E, Collins M, Donohoe K, Harris D, Murphy L, Quail MA, Samuel G, Skovsted IC, Kaltoft 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.
14. **Yother J.** 2011. Capsules of *Streptococcus pneumoniae* and other bacteria: paradigms for polysaccharide biosynthesis and regulation. Annu. Rev. Microbiol. **65**:563–581.
15. **Smith HE, de Vries R, van't Slot R, Smits MA.** 2000. The *cps* locus of *Streptococcus suis* serotype 2: genetic determinant for the synthesis of sialic acid. Microb. Pathog. **29**:127–134.
16. **Smith HE, Veenbergen V, van der Velde J, Damman M, Wisselink HJ, Smits MA.** 1999. The *cps* genes of *Streptococcus suis* serotypes 1, 2, and 9: development of rapid serotype-specific PCR assays. J. Clin. Microbiol. **37**:3146–3152.
17. **Wang K, Fan W, Cai L, Huang B, Lu C.** 2011. Genetic analysis of the capsular polysaccharide synthesis locus in 15 *Streptococcus suis* serotypes. FEMS Microbiol. Lett. **324**:117–124.
18. **Wang K, Fan W, Wisselink H, Lu C.** 2011. The *cps* locus of *Streptococcus suis* serotype 16: development of a serotype-specific PCR assay. Vet. Microbiol. **153**:403–406.

19. **Chen C, Tang J, Dong W, Wang C, Feng Y, Wang J, Zheng F, Pan X, Liu D, Li M, Song Y, Zhu X, Sun H, Feng T, Guo Z, Ju A, Ge J, Dong Y, Sun W, Jiang Y, Yan J, Yang H, Wang X, Gao GF, Yang R, Yu J.** 2007. A glimpse of streptococcal toxic shock syndrome from comparative genomics of *S. suis* 2 Chinese isolates. PLoS One **2**:e315.
20. **Holden MT, Hauser H, Sanders M, Ngo TH, Cherevach I, Cronin A, Goodhead I, Mungall K, Quail MA, Price C, Rabbinowitsch E, Sharp S, Croucher NJ, Chieu TB, Mai NT, Diep TS, Chinh NT, Kehoe M, Leigh JA, Ward PN, Dowson CG, Whatmore AM, Chanter N, Iversen P, Gottschalk M, Slater JD, Smith HE, Spratt BG, Xu J, Ye C, Bentley S, Barrell BG, Schultsz C, Maskell DJ, Parkhill J.** 2009. Rapid evolution of virulence and drug resistance in the emerging zoonotic pathogen *Streptococcus suis.* PLoS One **4**:e6072.
21. **Hu P, Yang M, Zhang A, Wu J, Chen B, Hua Y, Yu J, Chen H, Xiao J, Jin M.** 2011. Complete genome sequence of *Streptococcus suis* serotype 3 strain ST3. J. Bacteriol. **193**:3428–3429.
22. **Hu P, Yang M, Zhang A, Wu J, Chen B, Hua Y, Yu J, Xiao J, Jin M.** 2011. Complete genome sequence of *Streptococcus suis* serotype 14 strain JS14. J. Bacteriol. **193**:2375–2376.
23. **Ye C, Zheng H, Zhang J, Jing H, Wang L, Xiong Y, Wang W, Zhou Z, Sun Q, Luo X, Du H, Gottschalk M, Xu J.** 2009. Clinical, experimental, and genomic differences between intermediately pathogenic, highly pathogenic, and epidemic *Streptococcus suis.* J. Infect. Dis. **199**:97–107.
24. **Zhang A, Yang M, Hu P, Wu J, Chen B, Hua Y, Yu J, Chen H, Xiao J, Jin M.** 2011. Comparative genomic analysis of *Streptococcus suis* reveals significant genomic diversity among different serotypes. BMC Genomics **12**:523.
25. **Mogollon JD, Pijoan C, Murtaugh MP, Kaplan EL, Collins JE, Cleary PP.** 1990. Characterization of prototype and clinically defined strains of *Streptococcus suis* by genomic fingerprinting. J. Clin. Microbiol. **28**:2462–2466.
26. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. **18**:821–829.
27. **Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL.** 2009. BLAST+: architecture and applications. BMC Bioinformatics **10**:421.
28. **Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O.** 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics **9**:75.
29. **Sugawara H, Ohyama A, Mori H, Kurokawa K.** 2009. Microbial Genome Annotation Pipeline (MiGAP) for diverse users, abstr S001-1–2. Twentieth International Conference on Genome Informatics (GIW2009) Poster and Software Demonstrations, Yokohama, Japan.
30. **Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M.** 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. **34**:D32–D36.
31. **Zhao Y, Wu J, Yang J, Sun S, Xiao J, Yu J.** 2012. PGAP: pan-genomes analysis pipeline. Bioinformatics **28**:416–418.
32. **Enright AJ, Van Dongen S, Ouzounis CA.** 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. **30**:1575–1584.
33. **Eisen MB, Spellman PT, Brown PO, Botstein D.** 1998. Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. U. S. A. **95**:14863–14868.
34. **Saldanha AJ.** 2004. Java Treeview—extensible visualization of microarray data. Bioinformatics **20**:3246–3248.
35. **Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL.** 2008. NCBI BLAST: a better web interface. Nucleic Acids Res. **36**:W5–W9.
36. **Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, Parkhill J.** 2005. ACT: the Artemis Comparison Tool. Bioinformatics **21**:3422–3423.
37. **Thompson JD, Higgins DG, Gibson TJ.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4673–4680.
38. **Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S.** 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**:2731–2739.
39. **Chatellier S, Harel J, Zhang Y, Gottschalk M, Higgins R, Devriese LA, Brousseau R.** 1998. Phylogenetic diversity of *Streptococcus suis* strains of various serotypes as revealed by 16S rRNA gene sequence comparison. Int. J. Syst. Bacteriol. **48**:581–589.
40. **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 *Streptococcus orisratti.* Vet. Microbiol. **107**:63–69.
41. **Tien LHT, HT, Nishibori T, Nishitani Y, Nomoto R, Osawa R.** 2013. Reappraisal of the taxonomy of *Streptococcus suis* serotypes 20, 22, 26, and 33 based on DNA-DNA homology and *sodA* and *recN* phylogenies. Vet. Microbiol. **162**:842–849.
42. **Reeves PR, Hobbs M, Valvano MA, Skurnik M, Whitfield C, Coplin D, Kido N, Klena J, Maskell D, Raetz CR, Rick PD.** 1996. Bacterial polysaccharide synthesis and gene nomenclature. Trends Microbiol. **4**:495–503.
43. **Elliott SD, Tai JY.** 1978. The type-specific polysaccharides of *Streptococcus suis.* J. Exp. Med. **148**:1699–1704.
44. **Van Calsteren MR, Gagnon F, Lacouture S, Fittipaldi N, Gottschalk M.** 2010. Structure determination of *Streptococcus suis* serotype 2 capsular polysaccharide. Biochem. Cell Biol. **88**:513–525.
45. **Zheng X, Zheng H, Lan R, Ye C, Wang Y, Zhang J, Jing H, Chen C, Segura M, Gottschalk M, Xu J.** 2011. Identification of genes and genomic islands correlated with high pathogenicity in *Streptococcus suis* using whole genome tiling microarrays. PLoS One **6**:e17987.
46. **Van Calsteren MR, Gagnon F, Calzas C, Goyette-Desjardins G, Okura M, Takamatsu D, Gottschalk M, Segura M.** 2012. Structure determination of *Streptococcus suis* serotype 14 capsular polysaccharide. Biochem. Cell Biol. doi:10.1139/bcb-2012-0036.
47. **King SJ, Leigh JA, Heath PJ, Luque I, Tarradas C, Dowson CG, Whatmore AM.** 2002. Development of a multilocus sequence typing scheme for the pig pathogen *Streptococcus suis*: identification of virulent clones and potential capsular serotype exchange. J. Clin. Microbiol. **40**:3671–3680.
48. **Mavroidi A, Godoy D, Aanensen DM, Robinson DA, Hollingshead SK, Spratt BG.** 2004. Evolutionary genetics of the capsular locus of serogroup 6 pneumococci. J. Bacteriol. **186**:8181–8192.
49. **Lowe BA, Marsh TL, Isaacs-Cosgrove N, Kirkwood RN, Kiupel M, Mulks MH.** 2011. Microbial communities in the tonsils of healthy pigs. Vet. Microbiol. **147**:346–347.
50. **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.
51. **Anusak K, Dejsirilert S, Akeda Y, Sekizaki T, Hamada S, Gottschalk M, Oishi K.** 2012. Fifteen *Streptococcus suis* serotypes identified by multiplex PCR. J. Med. Microbiol. **61**:1669–1672.
52. **Silva LM, Baums CG, Rehm T, Wisselink HJ, Goethe R, Valentin-Weigand P.** 2006. Virulence-associated gene profiling of *Streptococcus suis* isolates by PCR. Vet. Microbiol. **115**:117–127.
53. **Wang K, Sun X, Lu C.** 2012. Development of rapid serotype-specific PCR assays for eight serotypes of *Streptococcus suis.* J. Clin. Microbiol. **50**:3329–3334.