Population Analysis of *Streptococcus suis* Isolates from Slaughtered Swine by Use of Minimum Core Genome Sequence Typing

Han Zheng, Shaobo Ji, Ruiting Lan, Zhijie Liu, Xuemei Bai, Wen Zhang, Marcelo Gottschalk, Jianguo Xu

Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, 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; School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, New South Wales, Australia; Groupe de Recherche sur les Maladies Infectieuses du Porc, Faculté de Médecine Vétérinaire, Université de Montréal, Montréal, Quebec, Canada

*Streptococcus suis*, an important zoonotic pathogen, is a highly diverse species with only a subset of strains that cause disease in humans. Our previous study proposed a minimum core genome (MCG) sequence typing method and defined seven MCG groups, with MCG group 1 as the prevalent group causing human infections. In this study, we identified a set of 10 single nucleotide polymorphisms (SNPs) distributed in six genes that were used to identify the seven MCG groups. The 10 SNPs were typed for 179 *S. suis* isolates collected from slaughtered pigs. The most prevalent groups among the tested isolates were MCG groups 6 and 7. Most of the isolates (147/179) were genotyped as mrp negative, epf negative, sly negative, and CDS2157 positive. The 179 isolates were also typed by multilocus sequence typing (MLST) and divided into 115 sequence types (STs), 111 of which were new. The 6 serotypes (29, 11, 5, 12, 30, and 2) represented 72.3% of the serotyped isolates. Our data show that the typing assay facilitates the application of genome data to the surveillance of *S. suis.*

*S. suis* is an important pathogen of pigs (1) and may cause serious human disease (2–4). However, little is known of the *S. suis* population structure. High genetic diversity within the species adds complexity to the understanding of the clonal relationships between strains, especially low-virulence strains (5–7). In a previous study, we developed a whole-genome sequence-based typing scheme to identify and classify *S. suis* populations as one of seven minimum core genome (MCG) groups (8). In that study, 13 epidemic strain-specific single nucleotide polymorphisms (SNPs), 553 group-specific SNPs, and 129 lineage-specific SNPs were found. None of them had reverse or parallel mutations. We also found that the isolates causing severe human infections, death, and outbreaks fell into MCG group 1.

In order to provide an interim means of MCG typing in laboratories where whole-genome sequencing is not yet available and an economical alternative for identification of the MCG groups, we developed a simplified MCG typing method based on core genome SNPs with a minimum of 10 SNPs distributed in six genes. Healthy pigs often are carriers of *S. suis*, which is considered a potential health hazard for workers in the pig and pork industry (9–15). In this study, we characterized 179 *S. suis* isolates collected from slaughtered pigs by MCG SNP typing, multilocus sequence typing (MLST), virulence marker profiles, and serotyping.

**MATERIALS AND METHODS**

**Specimen collection.** Throat swabs were collected from 3,000 different healthy pigs from January to December 2011 and from 1,500 different healthy pigs from May to August 2012. All of the samples were collected in China from different slaughter houses in Beijing and Jiangsu, Sichuan, and Guizhou provinces. This study was reviewed and approved by the ethics committee of the National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention. The rights and the welfare of the pigs used in the study were adequately protected.

**Isolation of *S. suis***. Selective Todd-Hewitt broth (THB) (Oxoid, Ltd., London, United Kingdom) with 3.75% fetal bovine serum (FBS) containing polymyxin B (10 μg/ml), nitrofurantoin (15 μg/ml), and aztreonam (50 μg/ml) was used to enrich the throat swab specimens. Twenty microliters of culture broth with growth spread on Todd-Hewitt agar with polymyxin B (10 μg/ml), nitrofurantoin (15 μg/ml), aztreonam (50 μg/ml), and neutral red (25 μg/ml). We obtained 80 *S. suis* isolates in 2011 and 99 *S. suis* isolates in 2012. Each isolate came from an individual pig.

**Bacterial isolates and chromosomal DNA preparation.** All of the isolates were confirmed to belong to the *S. suis* species using *gdh* gene PCR, 16S rRNA sequencing, and the biochemical identification system API 20 Strep (bioMérieux, Hazelwood, MO). Chromosomal DNA was prepared from all of the isolates as described previously (16). Thirty-five isolates were whole-genome sequenced by Solexa sequencing and were used in a previous study (8).

**PCR amplification and nucleotide sequence determination.** The genome sequence of strain GZ1 (GenBank accession number CP000837) was used as the reference in this study for primer design. The SNPs and primers used in this study are shown in Table 1. PCR was performed using standard conditions: 30 cycles at 94°C for 30 s, x°C for 30 s, and 72°C for 1 min per kilobase of the predicted product size (where x°C represents an annealing temperature appropriate for the particular primer set used) (Table 1). Details of all oligonucleotides used in this study are shown in Table 1. PCR products were directly sequenced by Sangon Biotech (Shanghai, People’s Republic of China). The sequences were analyzed using DNASTAR software.

**MLST analysis.** MLST was performed using PCR amplification and DNA sequencing of the genes *aroA*, *cpn60*, *dpr*, *gki*, *mutS*, *recA*, and *thrA*, as described previously (17). For each isolate, the alleles at each of the
and amplicon was 277 bp, as determined using agarose gel electrophoresis.

The length of the PCR performed at 94°C for 5 min, followed by 30 cycles at 94°C for 30 s, 53°C (reverse). Amplification was

for 30 s, and 72°C for 5 min. The length of the PCR

and 5

= -GAACAGTCCAGCCTCACG-3

was amplified and sequenced using a previously described method (19, 20). The CDS2157 gene (GenBank accession number JX978834) was amplified and sequenced using a previously described method (19, 20). The CDS2157 gene (GenBank accession number JX978834) was amplified and sequenced using a previously described method (19, 20).

Sequencing of mrp gene and PCR assay for sly, epf, and CDS2157.

The full-length mrp gene was amplified and sequenced using a previously described method (18) or extracted from the draft sequence (GenBank accession numbers PRJNA171448 to PRJNA171478 and PRJNA171480 to PRJNA171483), as done in our previous study (8). Amplification of the

and epf genes was performed using a previously described method (19, 20). The CDS2157 gene (GenBank accession number JX978834) was amplified using the primers 5’-CACCATTCCCTTTATCGC-3’ (forward) and 5’-GAAAGTGCAAGCCGAGGAG-3’ (reverse). Amplification was performed at 94°C for 5 min, followed by 30 cycles at 94°C for 30 s, 53°C for 30 s, and 72°C for 5 min, and then 72°C for 5 min. The length of the PCR amplicon was 277 bp, as determined using agarose gel electrophoresis.

Serotyping. The serotypes of the isolates were determined using a multiplex PCR (mpPCR) assay that identified 33 serotypes of

and CDS2157.

Targeted

Annealing

SNP position(s)

PCR product size (bp)

Table 1: Primers used in the study

<table>
<thead>
<tr>
<th>Primer no.</th>
<th>MCG group(s)</th>
<th>Sequence (5’ to 3’)</th>
<th>Targeted gene</th>
<th>Annealing temp (°C)</th>
<th>SNP position(s) (GZ1 genome)</th>
<th>PCR product size (bp)</th>
</tr>
</thead>
</table>
| 1          | 1            | Forward: TCCATTTGATAAGGACTC  
Reverse: AAGGAAGACAAGGCCCAAGG | SSGZ1_1981 | 48 | 2028696, 2028744 | 557 |
| 2          | 2/5          | Forward: GGCTGTCCTGGTATTATCA  
Reverse: TTCTATCTTCCCTCTCTCA | SSGZ1_0777 | 48 | 824818, 825000 | 626 |
| 3          | 3            | Forward: ACCAAGTAAATGCCGAAAAAGA  
Reverse: GCCAGAAACCCAGCACCA | SSGZ1_0776 | 51.7 | 822644 | 468 |
| 4          | 4            | Forward: ACCAAGCAGGAAAGACTGTTGACT  
Reverse: TACGCTTTGGCATCTACCT | SSGZ1_0114 | 53.7 | 107453 | 866 |
| 5          | 6/7-1/7-2    | Forward: GAGTAAAAAGCAAAGCTGCT  
Reverse: GAACCGTCAATAACCCAC | SSGZ1_0088 | 48 | 81404, 81419, 81999 | 795 |
| 6          | 7-3          | Forward: GCAAAGGACAGGAGATGCA  
Reverse: AGGCTGGAATACTGAACC | SSGZ1_0530 | 48 | 572576 | 457 |

Table 2: SNPs used in this study

<table>
<thead>
<tr>
<th>Position in GZ1 genome</th>
<th>E*</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
<th>Group 7-1</th>
<th>Group 7-2</th>
<th>Group 7-3</th>
<th>N*</th>
</tr>
</thead>
<tbody>
<tr>
<td>2028696</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>2028744</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>824818</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>822644</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>T</td>
<td>T</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>107453</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>825000</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>81999</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>T</td>
<td>G</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>81404</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>81419</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>572576</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>G</td>
<td>A</td>
<td>G</td>
</tr>
</tbody>
</table>

* E, epidemic strains (ST7 strains).

* N, ungroupable.

Results

Development of MCG SNP typing. In our previous study using genome sequencing, we identified 7 MCG groups. Based on the genome data of 85 isolates, we first identified the minimum number of SNPs needed to assign the 7 MCG groups. The SNPs chosen were based on the following criteria: (i) the target gene contained as many specific SNPs for different groups as possible and therefore required the fewest number of genes, (ii) the length of the amplified product was less than the reading length for the sequence reaction (900 bp), and (iii) the genes selected contained conserved regions without sequence variation among the 85 genome sequences available for S. suis (8) so that primers suitable for amplifying a diverse range of strains could be designed. MCG group 7 is defined by 3 subgroups, 7-1, 7-2, and 7-3, as there are no unique SNPs that define MCG group 7. To further differentiate the epidemic ST7 clone, we added an extra SNP to type this clone within MCG group 1. Therefore, a total of 10 SNPs distributed in six genes were selected (Table 2), including SSGZ1_0088 (encoding a putative preprotein translocase SecY protein, SNP positions 81404, 81419, and 81999), SSGZ1_0114 (encoding the glycosyltransferase family, SNP position 107453), SSGZ1_0530 (encoding a predicted membrane protein, SNP position 572576), SSGZ1_0776 (encoding the carbamoyl-phosphate synthase, large subunit, SNP position 822644), SSGZ1_0777 (encoding a SAM-dependent methyltransferase, SNP positions 824818 and 825000), and SSGZ1_1981 (encoding a hypothetical protein, SNP positions 2028696 and 2028744).

Six primer pairs were designed to identify the region covering the 10 SNPs. The primer pairs could be amplified in all the isolates tested. We used sequencing of the PCR product to identify the
SNPs of interest. The 169 isolates were categorized as belonging to 1 of the 7 MCG groups, and 10 were MCG ungroupable. The most prevalent group was MCG group 6, representing ≥65% of the isolates (118/179). The next most prevalent groups were MCG group 7 (7-1, two isolates; 7-2, five isolates; and 7-3, 17 isolates), MCG groups 4 and 5 (nine isolates each), MCG group 1 (six isolates), MCG group 2 (two isolates), and MCG group 3 (one isolate) (see Table S1 in the supplemental material).

**Relationship between MCGs and STs.** We also typed 179 isolates by MLST; 115 sequence types (STs) were found, among which 111 were newly assigned. Only four of the STs (ST1, ST17, ST28, and ST32) were reported previously, a considerably different observation from that of a previous study (21). Thirteen isolates were untypeable due to the deletion of one of the seven housekeeping genes (see Table S1 in the supplemental material).

A minimum spanning tree of the 435 STs in the *S. suis* MLST database identified 27 clonal complexes (defined as groups of profiles differing by no more than one gene from at least one other ST in the group) (see Fig. S1 in the supplemental material). Within the 27 clonal complexes, 15 included 27 STs and 48 isolates in this study.

MCG groups 1, 2, 3, 4, 5, 6, 7-1, 7-2, and 7-3 and MCG ungroupable contained 2, 1, 1, 2, 4, 80, 1, 4, 13, and 7 STs, respectively. The 15 clonal complexes were distributed in MCG groups 1 (2 complexes), 2 (1 complex), 4 (2 complexes), 5 (1 complex), 6 (7 complexes), and 7-3 (2 complexes) (see Fig. S1 in the supplemental material). Each ST and clonal complex were represented in only one MCG group or lineage. Isolates of the same ST and ST complex were clustered into the same MCG group.

**Genotypes of *mrp*, *epf*, and *sly* and their distribution in different MCGs.** Most of the isolates (150/179, 83.8%) were *mrp* negative. Twenty-nine isolates contained putative full-length *mrp* gene copies; five were MCG group 1, two were MCG group 2, seven were MCG group 4, nine were MCG group 5, two were MCG group 7-1, and four were MCG ungroupable. Based on the *mrp* subtypes reported in North America (18), the sequences of *mrp* were grouped into one of three subtypes, EU (four isolates), NA1 (22 isolates), or NA2 (three isolates). The three genotypes are very similar at the 5' and 3' ends, while variation has been observed in the central portion of the gene. Only nine isolates contained the *sly* gene. Six were from MCG group 1, two were from MCG group 4, and one was MCG ungroupable. Isolates positive for *epf* (*n* = 6) were all from MCG group 1.

There were eight genotypes of *mrp*, *epf*, and *sly*; primarily based on *mrp* variation: *mrp*<sup>EU</sup> *epf*<sup>−</sup> *sly*<sup>+</sup>; *mrp*<sup>NA1</sup> *epf*<sup>−</sup> *sly*<sup>+</sup>; *mrp*<sup>NA2</sup> *epf*<sup>−</sup> *sly*<sup>+</sup>; *mrp*<sup>NA2</sup> *epf* negative *sly*<sup>−</sup>; *mrp* negative *epf* negative *sly*<sup>−</sup>; *mrp*<sup>NA1</sup> *epf* negative *sly*<sup>−</sup>; and *mrp* negative *epf* negative *sly*<sup>−</sup>. Of the isolates, 82.1% (147/179) were genotyped *mrp* negative *epf* negative *sly* negative. Isolates from MCG group 1 were genotyped *mrp*<sup>EU</sup> *epf*<sup>−</sup> *sly*<sup>+</sup>, *mrp* negative *epf*<sup>−</sup> *sly*<sup>+</sup>, or *mrp*<sup>NA2</sup> *epf*<sup>−</sup> *sly*<sup>+</sup>. All the isolates from MCG group 2 harbored *mrp*<sup>NA2</sup> (see Table S2 in the supplemental material).

**All but ST1 isolates harbored the CDS2157 gene.** In our previous study, the CDS2157 gene annotated as RNA binding SI was present in all intermediate and weakly virulent (I/WV) and virulent (V) groups (except for the ST13 strain) but not in the epidemic and highly virulent (E/HV) group (22). We also tested for the presence of this gene in the 179 isolates. Our results showed that all except five ST1 isolates from MCG group 1 harbored the CDS2157 gene.

**Relationship between MCG groups and serotypes.** The serotypes of 83 isolates were identified using the mPCR assay (16). The result was confirmed using the seroagglutination test. Within the known serotypes, serotype 29 (*n* = 19) was the most prevalent, followed by serotypes 11 (*n* = 9), 2 (*n* = 8), 5 (*n* = 8), 12 (*n* = 8), 30 (*n* = 8), and 24 (*n* = 4). All of them but serotype 24 were found to be associated with multiple MCG groups. In addition, three isolates (each of serotypes 1/2, 7, and 10, two isolates (each) of serotypes 8 and 31, and a single isolate each of serotypes 9, 14, 15, 16, 21, and 22 were also identified (see Table S3 in the supplemental material).

A number of MCG groups which contain isolates of multiple serotypes were also identified. MCG groups 6 and 7 each contained 7 serotypes, MCG ungroupable included 4 serotypes, MCG groups 1 and 4 each contained 4 serotypes, and MCG ungroupable included 2 serotypes.

mPCR showed that 96 isolates were not assigned to a known type. Sequencing of the *cps* gene cluster showed that these isolates possess undescribed *wzy* genes compared to those of currently known serotypes (16, 23) (data not shown) and may represent new serotypes. Of the 96 untypeable isolates, 10 new *wzy* sequences were found. Their *cps* gene clusters were also greatly different from those of the known 33 serotypes (16, 23) (data not shown). Almost all of the untypeable isolates were from MCG group 6 (84.4%, 81/96) or 7 (14.6%, 14/96 isolates).

**DISCUSSION**

Studying the population structure of *S. suis* may reveal important information about the epidemiology of *S. suis* infection and facilitate the rapid identification of potentially virulent strains. The ability to cluster related isolates and differentiate unrelated isolates is crucial for understanding the population biology of *S. suis*. Whole-genome comparative analysis and collection of high-confidence global SNPs provide ideal typing targets for investigating the population structure of bacterial species (24–26). The reliability of phylogenetic inference based on MLST is adversely affected by the frequency of recombinations that obscure phylogenetic signals, as *S. suis* is a weakly clonal species. We showed previously that MCG groups are the ideal units for investigating the population structure of *S. suis* based on genome sequence analysis of 85 isolates (8). In this study, we translated genome-based MCG group assignment by identifying a set of 10 SNPs distributed in six genes which can accurately assign the isolates to a unique clade or lineage and distinguish epidemiologically associated strains. This approach offers a practical, rapid, and cost-effective assay that is suitable for laboratories where whole-genome sequencing is not yet available.

In the present study, we were unable to group 10 isolates. These isolates possibly had a significant recombinant history that cannot be reliably allocated. The most prevalent MCG groups among the tested isolates were groups 6 and 7. Only six isolates were from MCG group 1, to which the epidemic clone belongs, and no isolates associated with the epidemic clone were found. Using *Streptococcus pneumoniae* strain R6 as an out group, groups 6 and 7 diverged the earliest, while group 1 diverged most recently. This indicates that most isolates from carrier animals were from the groups that diverged earlier.

The presence of some genes, such as *mrp*, *epf*, and *sly*, has been...
associated with virulence (27, 28). We previously reported that MCG group 1 includes highly virulent isolates and contains the greatest number of virulence genes. Conversely, MCG groups 6 and 7 carry the lowest number of virulence genes (8). Here, we found that 98.6% of the isolates from MCG groups 6 and 7 were genotyped mrp negative, efp negative, sly negative, which is consistent with our previous findings.

Based on the presence/absence of genes using comparative genomic hybridization (CGH), we previously clustered 40 isolates into the E/HV, V, and I/WV groups (22). The virulence levels of the three groups decreased incrementally from the E/HV group to the I/WV group. There was a clear tendency in the association of isolates belonging to the E/HV group with MCG group 1. The genotype of mrp harbored by all isolates from the E/HV group was mrp$^{EU}$ (our unpublished observations) and CDS2157 negative. Of the 12 MCG group 1 isolates, 11 were genotyped mrp$^{EU}$ and CDS2157 negative. Only one isolate (NCTC 10446, serotype 15, ST81) was genotyped as mrp$^{NA2}$ and CDS2157 positive. The isolate was also clustered into the V group by CGH (22). We also found a similar isolate (YS12) in this study. These two isolates, although assigned to MCG group 1, may have diverged earliest (8) and shared some characteristics with MCG group 2, since MCG group 2 was predominantly of the genotype mrp$^{NA2}$. Studies on more isolates from diseased pigs and patients will be required to evaluate the public health significance of MCG group 2.

CDS2157 belongs to the Tex family of proteins that contain the S1 RNA-binding domain at the C terminus. We found in our previous study that this gene was specific to the I/WV group (22). We proposed that the presence of this gene may be negatively associated with virulence (22). In this study, we found that the gene was absent in all highly virulent ST1 isolates, further supporting our hypothesis.

The mechanisms used by S. suis to colonize the host are poorly understood. It is proposed that S. suis gains entry to the systemic circulation primarily through the palatal tonsils after adhesion and invasion of the epithelial cells (29). Once S. suis reaches deep tissues and/or the bloodstream, it is able to resist phagocytosis and persist in the blood at high concentrations with inflammatory consequences. CPS and sulisyn may play important roles in these stages (30–32). Most invasive S. suis diseases are caused by serotypes 1/2, 2, 3, 4, 8, 9, and 14 (18, 33–35). In our previous study, the blood bacterial loads of mice infected with strains belonging to these serotypes were higher than those of other serotype strains (22). We found in this study that <10% of the isolates from healthy pigs were sulisyn positive or belonged to serotypes 1/2, 2, 3, 4, 7, 8, 9, or 14. Therefore, most of the isolates from carrier pigs found in this study may be less capable of causing serious infections in humans.

Considerable genetic diversity between strains of the same serotype has been observed, and there is also far more serotype diversity with 10 novel wzy sequences from the 96 isolates that cannot be serotyped using current PCR molecular serotyping (16).

In conclusion, we developed an SNP-based typing method that can categorize isolates into the correct MCG groups, defined by whole-genome sequencing, for a robust, economical, and simple means of differentiating the population structure of S. suis. Our data show that MCG SNP typing assay facilitates the application of genome data to the surveillance of S. suis. Our analysis of 179 isolates isolated from pigs using MCG SNP typing and other molecular typing also expanded our knowledge on the population structure and genetic diversity of S. suis.

ACKNOWLEDGMENTS
This work was supported by the Ministry of Science and Technology, People’s Republic of China (grants 2012ZX10004-215, 2013ZX10004221, 2011ZX10002-001-002, and 81261120559).

We thank Yan Wang for technical assistance.

REFERENCES