



OPEN ACCESS

EDITED BY

Ben Pascoe,
University of Oxford, United Kingdom

REVIEWED BY

Prapas Patchanee,
Chiang Mai University, Thailand
Opeyemi U. Lawal,
University of Guelph, Canada

*CORRESPONDENCE

Tracy L. Nicholson
✉ tracy.nicholson@usda.gov

RECEIVED 29 June 2023

ACCEPTED 22 August 2023

PUBLISHED 21 September 2023

CITATION

Nicholson TL, Kalalah AA and Eppinger M (2023) Population structure and genetic diversity of *Streptococcus suis* isolates obtained from the United States. *Front. Microbiol.* 14:1250265. doi: 10.3389/fmicb.2023.1250265

COPYRIGHT

© 2023 Nicholson, Kalalah and Eppinger. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Population structure and genetic diversity of *Streptococcus suis* isolates obtained from the United States

Tracy L. Nicholson^{1*}, Anwar A. Kalalah^{2,3} and Mark Eppinger^{2,3}

¹National Animal Disease Center, Agricultural Research Service, United States Department of Agriculture, Ames, IA, United States, ²South Texas Center for Emerging Infectious Diseases (STCEID), The University of Texas at San Antonio, San Antonio, TX, United States, ³Department of Molecular Microbiology and Immunology (MMI), The University of Texas at San Antonio, San Antonio, TX, United States

Diseases caused by the zoonotic pathogen *Streptococcus suis* are an extensive economic problem as well as an animal welfare concern for the global swine industry. Previous studies have evaluated the genomic diversity and population structure of *S. suis* isolates, however, the majority of these studies utilized isolates obtained from countries other than the U.S. This study applied whole genome sequencing and cgMLST-based typing to evaluate the population structure and genetic relatedness among *S. suis* isolates obtained within the U.S. The established high-resolution phylogenomic framework revealed extensive genomic variation and diversity among the sampled *S. suis* isolates, with isolates from the U.S. and from countries outside the U.S. found interspersed in the phylogeny. *S. suis* isolates obtained within the U.S. did not cluster by state or geographic location, however, isolates with similar serotypes, both obtained from within and outside the U.S., generally clustered together. Average nucleotide identity (ANI) values determined for the *S. suis* genomes were extensively broad, approaching the recommended species demarcation value, and correlated with the phylogenetic group distribution of the cgMLST-based tree. Numerous antimicrobial resistance (AMR) elements were identified among both U.S. and non-U.S. isolates with *ble*, *tetO*, and *ermB* genes identified as the most prevalent. The *epf*, *mrp*, and *sly* genes, historically used as markers for virulence potential, were also observed in the genomes of isolates that grouped together forming a subclade of clonal complex 1 (CC1) isolates. Collectively, the data in this report provides critical information needed to address potential biosurveillance needs and insights into the genetic diversity and population structure of *S. suis* isolates obtained within the U.S.

KEYWORDS

Streptococcus suis, whole-genome sequencing (WGS), cgMLST-based typing, comparative genomics, population structure

Introduction

Streptococcus suis is one of the leading causes of bacterial infection in swine and contributes to significant economic losses to the swine industry worldwide (Feng et al., 2014; Segura et al., 2014a,b, 2017). *S. suis* causes a spectrum of clinical disease outcomes in pigs including pneumonia, endocarditis, septicemia, and meningitis. *S. suis* is also a zoonotic pathogen capable of causing diseases in humans, mainly arthritis, sepsis, and meningitis (Wertheim et al., 2009a,b; Feng et al., 2014; Segura et al., 2014a,b). Human infections have been either sporadic, and thought to be acquired from penetrating injuries associated with occupational exposure, or epidemic, and associated with consumption of raw or undercooked pork products, primarily in southeast Asia (Gottschalk et al., 2007; Dutkiewicz et al., 2018).

Virulence mechanisms used by *S. suis* are not extensively characterized (Fittipaldi et al., 2012; Segura et al., 2017). Studies addressing *S. suis* virulence have been confounded because different isolates cause a spectrum of disease outcomes ranging from lethal systemic disease to asymptomatic carriage (Segura et al., 2017; Votsch et al., 2018). Virulence-associated factors generally regarded as the most important contributors to disease are the capsular polysaccharide (CPS), muramidase-released protein (*mrp*), extracellular protein factor (*epf*), and suilysin (*sly*) (Fittipaldi et al., 2012; Segura et al., 2014a,b, 2017). However, none of these factors alone correlate with the ability to cause invasive systemic disease (Fittipaldi et al., 2012; Segura et al., 2017). Additionally, invasive clinical isolates obtained from both people and pigs frequently do not harbor all of these factors (Fittipaldi et al., 2012; Feng et al., 2014; Segura et al., 2014a,b, 2017). In addition to being considered a major virulence-factor for *S. suis*, CPS is the basis for the typing system of *S. suis* (Liu et al., 2013). Genotyping isolates using *epf* + *mrp* + *sly* + as markers was historically used in epidemiological studies to predict the virulence potential of *S. suis* isolates, particularly serotype 2 isolates obtained from European countries (Gottschalk et al., 1998; Wisselink et al., 2000; Fittipaldi et al., 2012).

Serotyping and multilocus sequence typing (MLST) are the most common methods used to differentiate *S. suis* isolates. Isolates are serotyped based on structural differences in the CPS and to date there are 27 recognized serotypes based on the antigenicity of CPSs and 28 novel CPS loci (Pan et al., 2015; Qiu et al., 2016; Zheng et al., 2017; Huang et al., 2019; Bojarska et al., 2020). Among these, serotype, serotype 2 is regarded as the most virulent and the most frequently isolated serotype from sick pigs (Goyette-Desjardins et al., 2014; Zheng et al., 2014). Additionally, two large outbreaks of human *S. suis* serotype 2 infection have elevated public health concerns (Tang et al., 2006; Yu et al., 2006). A standard MLST scheme has been developed for *S. suis* based on the nucleotide sequences of seven housekeeping genes (King et al., 2002), with isolates subsequently assigned a sequence type (ST) based upon their combination of unique allelic sequences. Worldwide ST1, ST25, and ST28 are the most prevalent STs isolated from swine, the predominance of each ST tends to vary with geographic location (Fittipaldi et al., 2011). For example, in Europe and Asia ST1 strains are more commonly isolated from swine, whereas in North America, ST25 and ST28 are more prevalent (de Greeff et al., 2011;

Fittipaldi et al., 2011; Onishi et al., 2012; Schultsz et al., 2012; Zhu et al., 2013; Estrada et al., 2019).

Phylogenetic studies employ whole-genome sequencing (WGS) and data analysis to better understand the epidemiology, origin, and evolution of bacteria (Eppinger et al., 2014; Sahibzada et al., 2017). WGS analysis can be used to assess isolate relatedness or determine genetic characteristics that define subsets of isolates allowing for the determination of phylogenetic relationships among bacterial isolates providing key data for epidemiological studies addressing biosurveillance and risk assessment (Eppinger et al., 2011, 2014; Rusconi et al., 2016; Sahibzada et al., 2017). Numerous studies have undertaken WGS in combination with comparative genomic approaches to evaluate the genomic diversity harbored by *S. suis* isolates (Weinert et al., 2015; Wileman et al., 2019; Hadjirin et al., 2021; Estrada et al., 2022). The U.S. is the third-largest producer and consumer of pork and pork products globally and unfortunately the vast majority of these studies utilize *S. suis* isolates derived from countries other than U.S. (Principalli et al., 2009; de Greeff et al., 2011; Wang et al., 2011; Goyette-Desjardins et al., 2014; Athey et al., 2015, 2016b; Weinert et al., 2015; Wileman et al., 2019; Scherrer et al., 2020; Guo et al., 2021; Hadjirin et al., 2021; Liang et al., 2021; Estrada et al., 2022). The goal of the current study was to begin to fill this gap by evaluating the genetic relatedness, population structure, and genetic diversity, relating to antimicrobial resistance (AMR) and virulence-associated factors, of *S. suis* isolates obtained within the U.S. using a core genome (cg) MLST-based methodology. Additionally, *S. suis* isolates obtained from various countries outside the U.S. were included in our analyses to provide a broader perspective and examine how the population structure of U.S. *S. suis* isolates compares with isolates obtained from various countries outside the U.S.

Materials and methods

Isolate information

Streptococcus suis isolates obtained from swine samples collected from pigs throughout the U.S. and submitted to the University of Minnesota Veterinary Diagnostic Laboratory between 2015 and 2017 were selected for the project with an effort to include isolates from as many states and an effort to include as many total isolates as possible without duplication ($n = 103$) (Supplementary Table 1; Nicholson et al., 2020; Nicholson and Shore, 2022). This set was complemented with international *S. suis* isolates for which complete genome assemblies were available in GenBank as of March 22, were selected for the project ($n = 45$) (Supplementary Table 1). All isolates were either obtained from samples collected as part of previous studies or were obtained from samples submitted as part of field case investigations and did not require Institutional Animal Care and Use Committee (IACUC) approval.

Whole genome sequencing, assembly, and annotation

Draft and complete genomes for *S. suis* isolates obtained from across the United States were generated as previously

described (Nicholson et al., 2020; Nicholson and Shore, 2022). Briefly, total genomic DNA was extracted from isolates grown in Trypticase Soy Broth (BD Biosciences, Sparks, MD, USA) using a MasterPure Gram Positive DNA Purification Kit (Lucigen Corporation, Middleton, WI, USA). For draft genome assemblies ($n = 94$), Illumina HiSeq data was obtained from libraries created using the NEB Ultra II DNA Library Prep Kit (New England Biolabs, Ipswich, MA, USA) sequenced on a HiSeq 3000 instrument generating 2×150 bp paired-end reads. Sequencing reads were assessed for quality using FastQC.¹ Reads were randomly subsampled using seqtk² to target a genome coverage of 150x considering an average expected genome length of 2.1 Mbases. Sequence data were assembled using MIRA v. 4.9.6³ (Chevreux et al., 1999). The average coverage obtained for each isolate is listed in **Supplementary Table 1**. To be retained in an assembly, contigs were required to be >1500 bp in length and have a coverage of >66% of the average coverage for the genome.

For closed genome assemblies ($n = 9$), whole genome sequencing was performed using the Pacific Biosciences (PacBio) platform. Library preparation for PacBio sequencing was performed following the PacBio 10-kb insert library preparation protocol. The 10 kb library for each strain was sequenced using the PacBio RSII platform with two SMRT[®] cells for each isolate. Closed whole-genome assemblies for isolates ISU1606, ISU2660, ISU2414, ISU2515, ISU2614, ISU2714, ISU2812, and SRD478 were generated using the hybrid assembler Unicycler v. 0.4.4 (Wick et al., 2017) software along with both the PacBio sequencing reads and the Illumina MiSeq platform paired-end sequencing reads, which were previously used to obtain draft assemblies for the nine U.S. isolates (Hau et al., 2015). Closed whole-genome assembly for isolate ISU2912 was generated using PacBio SMRT Analysis v 2.3.0,⁴ CANU v. 1.3 (Koren et al., 2017), and GARM v. 0.7.5 (Soto-Jimenez et al., 2014) along with PacBio sequencing reads and the Illumina MiSeq platform paired-end sequencing reads (Hau et al., 2015). Unless specified otherwise, default parameters were used for all software. Final annotations were completed using NCBI's Prokaryotic Genome Annotation Pipeline (PGAP) v 4.11 (Tatusova et al., 2016). Both draft and complete genomes and sequence reads are available from GenBank and the Sequence Read Archive (SRA). Accession numbers and genome statistics are summarized in **Supplementary Table 1**.

MLST and CPS serotype determination

Confirmation of isolates as *S. suis* based on nucleotide sequences of *recN*, determination of MLST, and CPS serotype were performed *in silico* using the automated pipeline developed by Athey et al. (2016a), which uses short-read sequencing data to assign sequence types (STs) based on the typing scheme developed by King et al. (2002) available from the Public Databases for Molecular Typing and Microbial Genome Diversity (Jolley et al., 2018). Amplicon sequencing was subsequently utilized to

determine ST for remaining isolates with an unknown and/or uncertain ST. These serotypes were further analyzed *in silico* based on the PCR typing schemes described by Liu et al. (2013) for the classical serotypes, and by Qiu et al. (2016) for the Novel CPS Loci (NCL) serotypes. The CPS locus was either incomplete or did not match any known serotypes for eight isolates (30076, 30831, 38529, 39631, 40426, 40437, 40471, SRD478), and thus designated as undefined (UND) in **Figure 1**. Using previously reported stringency criteria, a clonal complex (CC) was defined as a group of STs comprised of at least six identical alleles and containing at least three STs [Single Locus Variant (SLV) clustering] (Wang et al., 2015; Scherrer et al., 2020). STs that did not fall into any group were classified as singlets and STs that grouped with only one other ST were classified as doublets. A group of three STs that differ from each other by a single locus but lack other connections were not assigned a CC and classified as having no clear founder.

MLST schemas and phylogenetic analyses

Assembled genomes were imported into SeqSphere + v. 8.3 (Ridom GmbH, Münster, Germany) for gene-by-gene alignment, allele calling, and comparison (Junemann et al., 2013). A core genome (cg) MLST schema was developed using the closed chromosome of *S. suis* (serotype 2; ST 1) strain ISU1606 (GenBank accession CP030017) as seed. Core genome and accessory MLST targets were identified according to the inclusion/exclusion criteria of the SeqSphere + Target Definer. The allele information from core, accessory, and combined gene sets were used to establish phylogenetic hypotheses using the neighbor-joining method (Saitou and Nei, 1987). The resulting tree was visualized and decorated in iTol v. 6.7.4 (Letunic and Bork, 2021).

Average nucleotide identity

Average nucleotide identity (ANI) between the designated reference genome *S. suis* strain ISU1606 (serotype 2; ST 1) and the analyzed genomes were calculated using FastANI, based on MinHash mapping (Jain et al., 2018) in Galaxy (Goecks et al., 2010; Afgan et al., 2018).

Antimicrobial resistance gene profiling

Abricate (Seeman T, Abricate)⁵ was used to identify antimicrobial resistance genes using the AMR gene databases from the Comprehensive Antibiotic Resistance Database (CARD) (Jia et al., 2017), ResFinder [Center for Genomic Epidemiology, (Zankari et al., 2012)] and the NCBI Bacterial Antimicrobial Resistance Reference Gene Database (BioProject Accession PRJNA313047), which was downloaded in April 2019. A minimum percent identity threshold of 80% was used to identify AMR genes in the assembled genomes. Statistical analysis was completed using

1 <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

2 <https://github.com/lh3/seqtk>

3 <https://mira-assembler.sourceforge.net/docs/>

4 <https://www.pacb.com/support/software-downloads/>

5 Github <http://github.com/tseemann/abricate>.

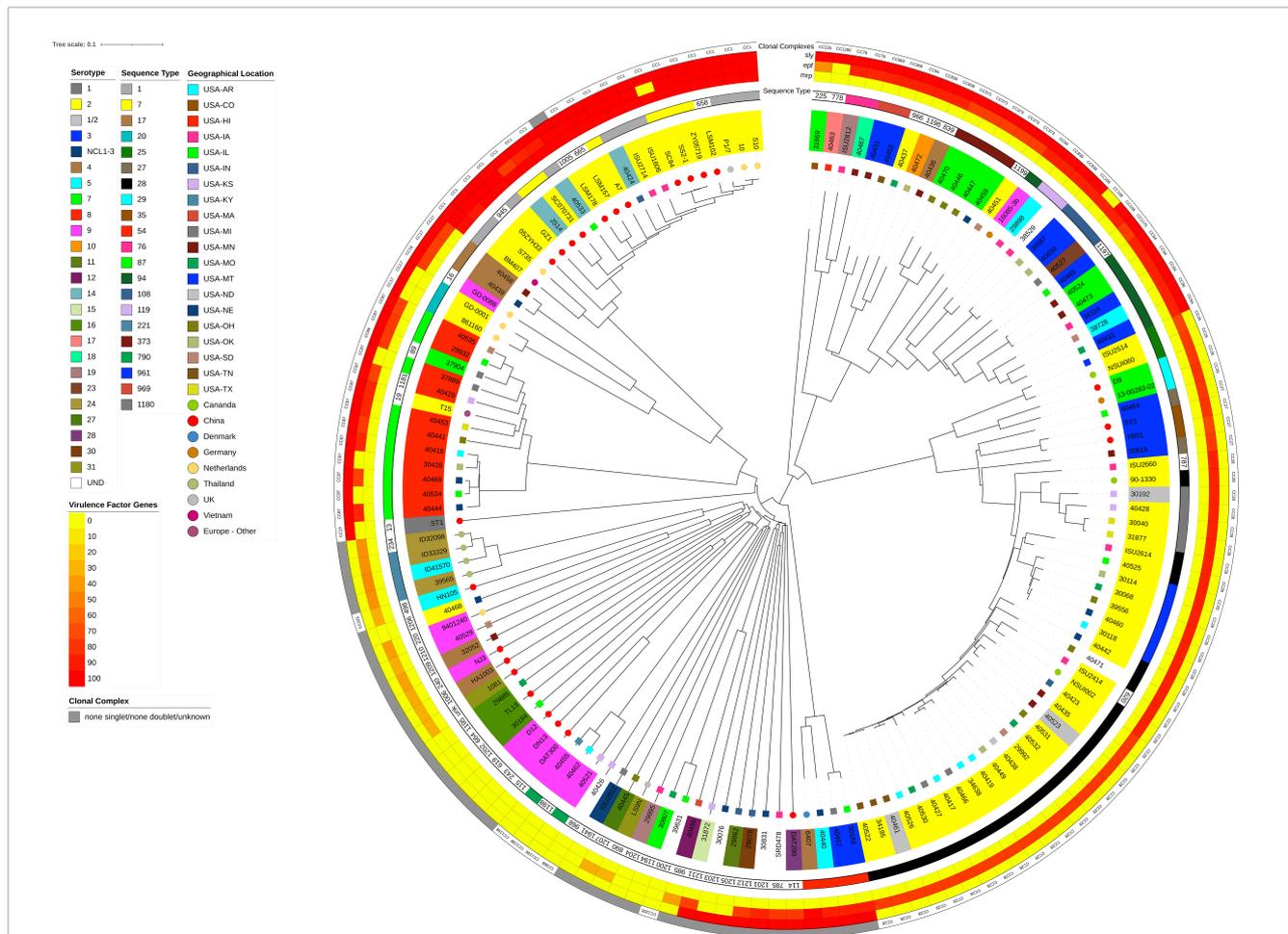


FIGURE 1
 Neighbor-Joining tree based on cgMLST analysis. The closed chromosome of *S. suis* strain ISU1606 (serotype 2; ST 1) served as seed in Ridom SeqSphere + (Junemann et al., 2013). The shared gene inventory was determined at 1,879 genes, comprised of 565 core and 1,314 accessory genes according to the inclusion/exclusion criteria of the SeqSphere + Target Definer. Distance values represent the number of genes with differing allele status in the network. Clonal complex, the nucleotide percentage identity for *epf*, *mrp*, and *sly* converted into a heatmap, ST-classification, serotype, and geographic source is shown from outside in. Serotype for isolates in which the capsule locus was either incomplete or did not match any known serotypes is indicated by UND.

GraphPad Prism 9.0 (GraphPad Software, Inc., La Jolla, CA, USA). Comparisons of AMR resistance gene frequency between U.S. and non-U.S. isolates were completed using Mann-Whitney tests. A *P*-value less than 0.05 was considered significant.

Comparative genomic assessment of virulence-associated genes

Genes *mrp* and *sly* were identified by BLASTN searches and the percent identity for each gene was determined for each isolate relative to the *S. suis* strain ISU1606 (serotype 2; ST 1) orthologs. *In silico* PCR using primers from Silva et al. (2006) was used to identify *epf* gene and distinguishment between *epf* or *epf*^{*}, the long form of extracellular protein factor, based on product size. Percent identity was determined by alignment with either the *epf* gene from strain D282 (accession X71881) or the *epf*^{*} gene from strain 1890 (accession X71881) (Smith et al., 1993).

Results

A total of 148 *S. suis* from the U.S. and various countries outside the U.S. were comprehensively analyzed in this study. International isolates were included if complete genome assemblies were publicly available, allowing for the *in silico* screening for AMR and virulence-factors, as well as sequence type (ST), CPS serotype, and clonal complex (CC) determination. Strain ISU1606 was used as the reference or seed strain because it shares numerous traits and characteristics with *S. suis* strain P1/7 (Holden et al., 2009), which is routinely used a reference strain. ISU1606 and P1/7 are ST1, serotype 2 isolates with similar chromosome size, total number of predicted protein coding sequences (CDSs), G + C content, global synteny, and a similar ability to cause similar clinical disease in swine (Nicholson et al., 2020). To begin evaluating population structure and relatedness among *S. suis* isolates obtained from the U.S. (*n* = 103) and from various countries outside the U.S. (*n* = 45), a neighbor-joining tree was constructed from cgMLST targets using the *S. suis* strain ISU1606 as the seed (Figure 1).

The combined cgMLST target gene inventory was determined at 1,879 genes and was comprised of 565 core and 1,314 accessory genes (Supplementary Table 2). The cgMLST-based tree topology shows a high degree of genome plasticity among all the *S. suis* isolates, with the isolates obtained from countries outside the U.S. interspersed throughout the tree (Figure 1). The extensive genome plasticity depicted by the tree topology is reflective of the high genomic variation and diversity previously described for *S. suis* isolates (Principalli et al., 2009; de Greeff et al., 2011; Weinert et al., 2015; Shelyakin et al., 2019; Guo et al., 2021). Additionally, the *S. suis* isolates obtained within the U.S. did not cluster by state or geographic location (Figure 1).

To examine the nucleotide-level relatedness among the isolates, ANI values were determined for all isolates when compared to *S. suis* strain ISU1606. The ANI values ranged from 99.99% (P1/7) to 94.42% (29995) with an average of 97.12% (Supplementary Table 5). The broad range of ANI values illustrates the extensive genomic variation among the sampled isolates given that ANI values of $\geq 95\%$ are typically observed for organisms belonging to the same species (Richter and Rossello-Mora, 2009; Kim et al., 2014). The ANI values among the sampled *S. suis* isolates correlate with the phylogenetic group distribution of the cgMLST-based tree (Figure 1). For example, isolates with ANI values ranging 99.99 to 99.70% (P1/7, S10, 10, S735, A7, ZY05719, ISU2714, GZ1, LSM157, 40533, 40424, SC84, SS2-1, SC070731, LSM102, LSM178, JS14, BM407, 05ZYH33) from a distinct cluster with a single branch point (Figure 1 and Table 1). Another example includes the isolates with ANI values ranging 94.70 to 94.42% (ISU2912, 40426, 40445, LS9N, 29995) that are grouped together with only few branch points (Figure 1 and Table 2).

Two additional cgMLST-based trees were constructed based on the core genes (Supplementary Figure 1) and the accessory genes only (Supplementary Figure 2). The resulting core-genome based tree resulted in an overall tree topology similar to the cgMLST-based tree (Figure 1 and Supplementary Figure 1). This finding is not surprising as the phylogenetic input is dominated by the shared core genes rather than genes classified as accessory. The accessory-based tree showed in general similar relatedness of isolates, however, reposition of some isolates were observed in the same branches (Figure 1 and Supplementary Figure 2). For example, isolates ZY05719, SS2-1, and SC84 are clustered together from a branch point in the cgMLST-based tree (Figure 1) and in the accessory-based tree these isolates are also clustered together from a branch point but in a repositioned order (Supplementary Figure 2).

Overall isolates with similar CCs and serotypes, obtained from within and outside the U.S., were observed to cluster together (Figure 1). For example, a group of 13 isolates mainly consisting of CC87 (ST19, ST87, and ST1181) were clustered together in Figure 1. Eleven of these 13 isolates were CPS serotype 8. The two non-serotype 8 isolates were T15 (serotype 2, ST19, and CC87) and 37904 (serotype 7, ST89, and CC87) (Figure 1). Additionally, serotype 2 isolates mainly clustered together and are partitioned in two distinct groups. One group largely consists of CC28 (ST28, ST620, ST787, ST961, and ST1180) isolates and a second group that mainly consisting of CC1 (ST1, ST7, ST658, ST665) isolates (Figure 1). The CC28- dominated group forms a large clade of 45 isolates, all serotype 2, except for nine (Figure 1). Of the nine non-serotype 2 isolates, D9 and 13-00285-02 (serotype

TABLE 1 Average nucleotide identity (ANI) values of sampled genomes.

Query genome	ANI value
P1/7	99.99
S10	99.99
10	99.99
S735	99.89
A7	99.88
ZY05719	99.86
ISU2714	99.85
GZ1	99.85
LSM157	99.84
40533	99.84
40424	99.84
SC84	99.82
SS2-1	99.82
SC070731	99.80
LSM102	99.78
LSM178	99.76
JS14	99.73
BM407	99.72
05ZYH33	99.70
ISU2912	94.70
40426	94.62
40445	94.52
LS9N	94.49
29995	94.42

7, ST29, and CC29) cluster together and next to four serotype 3 and CC27 isolates: 40464, ST3, YB51, and 30815 (Figure 1). These two small clusters separate isolates ISU2514 and NSU1060 both serotype 2, ST25, and CC25 from the other thirty-one isolates, which were serotype 2 and CC28. Twenty-one of these serotype 2 isolates are ST28 and five are ST961. Four isolates within this subclade of CC28 isolates are non-serotype 2 isolates. These non-serotype 2 isolates are 30192, 40523, and 40461, all serotype 1/2, and isolate 40471, which harbor an undetermined serotype (Figure 1).

The other group of serotype 2 isolates consisted of twenty-five isolates, which is further divided into two subclades. The smaller subclade consists of five CC17 (ST17 and ST20) and CC16 (ST16) isolates and included isolates 861160 and GD-0001 (both serotype 2, ST20, CC17), isolate GD-0088 (serotype 9, ST16, CC16), as well as isolates 40439 and 40458 (both serotype 4, ST17, CC17). The larger subclade consists of 20 isolates, all CC1 (ST1, ST7, ST658, ST665) except for isolate LSM178, an ST1005 isolate not assigned to a CC and therefore listed as a singlet. Four of the twenty isolates are U.S. isolates (40424, 40533, ISU1606, and ISU2714), while the majority (16/20) of isolates within this subclade are non-U.S. isolates. Additionally, 17 isolates of this subclade were serotype 2 and three isolates were serotype 14 (40424, 40533, and JS14). Isolates within this subclade were mainly ST1 (9/20) or ST7 (6/20).

TABLE 2 Isolates harboring *epf*, *mrp*, and *sly* genes.

Isolate	Country of origin	% Sequence identity			Serotype	ST	CC
		<i>epf</i>	<i>mrp</i>	<i>sly</i>			
10	Netherlands	99.92	100	100	2	1	CC1
05ZYH33	China	99.92	100	100	2	7	CC1
A7	China	99.92	100	100	2	7	CC1
BM407	Vietnam	95.65 ^b	100	99.93	2	1	CC1
GZ1	China	99.92	100	99.93	2	1	CC1
JS14	China	99.92	77.84 ^a	100	14	7	CC1
LSM102	China	99.92	100	100	2	658	CC1
LSM157	China	99.92	99.97 ^a	100	2	665	CC1
LSM178	China	99.92	99.97 ^a	100	2	1005	None-singlet
P1/7	United Kingdom	99.92	100	100	2	1	CC1
S10	Netherlands	99.92	100	100	2	1	CC1
S735	Netherlands	91.37 ^b	99.97	100	2	1	CC1
SC070731	China	99.88	86.2 ^a	100	2	7	CC1
SS2-1	China	99.92	100	100	2	7	CC1
ZY05719	China	99.92	100	100	2	7	CC1
40424	United States, IN	99.92	100	100	14	1	CC1
40533	United States, IL	99.92	89.18	100	14	1	CC1
ISU1606	United States, IA	99.90	100	100	2	1	CC1
ISU2714	United States, IA	99.90	100	100	2	1	CC1

^aPredicted pseudogene based on frameshift. ^b*epf** encoding the long form of extracellular protein factor (EP*).

Genomes were screened for chromosomal mutations and genes conferring antimicrobial resistance. For the U.S. isolates, the number of AMR elements harbored by an individual isolate ranged from one to twelve (Figure 2 and Supplementary Table 3). The average number of AMR elements harbored by an individual U.S. isolate was three (Figure 2 and Supplementary Table 3). The number of AMR elements harbored by an individual isolate ranged from one to nine for the non-U.S. isolates (Figure 2 and Supplementary Table 3). The average number of AMR elements harbored by an individual non-U.S. isolate was two, which was significantly ($P = 0.038$) lower than number of AMR elements harbored by an individual U.S. isolate (Figure 2 and Supplementary Table 3). While numerous AMR elements were identified, the most prevalent were *ble* (100% in both U.S. and non-U.S.) followed by *tetO* (87% U.S. and 49% non-U.S.), and *ermB* (65%, U.S. and 49% non-U.S.). Both the *tetO* and *ermB* loci are absent in 27 genomes, eleven of which were serotype 2 isolates grouped within the larger subclade CC1 isolates in Figure 1.

Genes *epf*, *mrp*, and *sly* have historically been used to predict the virulence potential of *S. suis* (Gottschalk et al., 1998; Wisselink et al., 2000; Fittipaldi et al., 2012). Thus, all genomes were screened for the presence of these loci and the percent identity relative to the reference orthologs was determined. The *epf* gene is present in four U.S. isolates (40424, 40533, ISU1606, and ISU2714) with a 99.9% identity to the reference ortholog. None of the four isolates encode the long form of extracellular protein factor (EP*), which can be expressed by some serotype 2 isolates and produce a high molecular weight variant (> 110 kDa) (Galina et al., 1996; Gottschalk et al.,

1998). This gene is further found in sixteen non-U.S. isolates with a sequence identity range from 91.37 to 99.9%. Two of these non-U.S. isolates, BM407 and S735, carry *epf** (Galina et al., 1996; Gottschalk et al., 1998). The *mrp* gene was found in 86 U.S. isolates and the nucleotide sequence diversity ranged from 31.47 to 100%, with an average sequence identity of 80.21% (Supplementary Table 4). Among non-U.S. isolates, the *mrp* gene was found in 37 isolates and, similar to U.S. isolates, a high nucleotide sequence diversity was observed ranging from 29.76 to 100% with an average 73.72% sequence identity (Supplementary Table 4). The *mrp* gene for nine (13-00283-02, D9, JS14, LSM157, LSM178, NSUI060, SC070731, ST3, T15) of the 37 isolates was predicted to encode a frameshift resulting in a pseudogene. Extensive nucleotide diversity with the *mrp* gene has been previously reported (Vecht et al., 1991; Galina et al., 1996; Baums and Valentin-Weigand, 2009). The *sly* gene was found in 45 U.S. isolates and is highly conserved, with a sequence identity range from 99.2–100% and an average 99.78% sequence identity (Supplementary Table 4). Twenty-four non-U.S. isolates carry the *sly* gene and, similar to U.S. isolates, feature high nucleotide sequence diversity with an average 99.9% sequence identity (99.4–100%) (Supplementary Table 4).

Among the total 148 analyzed isolates, 19 [15 non-U.S. and four U.S. isolates (Table 2) carry all three *epf*, *mrp*, and *sly* genes]. These nineteen isolates form a larger subclade mainly comprised of serotype 2 isolates, as previously mentioned (Figure 1). As stated above, all isolates are CC1 except for isolate LSM178, an ST1005 isolate not assigned to a CC and therefore listed as a singlet (Table 2). Sixteen of these isolates were serotype 2 and three isolates

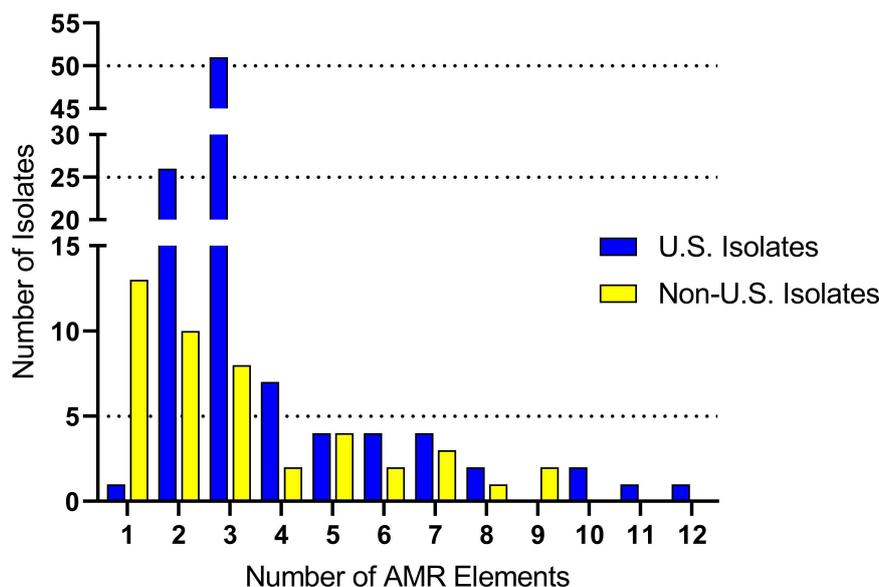


FIGURE 2

Antimicrobial resistance (AMR) gene frequency among *S. suis* isolates. The x-axis indicates the number of AMR genes harbored by a single isolate. The y-axis indicates the number of isolates identified harboring each the respective number of AMR genes.

were serotype 14 (40424, 40533, and JS14) and the majority of these isolates were ST1 (10/19) or ST7 (6/19) (Figure 1 and Table 2).

Discussion

Diseases caused by *S. suis* are a substantial economic problem as well as an animal welfare concern for the global swine industry. Additionally, there are increasing public health concerns regarding the zoonotic potential of *S. suis*. While numerous studies have employed a combination of comparative genomic approaches to assess the genomic diversity and population structure of *S. suis* isolates globally, the overwhelming majority of these studies utilize *S. suis* isolates obtained from countries other than the U.S. (Princivalli et al., 2009; de Greeff et al., 2011; Wang et al., 2011; Goyette-Desjardins et al., 2014; Athey et al., 2015, 2016b; Weinert et al., 2015; Wileman et al., 2019; Scherrer et al., 2020; Guo et al., 2021; Hadjirin et al., 2021; Liang et al., 2021; Estrada et al., 2022). In this report we examined the genomic diversity, phylogenetic relatedness, and population structure of *S. suis* isolates obtained within the U.S. by utilizing a cgMLST approach. Additionally, *S. suis* isolates obtained from various countries outside the U.S. were included in our analyses to provide a broader perspective and examine how the population structure of U.S. *S. suis* isolates compares with isolates obtained from countries outside the U.S.

The resulting phylogenetic tree from the cgMLST-based typing method demonstrated the extensive genomic variation and diversity among all the sampled *S. suis* isolates, with isolates obtained from countries outside the U.S. found interspersed throughout the cgMLST-based tree. The observed extensive genome plasticity in the established cgMLST-phylogeny is reflective of the high genomic variation and diversity previously described for *S. suis* (Princivalli et al., 2009; de Greeff et al., 2011; Weinert et al., 2015; Shelyakin et al., 2019; Guo et al., 2021). In fact, *S. suis*

is considered to have an incompletely defined or open pan-genome since the number of unique or accessory genes has increased and the number of conserved genes have decreased as more genome sequences become available (Wang et al., 2011; Shelyakin et al., 2019; Guo et al., 2021).

Average nucleotide identity has increasingly been utilized as a robust method for assessing genomic similarity between two genomes, with organisms belonging to the same species typically having ANI values of $\geq 95\%$. The ANI values determined for the *S. suis* isolates in this study were extensively broad, ranging from 99.99% (P1/7) to 94.42% (29995), and approach the recommended species demarcation value (Richter and Rossello-Mora, 2009; Kim et al., 2014). The broad range of ANI values observed illustrate the extensive genomic variation among the isolates and correlates with the group distribution in the cgMLST-based tree. The plasticity in genome composition for the sampled *S. suis* isolates is also evidenced in the genome size, ranging from ISU2912 with a size of 2,720,381 bp and thus 31.17% larger than ISU1606 to S735 with a size of 1,980,887 bp (4.49% smaller than ISU1606).

While the *S. suis* isolates obtained within the U.S. did not cluster by state or geographic location, in general all isolates with similar serotypes, regardless of their origins, cluster together in the cgMLST phylogeny. An exception to this general trend were the serotype 2 isolates, which clustered into two distinct groups. One group mostly consisted of ST28 and ST961 and contained three serotype 1/2 isolates and one isolate with an undetermined serotype. The second group, consisting of ST1 and ST7, had three serotype 14 isolates clustered within the group. The clustering of these isolates harboring the 1/2, 2, and 14 serotypes is reasonable and consistent with the high degree of similarity among these serotypes. Specifically, the CPS loci of serotypes 1 and 14 and

serotypes 2 and 1/2 are almost identical and most current PCR methods are unable to differentiate these specific serotypes (Okura et al., 2013). Their CPS structure differs by the substitution of only a galactose (Gal) for a N-acetylgalactosamine (GalNAc) due to a single nucleotide polymorphism in the glycosyltransferase *cpsK* gene (Van Calsteren et al., 2016; Roy et al., 2017). Further, serotype switching of field strains of serotypes 2 and 1/2, and 14 and 1, solely by replacing the amino acid 161 of CpsK has been demonstrated (Roy et al., 2017). In contrast, the CPS loci of all other serotypes are greatly different (Pan et al., 2015; Zheng et al., 2017; Huang et al., 2019).

The clustering of serotype 2 isolates into two distinct groups is supported by previous work that inferred the population structure using MLST and CC approaches (Okura et al., 2013; Hatrongjit et al., 2020; Liang et al., 2021). For example, one branch of serotype 2 isolates within the cgMLST-based tree consisting of ST25, ST27, ST35 isolates further branched into ST28, ST620, ST961, and ST1180, which are the STs comprising one of the serotype 2 groups. Based on CC methods, CC25 (consisting of ST25 isolates), CC27 (consisting of ST27 and ST35 isolates), and CC28 (consisting of ST28, ST620, ST961, ST1180, and ST787) are all linked together (Okura et al., 2013; Hatrongjit et al., 2020; Liang et al., 2021). Additionally, CC1 (consisting of ST1 isolates) and CC7 (consisting of ST7 isolates) are linked (Okura et al., 2013; Hatrongjit et al., 2020; Liang et al., 2021). Consistent with this, the second branch of serotype 2 isolates is comprised of ST1 and ST7 isolates. The results demonstrate that overall, there is a high degree of similarity and congruence between *S. suis* population structure based on previously CC methods and the cgMLST-based strategy utilized in this study. However, the established cgMLST phylogenies provide additional resolution as evident by the observed expanded branching and network.

In addition to the phylogenetic relatedness and population structure of *S. suis* isolates, AMR elements and virulence-associated genes were analyzed. Numerous AMR elements were identified among both U.S. and non-U.S., the most prevalent were *ble*, *tetO*, and *ermB*. However, the *tetO* and *ermB* genes were not observed in the genomes of isolates grouped within the larger subclade CC1 isolates shown in Figure 1. The *epf*, *mrp*, and *sly* genes, historically used as markers for virulence potential, were also observed in the genomes of isolates grouped within the larger subclade CC1 isolates shown in Figure 1.

Taken together, the data provided in this report demonstrate that the extensive genomic variation and diversity observed among *S. suis* isolates globally is also reflected among *S. suis* isolates obtained within the U.S. A high-resolution phylogenetic framework for *S. suis* can help provide data needed to improve risk assessments and address public health concerns associated with occupational exposure to *S. suis*.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Ethics statement

Ethical approval was not required for the study involving animals in accordance with the local legislation and institutional requirements because all isolates were either obtained from samples collected as part of previous studies or were obtained from samples submitted as part of field case investigations and did not require Institutional Animal Care and Use Committee (IACUC) approval.

Author contributions

All authors conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents, materials, and analysis tools, wrote the manuscript, approval of the final version to be published, and agreed to be accountable for all aspects of the work.

Funding

This work was provided by the United States Department of Agriculture, Agriculture Research Service project number 5030-32000-119-00-D, and the South Texas Center of Emerging Infectious Diseases (STCEID), the Department of Molecular Microbiology and Immunology and the High-Performance Computing Cluster operated by University Technology Solutions (UTS) at the University of Texas at San Antonio. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Numbers (SC1GM135110) to ME, and resources provided the SCINet project of the USDA Agricultural Research Service, ARS project number 0500-00093-001-00-D. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Mention of trade names or commercial products in this article is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the USDA. USDA is an equal opportunity provider and employer.

Acknowledgments

We are grateful for exceptional technical support provided by Sarah M. Shore.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1250265/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Core gene MLST. The closed chromosome of *S. suis* strain ISU1606 (serotype 2; ST 1) served as seed for all in Ridom SeqSphere + (Junemann et al., 2013). The core gene inventory was determined at 565 genes according to the inclusion/exclusion criteria of the SeqSphere + Target Definer. Distance values represent the number of genes with differing allele status in the network. ST-classification, serotype, and geographic source is shown from outside in. Serotype for isolates in which the capsule locus was either incomplete or did not match any known serotypes is indicated by UND.

SUPPLEMENTARY FIGURE 2

Accessory gene MLST. The closed chromosome of *S. suis* strain ISU1606 (serotype 2; ST 1) served as seed in Ridom SeqSphere + (Junemann et al., 2013). The accessory gene inventory was determined at 1,314 genes according to the inclusion/exclusion criteria of the SeqSphere + Target

Definer. Distance values represent the number of genes with differing allele status in the network. ST-classification, serotype, and geographic source is shown from outside in. Serotype for isolates in which the capsule locus was either incomplete or did not match any known serotypes is indicated by UND.

SUPPLEMENTARY TABLE 1

Isolate information. The table lists the *S. suis* isolates obtained from across the U.S. (sheet 1) and from various countries outside the U.S. (sheet 2), along with serotype, ST, accession numbers, and genome statistics. Serotype for isolates in which the capsule locus was either incomplete or did not match any known serotypes is indicated by UND.

SUPPLEMENTARY TABLE 2

cgMLST allelic changes. The table lists the total of 253,599 allelic changes identified in the 1,879 cgMLST gene inventory, identified by their locus tag numbers. Allelic change positions and annotations correspond to reference *S. suis* strain ISU1606.

SUPPLEMENTARY TABLE 3

Antimicrobial resistance (AMR) gene distribution. The table lists AMR gene elements identified in all *S. suis* isolates obtained from across the U.S. (sheet 1) and from various countries outside the U.S. (sheet 2).

SUPPLEMENTARY TABLE 4

Virulence-associated gene distribution. The table lists virulence-associated genes (*epf*, *mrp*, and *sly*) identified in all *S. suis* isolates obtained from across the U.S. (sheet 1) and from various countries outside the U.S. (sheet 2).

SUPPLEMENTARY TABLE 5

Average nucleotide identity (ANI) values of cgMLST sampled genomes. The table lists ANI values between the designated reference genome *S. suis* strain ISU1606 and all queried genomes.

References

- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* 46, W537–W544. doi: 10.1093/nar/gky379
- Athey, T. B., Auger, J. P., Teatero, S., Dumesnil, A., Takamatsu, D., Wasserscheid, J., et al. (2015). Complex population structure and virulence differences among serotype 2 *Streptococcus suis* strains belonging to sequence type 28. *PLoS One* 10:e0137760. doi: 10.1371/journal.pone.0137760
- Athey, T. B., Teatero, S., Lacouture, S., Takamatsu, D., Gottschalk, M., and Fittipaldi, N. (2016a). Determining *Streptococcus suis* serotype from short-read whole-genome sequencing data. *BMC Microbiol.* 16:162. doi: 10.1186/s12866-016-0782-8
- Athey, T. B., Teatero, S., Takamatsu, D., Wasserscheid, J., Dewar, K., Gottschalk, M., et al. (2016b). Population Structure and Antimicrobial Resistance Profiles of *Streptococcus suis* Serotype 2 Sequence Type 25 Strains. *PLoS One* 11:e0150908. doi: 10.1371/journal.pone.0150908
- Baums, C. G., and Valentin-Weigand, P. (2009). Surface-associated and secreted factors of *Streptococcus suis* in epidemiology, pathogenesis and vaccine development. *Anim. Health Res. Rev.* 10, 65–83. doi: 10.1017/S146625230999003X
- Bojarska, A., Janas, K., Pejsak, Z., Otulak-Koziel, K., Garbaczewska, G., Hryniewicz, W., et al. (2020). Diversity of serotypes and new cps loci variants among *Streptococcus suis* isolates from pigs in Poland and Belarus. *Vet. Microbiol.* 240:108534. doi: 10.1016/j.vetmic.2019.108534
- Chevreaux, B., Wetter, T., and Suhai, S. (1999). Genome sequence assembly using trace signals and additional sequence information. *Comput. Sci. Biol.* 99, 45–56.
- de Greeff, A., Wisselink, H. J., de Bree, F. M., Schultsz, C., Baums, C. G., Thi, H. N., et al. (2011). Genetic diversity of *Streptococcus suis* isolates as determined by comparative genome hybridization. *BMC Microbiol.* 11:161. doi: 10.1186/1471-2180-11-161
- Dutkiewicz, J., Zajac, V., Sroka, J., Wasinski, B., Cisak, E., Sawczyn, A., et al. (2018). *Streptococcus suis*: A re-emerging pathogen associated with occupational exposure to pigs or pork products. Part II - Pathogenesis. *Ann. Agric. Environ. Med.* 25, 186–203. doi: 10.26444/aaem/85651
- Eppinger, M., Mammel, M. K., Leclerc, J. E., Ravel, J., and Cebula, T. A. (2011). Genomic anatomy of *Escherichia coli* O157:H7 outbreaks. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20142–20147. doi: 10.1073/pnas.1107176108
- Eppinger, M., Pearson, T., Koenig, S. S., Pearson, O., Hicks, N., Agrawal, S., et al. (2014). Genomic epidemiology of the Haitian cholera outbreak: A single introduction followed by rapid, extensive, and continued spread characterized the onset of the epidemic. *mBio* 5:e01721. doi: 10.1128/mBio.01721-14
- Estrada, A. A., Gottschalk, M., Gebhart, C. J., and Marthaler, D. G. (2022). Comparative analysis of *Streptococcus suis* genomes identifies novel candidate virulence-associated genes in North American isolates. *Vet. Res.* 53:23. doi: 10.1186/s13567-022-01039-8
- Estrada, A. A., Gottschalk, M., Rossow, S., Rendahl, A., Gebhart, C., and Marthaler, D. G. (2019). Serotype and Genotype (Multilocus Sequence Type) of *Streptococcus suis* Isolates from the United States Serve as Predictors of Pathotype. *J. Clin. Microbiol.* 57, e377–e319. doi: 10.1128/JCM.00377-19
- Feng, Y., Zhang, H., Wu, Z., Wang, S., Cao, M., Hu, D., et al. (2014). *Streptococcus suis* infection: An emerging/reemerging challenge of bacterial infectious diseases? *Virulence* 5, 477–497. doi: 10.4161/viru.28595
- Fittipaldi, N., Segura, M., Grenier, D., and Gottschalk, M. (2012). Virulence factors involved in the pathogenesis of the infection caused by the swine pathogen and zoonotic agent *Streptococcus suis*. *Future Microbiol.* 7, 259–279. doi: 10.2217/fmb.11.149
- Fittipaldi, N., Xu, J., Lacouture, S., Tharavichitkul, P., Osaki, M., Sekizaki, T., et al. (2011). Lineage and virulence of *Streptococcus suis* serotype 2 isolates from North America. *Emerg. Infect. Dis.* 17, 2239–2244. doi: 10.3201/eid1712.110609
- Galina, L., Vecht, U., Wisselink, H. J., and Pijoan, C. (1996). Prevalence of various phenotypes of *Streptococcus suis* isolated from swine in the U.S.A. based on the presence of muraminidase-released protein and extracellular factor. *Can. J. Vet. Res.* 60, 72–74.
- Goecks, J., Nekrutenko, A., Taylor, J., and Galaxy, T. (2010). Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11:R86. doi: 10.1186/gb-2010-11-8-r86
- Gottschalk, M., Lebrun, A., Wisselink, H., Dubreuil, J. D., Smith, H., and Vecht, U. (1998). Production of virulence-related proteins by Canadian strains of *Streptococcus suis* capsular type 2. *Can. J. Vet. Res.* 62, 75–79.
- Gottschalk, M., Segura, M., and Xu, J. (2007). *Streptococcus suis* infections in humans: The Chinese experience and the situation in North America. *Anim. Health Res. Rev.* 8, 29–45. doi: 10.1017/S1466252307001247
- Goyette-Desjardins, G., Auger, J. P., Xu, J., Segura, M., and Gottschalk, M. (2014). *Streptococcus suis*, an important pig pathogen and emerging zoonotic agent—an update on the worldwide distribution based on serotyping and sequence typing. *Emerg. Microbes Infect.* 3:e45. doi: 10.1038/emi.2014.45

- Guo, G., Du, D., Yu, Y., Zhang, Y., Qian, Y., and Zhang, W. (2021). Pan-genome analysis of *Streptococcus suis* serotype 2 revealed genomic diversity among strains of different virulence. *Transbound Emerg. Dis.* 68, 637–647. doi: 10.1111/tbed.13725
- Hadjirin, N. F., Miller, E. L., Murray, G. G. R., Yen, P. L. K., Phuc, H. D., Wileman, T. M., et al. (2021). Large-scale genomic analysis of antimicrobial resistance in the zoonotic pathogen *Streptococcus suis*. *BMC Biol.* 19:191. doi: 10.1186/s12915-021-01094-1
- Hatrongjit, R., Fittipaldi, N., Gottschalk, M., and Kerdsin, A. (2020). Tools for Molecular Epidemiology of *Streptococcus suis*. *Pathogens* 9:81. doi: 10.3390/pathogens9020081
- Hau, S. J., Bayles, D. O., Alt, D. P., Brockmeier, S. L., Frana, T. S., and Nicholson, T. L. (2015). Draft Genome Sequences of Nine *Streptococcus suis* Strains Isolated in the United States. *Genome Announc.* 3, e1301–e1315. doi: 10.1128/genomeA.01301-15
- Holden, M. T., Hauser, H., Sanders, M., Ngo, T. H., Cherevach, I., Cronin, A., et al. (2009). Rapid evolution of virulence and drug resistance in the emerging zoonotic pathogen *Streptococcus suis*. *PLoS One* 4:e6072. doi: 10.1371/journal.pone.0006072
- Huang, J., Liu, X., Chen, H., Chen, L., Gao, X., Pan, Z., et al. (2019). Identification of six novel capsular polysaccharide loci (NCL) from *Streptococcus suis* multidrug resistant non-typeable strains and the pathogenic characteristic of strains carrying new NCLs. *Transbound Emerg. Dis.* 66, 995–1003. doi: 10.1111/tbed.13123
- Jain, C., Rodriguez, R. L., Phillippy, A. M., Konstantinidis, K. T., and Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat. Commun.* 9:5114. doi: 10.1038/s41467-018-07641-9
- Jia, B., Raphenya, A. R., Alcock, B., Waglechner, N., Guo, P., Tsang, K. K., et al. (2017). CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* 45, D566–D573. doi: 10.1093/nar/gkw1004
- Jolley, K. A., Bray, J. E., and Maiden, M. C. J. (2018). Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res.* 3:124. doi: 10.12688/wellcomeopenres.14826.1
- Junemann, S., Sedlazeck, F. J., Prior, K., Albersmeier, A., John, U., Kalinowski, J., et al. (2013). Updating benchtop sequencing performance comparison. *Nat. Biotechnol.* 31, 294–296. doi: 10.1038/nbt.2522
- Kim, M., Oh, H. S., Park, S. C., and Chun, J. (2014). Towards a taxonomic coherence between average nucleotide identity and 16S rRNA gene sequence similarity for species demarcation of prokaryotes. *Int. J. Syst. Evol. Microbiol.* 64, 346–351. doi: 10.1099/ijs.0.059774-0
- King, S. J., Leigh, J. A., Heath, P. J., Luque, I., Tarradas, C., Dowson, C. G., et al. (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. doi: 10.1128/jcm.40.10.3671-3680.2002
- Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., and Phillippy, A. M. (2017). Canu: Scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* 27, 722–736. doi: 10.1101/gr.215087.116
- Leticun, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* 49, W293–W296. doi: 10.1093/nar/gkab301
- Liang, P., Wang, M., Gottschalk, M., Vela, A. I., Estrada, A. A., Wang, J., et al. (2021). Genomic and pathogenic investigations of *Streptococcus suis* serotype 7 population derived from a human patient and pigs. *Emerg. Microbes Infect.* 10, 1960–1974. doi: 10.1080/22221751.2021.1988725
- Liu, Z., Zheng, H., Gottschalk, M., Bai, X., Lan, R., Ji, S., et al. (2013). Development of multiplex PCR assays for the identification of the 33 serotypes of *Streptococcus suis*. *PLoS One* 8:e72070. doi: 10.1371/journal.pone.0072070
- Nicholson, T. L., and Shore, S. M. (2022). Genome Sequences of *Streptococcus suis* Isolates Obtained from Pigs in the United States between 2015 and 2017. *Microbiol. Resour. Announc.* 11:e0106721. doi: 10.1128/mra.01067-21
- Nicholson, T. L., Waack, U., Anderson, T. K., Bayles, D. O., Zaia, S. R., Goertz, I., et al. (2020). Comparative Virulence and Genomic Analysis of *Streptococcus suis* Isolates. *Front. Microbiol.* 11:620843. doi: 10.3389/fmicb.2020.620843
- Okura, M., Takamatsu, D., Maruyama, F., Nozawa, T., Nakagawa, I., Osaki, M., et al. (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. doi: 10.1128/AEM.03742-12
- Onishi, H., Sugawara, M., Okura, M., Osaki, M., and Takamatsu, D. (2012). Prevalence of *Streptococcus suis* genotypes in isolates from porcine endocarditis in East Japan. *J. Vet. Med. Sci.* 74, 1681–1684. doi: 10.1292/jvms.12-0301
- Pan, Z., Ma, J., Dong, W., Song, W., Wang, K., Lu, C., et al. (2015). Novel variant serotype of *Streptococcus suis* isolated from piglets with meningitis. *Appl. Environ. Microbiol.* 81, 976–985. doi: 10.1128/AEM.02962-14
- Princivalli, M. S., Palmieri, C., Magi, G., Vignaroli, C., Manzin, A., Camporese, A., et al. (2009). Genetic diversity of *Streptococcus suis* clinical isolates from pigs and humans in Italy (2003–2007). *Euro Surveill.* 14:19310. doi: 10.2807/esc.14.33.19310-en
- Qiu, X., Bai, X., Lan, R., Zheng, H., and 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
- Richter, M., and Rossello-Mora, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* 106, 19126–19131. doi: 10.1073/pnas.0906412106
- Roy, D., Athey, T. B. T., Auger, J. P., Goyette-Desjardins, G., Van Calsteren, M. R., Takamatsu, D., et al. (2017). A single amino acid polymorphism in the glycosyltransferase CpsK defines four *Streptococcus suis* serotypes. *Sci. Rep.* 7:4066. doi: 10.1038/s41598-017-04403-3
- Rusconi, B., Sanjar, F., Koenig, S. S., Mammel, M. K., Tarr, P. I., and Eppinger, M. (2016). Whole Genome Sequencing for Genomics-Guided Investigations of *Escherichia coli* O157:H7 Outbreaks. *Front. Microbiol.* 7:985. doi: 10.3389/fmicb.2016.00985
- Sahibzada, S., Abraham, S., Coombs, G. W., Pang, S., Hernandez-Jover, M., Jordan, D., et al. (2017). Transmission of highly virulent community-associated MRSA ST93 and livestock-associated MRSA ST398 between humans and pigs in Australia. *Sci. Rep.* 7:5273. doi: 10.1038/s41598-017-04789-0
- Saitou, N., and Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425. doi: 10.1093/oxfordjournals.molbev.a040454
- Scherrer, S., Rosato, G., Spoerry Serrano, N., Stevens, M. J. A., Rademacher, F., Schrenzel, J., et al. (2020). Population structure, genetic diversity and pathotypes of *Streptococcus suis* isolated during the last 13 years from diseased pigs in Switzerland. *Vet. Res.* 51:85. doi: 10.1186/s13567-020-00813-w
- Schultz, C., Jansen, E., Keijzers, W., Rothkamp, A., Duim, B., Wagenaar, J. A., et al. (2012). Differences in the population structure of invasive *Streptococcus suis* strains isolated from pigs and from humans in The Netherlands. *PLoS One* 7:e33854. doi: 10.1371/journal.pone.0033854
- Segura, M., Fittipaldi, N., Calzas, C., and Gottschalk, M. (2017). Critical *Streptococcus suis* Virulence Factors: Are They All Really Critical? *Trends Microbiol.* 25, 585–599. doi: 10.1016/j.tim.2017.02.005
- Segura, M., Zheng, H., de Greeff, A., Gao, G. F., Grenier, D., Jiang, Y., et al. (2014a). Latest developments on *Streptococcus suis*: An emerging zoonotic pathogen: Part 1. *Future Microbiol.* 9, 441–444. doi: 10.2217/fmb.14.14
- Segura, M., Zheng, H., de Greeff, A., Gao, G. F., Grenier, D., Jiang, Y., et al. (2014b). Latest developments on *Streptococcus suis*: An emerging zoonotic pathogen: Part 2. *Future Microbiol.* 9, 587–591. doi: 10.2217/fmb.14.15
- Shelyakin, P. V., Bochkareva, O. O., Karan, A. A., and Gelfand, M. S. (2019). Microevolution of three *Streptococcus suis* species: Selection, antigenic variation, and horizontal gene inflow. *BMC Evol. Biol.* 19:83. doi: 10.1186/s12862-019-1403-6
- Silva, L. M., Baums, C. G., Rehm, T., Wisselink, H. J., Goethe, R., and Valentin-Weigand, P. (2006). Virulence-associated gene profiling of *Streptococcus suis* isolates by PCR. *Vet. Microbiol.* 115, 117–127. doi: 10.1016/j.vetmic.2005.12.013
- Smith, H. E., Reek, F. H., Vecht, U., Gielkens, A. L., and Smits, M. A. (1993). Repeats in an extracellular protein of weakly pathogenic strains of *Streptococcus suis* type 2 are absent in pathogenic strains. *Infect. Immun.* 61, 3318–3326. doi: 10.1128/iai.61.8.3318-3326.1993
- Soto-Jimenez, L. M., Estrada, K., and Sanchez-Flores, A. (2014). GARM: Genome assembly, reconciliation and merging pipeline. *Curr. Top. Med. Chem.* 14, 418–424. doi: 10.2174/1568026613666131204110628
- Tang, J., Wang, C., Feng, Y., Yang, W., Song, H., Chen, Z., et al. (2006). Streptococcal toxic shock syndrome caused by *Streptococcus suis* serotype 2. *PLoS Med.* 3:e151. doi: 10.1371/journal.pmed.0030151
- Tatusova, T., DiCuccio, M., Badretdin, A., Chetvernin, V., Nawrocki, E. P., Zaslavsky, L., et al. (2016). NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 44, 6614–6624. doi: 10.1093/nar/gkw569
- Van Calsteren, M. R., Goyette-Desjardins, G., Gagnon, F., Okura, M., Takamatsu, D., Roy, R., et al. (2016). Explaining the Serological Characteristics of *Streptococcus suis* Serotypes 1 and 1/2 from Their Capsular Polysaccharide Structure and Biosynthesis. *J. Biol. Chem.* 291, 8387–8398. doi: 10.1074/jbc.M115.700716
- Vecht, U., Wisselink, H. J., Jellema, M. L., and Smith, H. E. (1991). Identification of two proteins associated with virulence of *Streptococcus suis* type 2. *Infect. Immun.* 59, 3156–3162.
- Votsch, D., Willenborg, M., Weldearegay, Y. B., and Valentin-Weigand, P. (2018). *Streptococcus suis* - The “Two Faces” of a Pathobiont in the Porcine Respiratory Tract. *Front. Microbiol.* 9:480. doi: 10.3389/fmicb.2018.00480
- Wang, S., Gao, M., An, T., Liu, Y., Jin, J., Wang, G., et al. (2015). Genetic diversity and virulence of novel sequence types of *Streptococcus suis* from diseased and healthy pigs in China. *Front. Microbiol.* 6:173. doi: 10.3389/fmicb.2015.00173
- Wang, Y., Zhang, W., Wu, Z., and Lu, C. (2011). Reduced virulence is an important characteristic of biofilm infection of *Streptococcus suis*. *FEMS Microbiol. Lett.* 316, 36–43. doi: 10.1111/j.1574-6968.2010.02189.x
- Weinert, L. A., Chaudhuri, R. R., Wang, J., Peters, S. E., Corander, J., Jombart, T., et al. (2015). Genomic signatures of human and animal disease in the zoonotic pathogen *Streptococcus suis*. *Nat. Commun.* 6:6740. doi: 10.1038/ncomms7740

- Wertheim, H. F., Nghia, H. D., Taylor, W., and Schultz, C. (2009a). *Streptococcus suis*: An emerging human pathogen. *Clin. Infect. Dis.* 48, 617–625. doi: 10.1086/596763
- Wertheim, H. F., Nguyen, H. N., Taylor, W., Lien, T. T., Ngo, H. T., Nguyen, T. Q., et al. (2009b). *Streptococcus suis*, an important cause of adult bacterial meningitis in northern Vietnam. *PLoS One* 4:e5973. doi: 10.1371/journal.pone.0005973
- Wick, R. R., Judd, L. M., Gorrie, C. L., and Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput. Biol.* 13:e1005595. doi: 10.1371/journal.pcbi.1005595
- Wileman, T. M., Weinert, L. A., Howell, K. J., Wang, J., Peters, S. E., Williamson, S. M., et al. (2019). Pathotyping the Zoonotic Pathogen *Streptococcus suis*: Novel Genetic Markers To Differentiate Invasive Disease-Associated Isolates from Non-Disease-Associated Isolates from England and Wales. *J. Clin. Microbiol.* 57, e1712–e1718. doi: 10.1128/JCM.01712-18
- Wisselink, H. J., Smith, H. E., Stockhofe-Zurwieden, N., Peperkamp, K., and Vecht, U. (2000). Distribution of capsular types and production of muramidase-released protein (MRP) and extracellular factor (EF) of *Streptococcus suis* strains isolated from diseased pigs in seven European countries. *Vet. Microbiol.* 74, 237–248. doi: 10.1016/s0378-1135(00)00188-7
- Yu, H., Jing, H., Chen, Z., Zheng, H., Zhu, X., Wang, H., et al. (2006). Human *Streptococcus suis* outbreak, Sichuan, China. *Emerg. Infect. Dis.* 12, 914–920. doi: 10.3201/eid1206.051194
- Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012). Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–2644. doi: 10.1093/jac/dks261
- Zheng, H., Ji, S., Lan, R., Liu, Z., Bai, X., Zhang, W., et al. (2014). Population analysis of *Streptococcus suis* isolates from slaughtered swine by use of minimum core genome sequence typing. *J. Clin. Microbiol.* 52, 3568–3572. doi: 10.1128/JCM.00536-14
- Zheng, H., Qiu, X., Roy, D., Segura, M., Du, P., Xu, J., et al. (2017). Genotyping and investigating capsular polysaccharide synthesis gene loci of non-serotypeable *Streptococcus suis* isolated from diseased pigs in Canada. *Vet. Res.* 48:10. doi: 10.1186/s13567-017-0417-6
- Zhu, W., Wu, C., Sun, X., Zhang, A., Zhu, J., Hua, Y., et al. (2013). Characterization of *Streptococcus suis* serotype 2 isolates from China. *Vet. Microbiol.* 166, 527–534. doi: 10.1016/j.vetmic.2013.06.009