Veterinary Microbiology 284 (2023) 109839

Contents lists available at ScienceDirect

Veterinary Microbiology

journal homepage: www.elsevier.com/locate/vetmic

Genome characteristics related to the virulence of *Streptococcus suis* in Swedish pigs

Anna Werinder [a,*], Anna Aspán [b], Magdalena Jacobson [a], Annette Backhans [c], Marie Sjölund [a,c], Bengt Guss [d], Robert Söderlund [b]

[a] Swedish University of Agricultural Sciences (SLU), Department of Clinical Sciences, Box 7054, 750 07 Uppsala, Sweden
[b] National Veterinary Institute (SVA), Department of Microbiology, 751 89 Uppsala, Sweden
[c] National Veterinary Institute (SVA), Department of Animal Health and Antimicrobial Strategies, 751 89 Uppsala, Sweden
[d] Swedish University of Agricultural Sciences (SLU), Department of Biomedical Science and Veterinary Public Health, Box 7036, 750 07 Uppsala, Sweden

ARTICLE INFO

Keywords:
GWAS
Meningitis
MLST
Streptococcus suis
Virulence-associated genes
Whole-genome sequencing
Zoonosis

ABSTRACT

The impact of *S. suis* on Swedish pig production has increased in recent years, and characterization of the strains present in the pig population is needed to aid in surveillance and prevention. Therefore, the aim of this study was to identify and characterize differences in the genomes between Swedish *S. suis* isolates associated with disease and isolates from healthy animals. Isolates categorized as being pathogenic (n = 100) or non-pathogenic (n = 117) were whole-genome sequenced, serotyped *in silico*, and sequence-typed using traditional MLST and core-genome MLST, and a genome-wide association study was performed to identify virulence-associated genes. In decreasing order, serotypes 2, 1, and 7 were the most common in the pathogenic group, and serotypes 15 and 12 were the most common in the non-pathogenic group. Among the commonly disease-associated sequence types, ST28 and ST25 were identified, whereas ST1 was scarcely found. The majority of isolates belonged to novel sequence types, revealing differences between Swedish isolates and those reported from other countries. The genomes of the pathogenic isolates were on average smaller and less heterogenic as compared to those of the non-pathogenic isolates. Although a majority of the previously published virulence-associated genes included in the study were found in the genomes of both pathogenic and non-pathogenic isolates, several new, significantly virulence-associated genes were identified.

1. Introduction

Streptococcus (S.) suis is an important bacterial pathogen that may cause arthritis, meningitis, and pneumonia in both pigs and humans (Gottschalk and Segura, 2019). There is a marked phenotypic and genotypic diversity within the species, and 29 serotypes of varying virulence are currently recognized (Okura et al., 2016). The prevalence of *S. suis* in pigs may approach 100%, with one or several serotypes carried predominantly in the tonsils without causing any obvious clinical signs (Werinder et al., 2020). Serotypes 1/2, 2, 3, 7, and 9 are most often associated with disease in pigs, although there are geographical variations in the relative frequencies (Goyette-Desjardins et al., 2014; Segura et al., 2020). However, while certain serotypes are over-represented among clinical cases, the serotype of an isolate is not a definitive indicator of virulence. Thus, further characterization of pathogenic strains is needed.

Whole-genome sequencing (WGS) has been increasingly used to find reliable markers of virulence in *S. suis* genomes. In addition to the well-studied genes *mrp*, *epf*, and *sly*, many new, virulence-associated genes have been proposed, but so far none have been established as consistently required for pathogenicity (Fittipaldi et al., 2012; Segura et al., 2017; Estrada et al., 2021, 2022). Further, seven-gene multilocus sequence typing (MLST) and core genome MLST (cgMLST) may be used to make highly discriminatory comparisons of genomes, and certain sequence types (STs) are reported to be associated with virulence (King et al., 2002; Segura et al., 2020). Additionally, a reduction of the genome size has been associated with pathogenicity in *S. suis* (Murray et al., 2020).

The impact of *S. suis* on Swedish pig production has increased in recent years, and characterization of the strains present in the pig population is vital to developing effective approaches to the diagnosis, treatment, and prevention of the disease. Therefore, the aim of this study

* Corresponding author.
E-mail address: anna.werinder@slu.se (A. Werinder).

https://doi.org/10.1016/j.vetmic.2023.109839
Received 24 March 2023; Received in revised form 12 June 2023; Accepted 27 July 2023
Available online 28 July 2023

was to identify and characterize differences in the genomes between *S. suis* isolates associated with disease and isolates from healthy animals in Sweden, using WGS.

2. Materials and methods

2.1. Ethics statement

Ethical approval for tonsil swab sampling of live animals and euthanasia of diseased pigs for necropsy was granted by the Ethics Committee for Animal Experimentation, Uppsala, Sweden (Dnr 5.8.18–15404), and informed consent was obtained from the farm owners.

2.2. Bacterial isolates

A total of 217 *S. suis* isolates were included in the study and categorized as "pathogenic" or "non-pathogenic" (Supplementary file 1). The pathogenic isolates (n = 100) originated from affected tissues of pigs of varying ages, diagnosed with disease caused by *S. suis*. The isolates were obtained from clinical submissions made to the Swedish National Veterinary Institute during 1985–2020, and to the Department of Pathology at the Swedish University of Agricultural Sciences during 2018–2020. The non-pathogenic isolates (n = 117) were cultured from the tonsils of the soft palate of healthy grower pigs from 20 commercial pig farms in Sweden during 2018–2019, as previously described (Werinder et al., 2020). All isolates included in the study were determined to be *S. suis* by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) using a Microflex LT system (Bruker Daltonik GmbH) with a cut-off score of 2.00.

2.3. Whole-genome sequencing

DNA was extracted using protease K and lysozyme before purification using an EZ1 Advanced XL robotic instrument (Qiagen Inc.) with an EZ1 DNA tissue kit (Qiagen Inc.), as previously described (Werinder et al., 2021). Whole-genome sequencing was performed using an Illumina NextSeq 500 sequencing system (Illumina, Inc.) with the Nextera XT DNA library preparation kit (Illumina, Inc.), and the NextSeq 500/550 mid-output kit v. 2.5 (lllumina, Inc.) to sequence 2 × 150 bp paired-end reads to a minimum coverage of 25 × . The sequence data used in this study are available in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB45445.

The reads were quality-checked with FastQC v.0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and trimmed with Trimmomatic v.0.39 (Bolger et al., 2014) using the sliding window option. Draft genomes were assembled using SPAdes v.3.14.1 (Bankevich et al., 2012) with the "careful" option, and evaluated using Quast v.5.0.2 (Gurevich et al., 2013). Correction was performed using Bowtie2 v.2.4.1 (Langmead and Salzberg, 2012) with Pilon v.1.23 (Walker et al., 2014). Contigs of > 500 bp in length and with a kmer coverage of > 10 × were kept for further analysis.

2.4. Species confirmation

Species confirmation was performed using PYANI v.0.2.10 (Pritchard et al., 2016) to ascertain that all genomes had an average nucleotide identity (ANI) of ≥ 94% to the well-annotated genome *S. suis* SC84 [NCBI accession number NC_012924.1] (Werinder et al., 2021). Additionally, BLAST+ v.2.10.1 (Camacho et al., 2009) with a cut-off of 95% nucleotide identity was used to confirm the presence of a 16S rRNA gene with ≥ 98.7% sequence similarity to the *S. suis* 16S rRNA gene, and the presence of a *recN* gene with ≥ 95% sequence similarity to the full 1, 662-bp sequence of the *recN* gene from a serotype 2 strain (Glazunova et al., 2010).

2.5. Genome size

The pan-genome, *i.e.* the entire set of genes present in the 217 isolates, was calculated using Roary v.3.13.0 (Page et al., 2015), with a 90% BLASTp minimum percentage identity cut-off, and the output was used to calculate the number of genes present in the genomes. Quast was used to calculate genome sizes. Results were compared using two-tailed Welch's two-sample t-tests in R v.4.1.2 (R Core Team 2021, https://www.R-project.org/) with RStudio v.2022.02.3 (http://www.rstudio.com/). A value of $P < 0.05$ was considered significant.

2.6. Serotyping

Serotyping of the 217 isolates was done *in silico* using the pipeline described by Athey et al. (2016). To confirm the differentiation between serotypes 1 and 14, and serotypes 2 and 1/2, that differ at a single position in the capsular gene *cpsK* only, the *cpsK* gene sequences were extracted using BLAST+, translated into amino acids, and manually inspected after multiple alignment with MUSCLE in the software MEGA X (Kumar et al., 2018). Isolates initially identified as serotype 2 by the pipeline, but with cysteine (C) instead of tryptophan (W) at position 161, were determined to be serotype 1/2. Similarly, isolates initially identified as serotype 14, that had cysteine instead of tryptophan at position 161, were determined to be serotype 1.

2.7. Multilocus sequence typing

Traditional seven-gene MLST, using fragments of the genes *cpn60*, *dpr*, *recA*, *aroA*, *thrA*, *gki*, and *mutS* (King et al., 2002), was performed using FastMLST v.0.0.15 (Guerrero-Araya et al., 2021) with the PubMLST database (https://pubmlst.org/) containing 1962 STs. Hierarchical clustering of cgMLST (HierCC) was performed using Enterobase (https://enterobase.warwick.ac.uk/) with the cgMLSTv1 scheme for streptococci (Zhou et al., 2021). A minimum-spanning tree was then generated using the MSTreeV2 algorithm and annotated using GrapeTree (Zhou et al., 2018).

2.8. Virulence-associated genes

To identify genes associated with virulence, the draft assemblies were first annotated with Prokka v.1.14.6 (Seemann, 2014) using default settings and with the *Streptococcus suis* SC84 genome used as reference. Analysis of the pan-genome was done using Roary with a 90% BLASTp identity cut-off. Scoary v.1.6.16 (Brynildsrud et al., 2016) was used to identify genes that were significantly more frequently observed in pathogenic isolates (Benjamini-Hochberg adjusted $P < 0.05$), as compared to non-pathogenic isolates. Genes that were present in ≥ 75% of pathogenic isolates, and missing from ≥ 75% of the non-pathogenic isolates, were selected for further study. Additionally, the presence of previously proposed virulence-associated genes, including the commonly described genes *epf*, *mrp*, and *sly*, was investigated using BLAST+ with cut-offs of 90% sequence identity and 60% coverage.

P-values for the gene profiles were calculated using Fisher's exact test in R with RStudio. The presence or absence of the new, as well as the previously published, virulence-associated genes was visualized in a heatmap using R with RStudio and the package pheatmap (v. 1.0.12; https://CRAN.R-project.org/package=pheatmap).

3. Results

3.1. Genome size

The isolates had an open pan-genome, meaning that the number of genes in the pan-genome increased as new isolates were sequenced and added to the analysis. The core genome consisted of 903 genes that were present in ≥ 99% of the isolates, and the accessory genome included

13,618 genes (Supplementary figure 1). The core genome of the pathogenic isolates consisted of 1054 genes, while the core genome of the non-pathogenic isolates consisted of 1010 genes.

The genomes of the pathogenic isolates were on average smaller and contained fewer genes than the genomes of the non-pathogenic isolates (Fig. 1 A and B). The mean genome size of the pathogenic isolates was 2.14 Mbp (SD 0.10), while the mean genome size of the non-pathogenic isolates was 2.30 Mbp (SD 0.12). The difference between the groups was 0.16 Mbp ($P < 0.001$, 95% CI 0.14–0.20). Similarly, the mean number of genes in the pathogenic isolates was 2007 (SD 88.6), while the mean number of genes in the non-pathogenic isolates was 2119 (SD 108.7). Thus, the non-pathogenic isolates contained on average 112 more genes ($P < 0.001$, 95% CI 85.81–138.63).

The mean GC content also differed between the groups ($P < 0.001$, 95% CI 0.07–0.16). The mean was 41.22% (SD 0.20) in the pathogenic isolates, as compared to 41.34% (SD 0.16) in the non-pathogenic isolates.

3.2. Serotyping

The serotyping pipeline initially classified a total of 37 isolates as serotype 2 and none as serotype 1/2. A multiple alignment of the extracted *cpsK* genes translated into amino acids nevertheless revealed differences in the amino acid at position 161, which indicated that three isolates instead belonged to serotype 1/2. Similarly, the pipeline classified 25 serotypes as serotype 14 and none as serotype 1. Manual inspection of the amino acid sequences revealed that 24 isolates were identical to the serotype 1 reference at position 161, while only one isolate was identified as serotype 14. Thus, a total of 25 serotypes were identified among the 217 isolates (Fig. 1 C).

The most abundant serotypes in the pathogenic group, in decreasing order, were 2, 1, and 7, which constituted 24% (24/100), 34% (34/100), and 12% (12/100) of the pathogenic isolates, respectively. The most abundant serotypes in the non-pathogenic group, in decreasing order, were serotypes 15 and 12, constituting 13% (15/117) and 10% (12/117) of the isolates, respectively. Eleven serotypes; 3, 5, 9, 11, 16, 21, 23, 28, 29, 30, and 31, were present in both the pathogenic and non-pathogenic groups. Four of the pathogenic isolates (4%, 4/100) were non-typeable using the *in silico* pipeline, in contrast to the non-pathogenic group, where 40 isolates (34%, 40/117) were non-typeable.

3.3. Multilocus sequence typing

Seven-gene MLST revealed that 83 isolates belonged to one of 13 previously known sequence types; ST1, ST13, ST25, ST28, ST29, ST38, ST55, ST87, ST117, ST483, ST911, ST1112, and ST1805. The most abundant types were ST13, ST28, and ST25. Another 29 isolates contained known alleles in 20 novel combinations, that were submitted to the PubMLST database and assigned permanent ST numbers. In a further 102 isolates, novel alleles of *cpn60*, *dpr*, *recA*, *aroA*, *thrA*, *gki*, or *mutS* were present (Table 1). The alleles were also submitted to PubMLST and assigned permanent allele numbers, after which the resulting novel profiles received permanent ST numbers, resulting in a further 70 novel STs discovered.

Two isolates, non-serotypeable and originating from heart lesions of two diseased pigs, completely lacked an *aroA* allele and could therefore not receive an ST. The isolates had 5 out of 7 alleles that were identical to each other and to another isolate present in the database. That isolate also lacked *aroA* and originated from the heart of a diseased pig in North America.

According to the cgMLST analysis, the 217 isolates included in the study were separated into a total of 199 different sequence types, all of which were novel (Supplementary file 1). The proportions of unique cgSTs, 92% (92/100) and 91% (107/117), respectively, were similar in the pathogenic and the non-pathogenic group. In the pathogenic group, only five STs were shared by multiple isolates; five isolates were cgST-61734, while two isolates each were cgST-61724, cgST-61729, cgST-61748, and cgST-61750. Among the non-pathogenic isolates, seven STs were shared by more than one isolate, with three isolates each of cgST-61832, cgST-61859, and cgST-61897, and two isolates each of cgST- 61812, cgST-61827, and cgST-61862.

The hierarchical cgMLST clustering analysis (HierCC) confirmed the species identity by placing all isolates in the HierCC cluster HC363_1 (*S. suis*). The 217 isolates belonged to 81 distinct HC100 clusters, or cgMLST eBurstGroups (ceBGs). The pathogenic isolates belonged to 17 separate ceBGs, while the non-pathogenic isolates belonged to 70 ceBGs, thus showing a greater diversity. The main ceBG identified was group 1, which contained 54 isolates, all but one originating from the pathogenic group. These isolates belonged to serotypes 2, 1/2, 3, and 7, and three isolates were non-typeable. The second largest ceBG, number 237, contained 25 isolates, all of which belonged to the pathogenic group and were serotype 1 or 14. With few exceptions, the pathogenic isolates clustered together in a minimum spanning tree constructed from

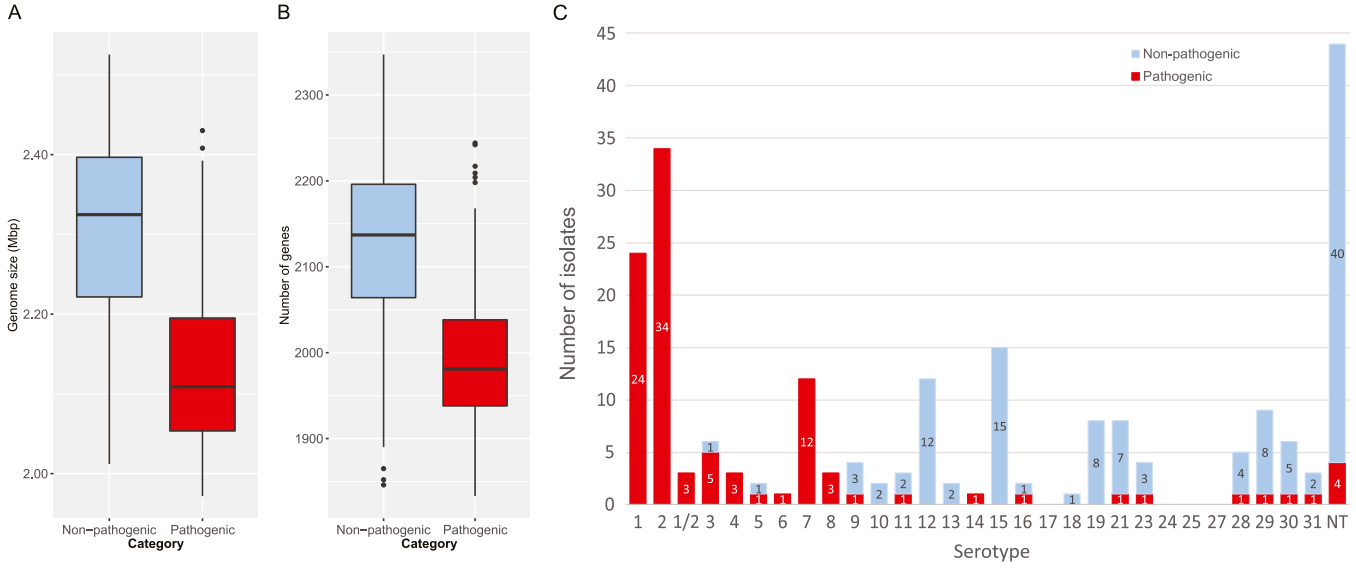

Fig. 1. (A) Comparison of total genome size between pathogenic (red) and non-pathogenic (blue) Swedish *Streptococcus suis* isolates. (B) Comparison of the number of genes contained within the genomes of pathogenic and non-pathogenic isolates. (C) Serotypes of 217 Swedish *S. suis* isolates, as determined by *in silico* serotyping.

Table 1
Results from multilocus sequence typing (MLST) analysis of 217 Swedish *Streptococcus suis* isolates.

Serotype	ST[a] (newly assigned numbers)	Number of isolates
1	13	24
2	1, 25, 28, 1805, novel STs (1999, 2000)	34
1/2	28	3
3	117, novel ST (2043)	6
4	911	3
5	Novel STs (1993, 2048)	2
6	55	1
7	29	12
8	87, novel ST (1976)	3
9	Novel STs (1966, 1995, 2033)	4
10	Novel STs (2022, 2039)	2
11	Novel STs (1964, 1965, 2006)	3
12	Novel STs (1984, 1985, 1986, 2012, 2047)	12
13	Novel STs (1992, 2002)	2
14	13	1
15	1112, novel STs (1963, 1967, 1968, 1969, 1970, 1978, 1980, 1982, 2007, 2017, 2028, 2019, 2030)	15
16	Novel STs (1974, 1991)	2
17	-	0
18	Novel ST (1990)	1
19	Novel STs (1975, 1979, 2005, 2008)	8
21	Novel STs (1973, 1987, 2018, 2038, 2042)	8
23	483, novel STs (2040, 2044)	4
24	-	0
25	-	0
27	-	0
28	38, novel STs (1996, 1997, 2013)	5
29	Novel STs (1977, 1983, 1994, 2004, 2031, 2036)	9
30	Novel STs (2009, 2010, 2023, 2025, 2050)	6
31	Novel STs (1971, 1972, 2052)	3
Non-typeable	28, novel STs (1981, 1984, 1988, 1989, 1998, 2001, 2003, 2011, 2015, 2016, 2019, 2020, 2021, 2024, 2026, 2027, 2032, 2034, 2035, 2037, 2041, 2045, 2049, 2051), non-typeable	44
Total		**217**

[a] ST = sequence type

cgMLST data. Isolates of the same serotypes or organs of isolation clustered together to a lesser extent. However, isolates originating from the brain, heart, and lungs, respectively, clustered near each other (Fig. 2).

3.4. Virulence-associated genes

The well-studied virulence-associated genes *epf*, *mrp*, and *sly* were found in 1.8%, 42.9%, and 28.1% of the isolates, respectively. *Mrp* and *sly* were significantly more often present in the pathogenic isolates than in the non-pathogenic isolates, while no difference was found between the pathogenic and non-pathogenic groups regarding the presence of *epf*. The genotypic profiles *epf*−/*mrp*+/*sly*− and *epf*−/*mrp*−/*sly*+ were significantly more often present in the pathogenic isolates, while the profile *epf*−/*mrp*−/*sly*− was significantly more common in the non-pathogenic isolates (Table 2). Among the serotype 2 isolates, 97% (33/34) of the isolates were *mrp*+, and 88% (30/34) had the genotypic profile *epf*−/*mrp*+/*sly*−. Among the 25 isolates belonging to serotype 1 or 14, all had the genotype *epf*−/*mrp*−/*sly*+.

The recently proposed genotypic profile of *SSU_RS09525*+/*SSU_RS09155*+/*SSU_RS03100*+, published in a recent study on North American isolates (Estrada et al., 2022), was significantly more often present in the pathogenic isolates (86%, 86/100) than in the non-pathogenic isolates (5.1%, 6/117) in the present study (Table 2). Out of the isolates containing the profile, 93.4% (86 out of 92) were pathogenic.

The genome-wide association study indicated 34 genes whose presence differed significantly between pathogenic and non-pathogenic isolates ($P < 0.05$), and that were present in ≥ 75% of the pathogenic isolates and absent from ≥ 75% of the non-pathogenic isolates. Of these, two were identified as *srtC1* (sortase) and *lacG_1* (hypothetical protein), respectively, and 32 as genes encoding hypothetical proteins without annotation. The presence/absence of the identified genes is visualized in a heatmap. The majority of the previously published virulence-associated genes were present in most genomes, pathogenic as well as non-pathogenic (Fig. 3).

4. Discussion

The results from this study showed that certain differences between pathogenic and non-pathogenic isolates could be identified. Differences in the genomes regarding genome size and gene content were demonstrated, and based on cgMLST analysis, pathogenic isolates predominantly clustered together, while there was a higher phylogenetic diversity among the non-pathogenic isolates. The majority of the previously published, virulence-associated genes were identified in both pathogenic and non-pathogenic genomes and were not useful as markers of virulence in the Swedish *S. suis* isolates. However, new candidate genes were identified which were significantly more often present in pathogenic as compared to non-pathogenic isolates, and warrant further studies.

4.1. Genome size

The pan-genome of the Swedish isolates was open, *i.e.* the number of accessory genes increased as more genomes were sequenced and added to the analysis, which is consistent with the findings of previous genomic studies on this highly diverse species (Murray et al., 2020; Estrada et al., 2022).

Analysis of the present isolates revealed a core genome of 903 genes, which is fewer as compared to results from studies in North America and China (Zhang et al., 2011; Estrada et al., 2022), although differences in sequencing strategy may have an impact on the results. A smaller core genome indicates heterogeneity among the present isolates, despite the use of strict species-identification criteria. There was no obvious difference in the number of core genes between the pathogenic and non-pathogenic isolates, which contained 1054 and 1010 genes, respectively.

The mean total genome size of the pathogenic isolates was smaller as compared to the non-pathogenic isolates, indicating that genome reduction had taken place in the former group, or expansion in the latter. Thus, the present results support previous studies which have established a smaller genome as a feature of pathogenic *S. suis* (Weinert et al., 2015; Murray et al., 2020). The difference in GC content between the groups, although statistically significant, was small and well within the range that may occur within a species (Rosselló-Mora and Amann, 2001).

4.2. Serotypes

The most common serotypes among the pathogenic isolates in this study were, in decreasing order, serotypes 2, 1, and 7. This is consistent with the dominance of serotype 2 and prominence of serotype 7 among disease-associated isolates in both pigs and humans reported globally, but serotype 1 appears more common in Swedish pigs than in pig in many other countries (Goyette-Desjardins et al., 2014). The possibly opportunistic role of *S. suis* in respiratory disease makes the classification of isolates originating from lung tissue challenging, and may partly explain the presence of certain serotypes in both the pathogenic and non-pathogenic groups. In our hands, the serotyping pipeline initially misidentified the serotype 1 isolates as serotype 14, which is notable, but may be an issue related to software compatibility.

In notable contrast to other European countries where serotype 9 is the most commonly isolated serotype from diseased pigs (Schultsz et al.,

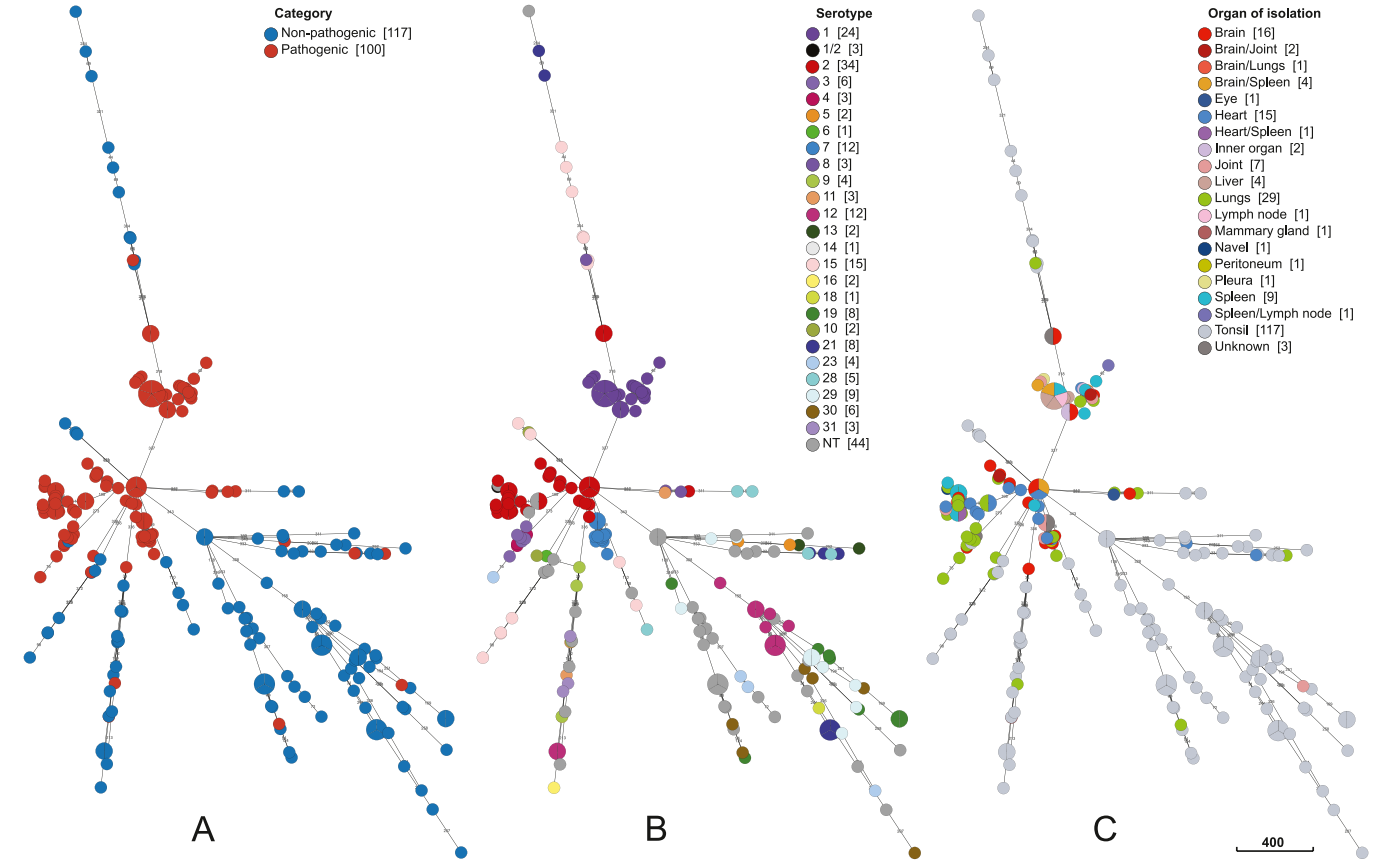

Fig. 2. Minimum spanning tree of core-genome multi-locus sequence typing (cgMLST) data from 217 *Streptococcus suis* isolates from healthy and diseased Swedish pigs. The same tree is annotated according to (A) pathogenicity; (B) organ of isolation; and (C) serotype.

Table 2
The presence of the well-studied genes *epf*, *mrp*, and *sly*, and the previously proposed virulence-associated genes *SSU_RS09525*, *SSU_RS09155*, *SSU_RS03100*, and genotypic profiles in pathogenic and non-pathogenic Swedish *Streptococcus suis* isolates. A significant difference was considered if $P < 0.05$.

Virulence-associated gene or genotype	Total isolates (n = 217)	Pathogenic isolates (n = 100)	Non-pathogenic isolates (n = 117)	P-value
epf	4 (1.8%)	2 (2.0%)	2 (1.7%)	1
mrp	93 (42.9%)	64 (64.0%)	29 (24.8%)	< 0.001
sly	61 (28.1%)	37 (37.0%)	24 (20.5%)	< 0.001
epf+/mrp+/sly+	4 (1.8%)	2 (2.0%)	2 (1.7%)	1
epf−/mrp+/sly−	57 (26.3%)	52 (52.0%)	5 (4.3%)	< 0.001
epf−/mrp+/sly+	32 (14.7%)	10 (10.0%)	22 (18.8%)	0.1
epf−/mrp−/sly−	99 (45.6%)	11 (11.0%)	88 (75.2%)	< 0.001
epf−/mrp−/sly+	25 (11.5%)	25 (25.0%)	0 (0%)	< 0.001
epf+/mrp+/sly−	0 (0%)	0 (0%)	0 (0%)	na
epf+/mrp−/sly+	0 (0%)	0 (0%)	0 (0%)	na
epf+/mrp−/sly−	0 (0%)	0 (0%)	0 (0%)	na
SSU_RS09525	162 (74.7%)	95 (95%)	67 (57.3%)	< 0.001
SSU_RS09155	98 (45.2%)	87 (87%)	11 (9.4%)	< 0.001
SSU_RS03100	106 (48.8%)	90 (90%)	16 (13.7%)	< 0.001
SSU_RS09525+/ SSU_RS09155+/ SSU_RS03100+	92 (42.4%)	86 (86%)	6 (5.1%)	< 0.001

2012), few isolates of serotype 9 were identified in the Swedish material, and three out of four isolates were in the non-pathogenic group. The shift in dominance towards serotype 9 in other European countries has been speculated to result from the use of vaccines or autogenous bacterins primarily targeting serotype 2 (Prüfer et al., 2019), a practice that has been very limited in Sweden. However, care should be taken when extrapolating the results to a wider population. Additionally, before 1995 not all of the currently recognized 29 serotypes had been described, and antisera for serotyping were thus not available.

4.3. Multilocus sequence typing analyses

Traditional MLST, using fragments of seven housekeeping genes; *cpn60*, *dpr*, *recA*, *aroA*, *thrA*, *gki*, and *mutS* (King et al., 2002), is a standard method of studying the structure of *S. suis* populations. MLST analysis revealed that several isolates of serotype 2 belonged to ST28 and ST25, a combination commonly associated with virulence primarily in North America and Asia (Segura et al., 2020). ST1 was identified in only 2% of the Swedish isolates, in contrast to the high prevalence reported in studies from other countries, such as Spain and the Netherlands, and in particular the UK (de Greeff et al., 2011; Schultsz et al., 2012; Wileman et al., 2019). Additionally, the large proportion (131/217) of isolates that were identified as new STs or contained new MLST alleles indicate genotypic differences between Swedish isolates and those from other countries. A possible reason for these differences in genetic structure is that the import of live pigs to Sweden is very restricted (G. Parsons, Swedish Board of Agriculture, personal communication, 16 November 2022).

Interestingly, two isolates that lacked the *aroA* gene fragment required for MLST were very similar to a North American isolate that also lacked this fragment, and all were isolated from heart lesions. One

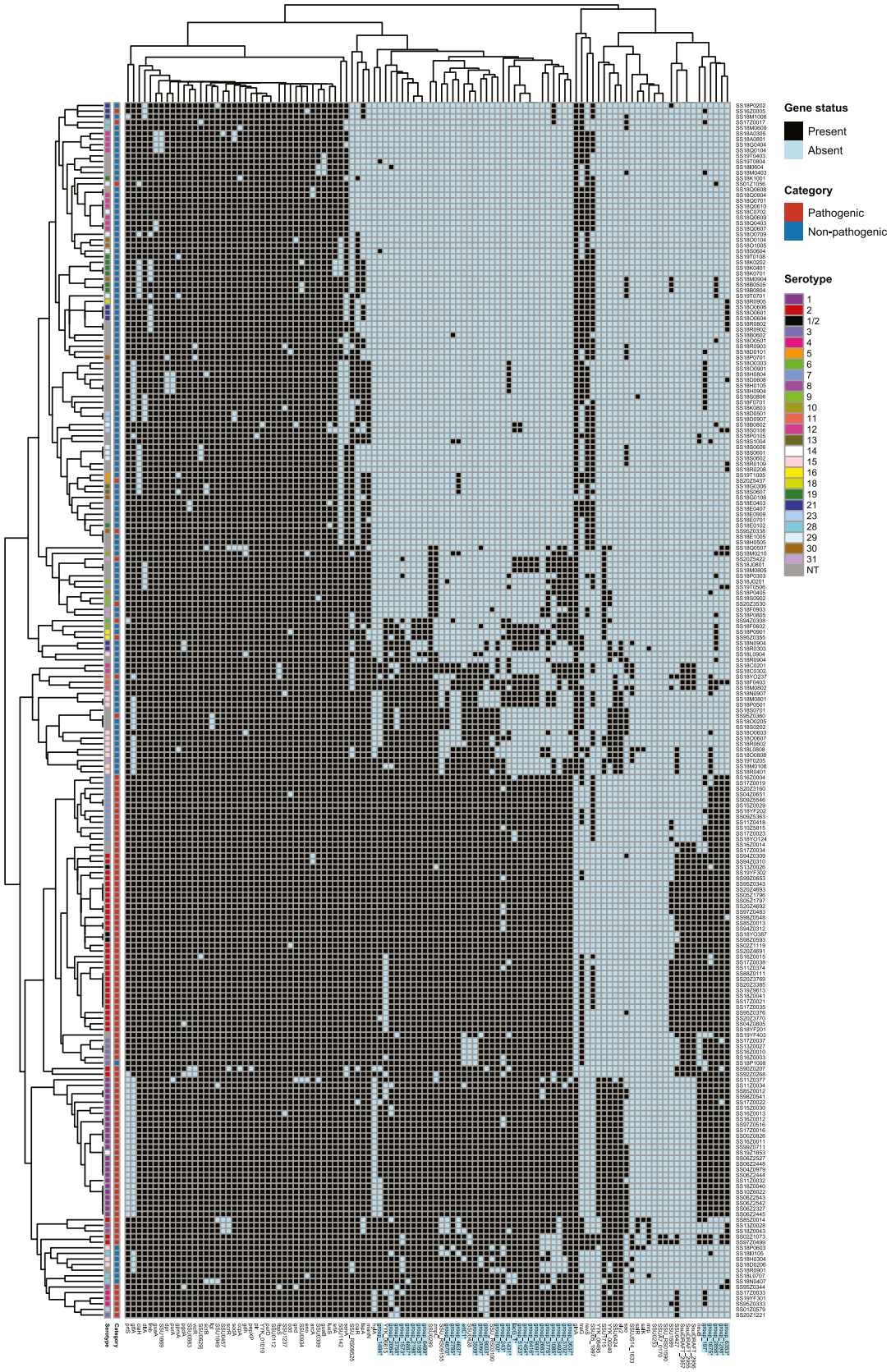

Fig. 3. Heatmap illustrating the presence (black) and absence (blue) of new and previously published putative virulence-associated genes in 100 pathogenic and 117 non-pathogenic *Streptococcus suis* isolates. Genes and isolates are clustered according to their profile similarities. Newly identified genes are marked blue on the x-axis, and isolate identities, annotated with pathogenic (red) and non-pathogenic (blue) and serotype (coloured according to the legend), are given on the y-axis.

of the present isolates originated from a farm that repeatedly suffered from *S. suis*-related endocarditis in the pigs.

A higher resolution can be achieved with cgMLST as compared to traditional MLST because of the greater number of genes included. The pathogenic isolates, with few exceptions, clustered together in a minimum spanning tree constructed from cgMLST data, suggesting that they are relatively closely related. The Swedish isolates did not match any cgSTs already present in Enterobase but clustered close to the other *S. suis* in the database. The HierCC clustering results showed that the isolates belonged to 81 ceBGs, with the pathogenic isolates represented in fewer groups than the non-pathogenic isolates. Clustering of isolates of the same serotype was also evident, but to a lesser extent, as was clustering based on the organ of isolation. The cgMLST scheme (cgMLSTv1) in place at Enterobase is currently composed of 372 loci from different species of streptococci. A defined *S. suis*-specific cgMLST scheme is not yet available, but would likely provide more detailed information regarding the population structure of this highly diverse species. However, traditional MLST still has an important role in the study of *S. suis*, particularly considering the steady expansion of the PubMLST database of STs.

4.4. Virulence-associated genes

The majority of the previously published virulence-associated genes were not associated with the pathogenic pathotype in this study, since they were either found in the majority of all isolates, or were absent from the majority of the isolates. *Mrp* and *sly* were more common in the pathogenic isolates, while *epf* was very uncommon in the material as a whole. Neither *mrp* nor *sly* were present in $\geq$ 75% of the pathogenic isolates in total, but *mrp* was present in 97% (33/34) of the serotype 2 isolates and *sly* in 100% (24/24) of the serotype 1 isolates.

The recently proposed virulence-associated genotypic profile of *SSU_RS09525+/SSU_RS09155+/SSU_RS03100+*, however, was present in 86% of the pathogenic isolates, and 5.1% of the non-pathogenic isolates in the present study. While the presence of the profile was lower than the 96.4% observed in the original study (Estrada et al., 2022), this indicates that the profile is potentially useful in Swedish isolates.

The other previously published virulence-associated genes were not significantly associated with virulence in the present material. Virulence is likely the result of mechanisms involving several genes or sets of genes, and putative virulence-associated genes may thus be clonal markers instead of being directly linked to virulence. The results show that virulence-associated genes discovered in isolates from limited geographical areas do not definitely apply to the general *S. suis* population and that studies from different geographical areas, as well as on a larger international scale, are necessary to broaden the understanding of virulence in *S. suis*.

5. Conclusion

Differences were found between the genomes of pathogenic and non-pathogenic *S. suis* isolates from Swedish pigs. The genomes of pathogenic isolates were on average smaller and less heterogenic as compared to those of non-pathogenic isolates. A majority of the previously published virulence-associated genes included in the study were found in both the pathogenic and the non-pathogenic genomes, and instead, a new set of significantly virulence-associated genes was identified.

Additional information

Ethics approval and consent to participate

Ethical approval for tonsil swab sampling of live animals and euthanasia of diseased pigs for necropsy was granted by the Ethics Committee for Animal Experimentation, Uppsala, Sweden [Dnr 5.8.18–15404], and informed consent was obtained from the farm owners.

Funding

This work was supported by FORMAS [grant number 2016–01118], the Swedish Pig Farmers' Research Foundation [grant number 201906], and the Ivar and Elsa Sandberg Foundation [2019-01-30]. The funding bodies had no role in the study design, collection of samples, analysis, interpretation, preparation of the manuscript, or the decision to submit the work for publication.

CRediT authorship contribution statement

All authors contributed to the design of the study. A.W. and M.J. carried out the sampling. A.W. performed the laboratory and bioinformatics work in collaboration with A.A. and R.S. A.W. wrote the manuscript with support from M.J and RS. A.A., A.B., M.S., and B.G. contributed to finalizing the manuscript. All authors read and approved the final manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are grateful to Boel Harbom and Sara Frosth for their kind assistance.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.vetmic.2023.109839.

References

Athey, T.B.T., Teatero, S., Lacouture, S., Takamatsu, D., Gottschalk, M., Fittipaldi, N., 2016. Determining *Streptococcus suis* serotype from short-read whole-genome sequencing data. BMC Microbiol 16, 162. https://doi.org/10.1186/s12866-016-0782-8.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brynildsrud, O., Bohlin, J., Scheffer, L., Eldholm, V., 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. Genome Biol. 17, 238. https://doi.org/10.1186/s13059-016-1108-8.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinform 10, 421. https://doi.org/10.1186/1471-2105-10-421.

de Greeff, A., Wisselink, H.J., de Bree, F.M., Schultsz, C., Baums, C.G., Thi, H.N., Stockhofe-Zurwieden, N., Smith, H.E., 2011. Genetic diversity of *Streptococcus suis* isolates as determined by comparative genome hybridization. BMC Microbiol 11, 161. https://doi.org/10.1186/1471-2180-11-161.

Estrada, A.A., Gottschalk, M., Gebhart, C.J., Marthaler, D.G., 2022. Comparative analysis of *Streptococcus suis* genomes identifies novel candidate virulence-associated genes in North American isolates. Vet. Res. 53, 23. https://doi.org/10.1186/s13567-022-01039-8.

Estrada, A.A., Gottschalk, M., Rendahl, A., Rossow, S., Marshall-Lund, L., Marthaler, D. G., Gebhart, C.J., 2021. Proposed virulence-associated genes of *Streptococcus suis* isolates from the United States serve as predictors of pathogenicity. Porc Health Manag 7, 22. https://doi.org/10.1186/s40813-021-00201-6.

Fittipaldi, N., Segura, M., Grenier, D., 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. https://doi.org/10.2217/fmb.11.149.

Glazunova, O.O., Raoult, D., Roux, V., 2010. Partial *recN* gene sequencing: a new tool for identification and phylogeny within the genus *Streptococcus*. Int J. Syst. Evol. Bacteriol. 60, 2140–2148. https://doi.org/10.1099/ijs.0.018176-0.

Gottschalk, M., Segura, M., 2019. Streptococcosis. In: Zimmerman, J.J. (Ed.), Diseases of Swine. Hoboken, NJ: John Wiley & Sons, Ltd, pp. 934–950. https://doi.org/10.1002/9781119350927.ch61.

Goyette-Desjardins, G., Auger, J.P., Xu, J., Segura, M., 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 https://doi.org/10.1038/emi.2014.45.

Guerrero-Araya, E., Muñoz, M., Rodríguez, C., Paredes-Sabja, D., 2021. FastMLST: a multi-core tool for multilocus sequence typing of draft genome assemblies. Bioinform Biol. Insights 15. https://doi.org/10.1177/11779322211059238.

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086.

King, S.J., Leigh, J.A., Heath, P.J., Luque, I., Tarradas, C., Dowson, C.G., Whatmore, A. M., 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. https://doi.org/10.1128/jcm.40.10.3671-3680.2002.

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.

Murray, G.G.R., Charlesworth, J., Miller, E.L., Casey, M.J., Lloyd, C.T., Gottschalk, M., Tucker, A.W., (Dan), Welch, J.J., Weinert, L.A., 2020. Genome reduction is associated with bacterial pathogenicity across different scales of temporal and ecological divergence. Mol. Biol. Evol. 38, 1570–1579. https://doi.org/10.1093/molbev/msaa323.

Okura, M., Osaki, M., Nomoto, R., Arai, S., Osawa, R., Sekizaki, T., Takamatsu, D., 2016. Current taxonomical situation of *Streptococcus suis*. Pathogens 5. https://doi.org/10.3390/pathogens5030045.

Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., Fookes, M., Falush, D., Keane, J.A., Parkhill, J., 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31, 3691–3693. https://doi.org/10.1093/bioinformatics/btv421.

Pritchard, L., Glover, H., Humphris, R., G, S., Elphinstone, J.K., Toth, I., 2016. Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal. Methods 8, 12–24. https://doi.org/10.1039/C5AY02550H.

Prüfer, T., Rohde, J., Verspohl, J., Rohde, M., Greeff, A., de, Willenborg, J., Valentin-Weigand, P., 2019. Molecular typing of *Streptococcus suis* strains isolated from diseased and healthy pigs between 1996-2016. PloS One. https://doi.org/10.1371/journal.pone.0210801.

Rosselló-Mora, R., Amann, R., 2001. The species concept for prokaryotes. FEMS Microbiol Rev. 25, 39–67. https://doi.org/10.1111/j.1574-6976.2001.tb00571.x.

Schultsz, C., Jansen, E., Keijzers, W., Rothkamp, A., Duim, B., Wagenaar, J.A., van der Ende, A., 2012. Differences in the population structure of invasive *Streptococcus suis* strains isolated from pigs and from humans in the Netherlands. PLoS One 7, e33854. https://doi.org/10.1371/journal.pone.0033854.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Segura, M., Aragon, V., Brockmeier, S.L., Gebhart, C., Greeff, A., de, Kerdsin, A., O'Dea, M.A., Okura, M., Saléry, M., Schultsz, C., Valentin-Weigand, P., Weinert, L. A., Wells, J.M., Gottschalk, M., 2020. Update on *Streptococcus suis* research and prevention in the era of antimicrobial restriction: 4th international workshop on *S. suis*. Pathogens 9, 374. https://doi.org/10.3390/pathogens9050374.

Segura, M., Fittipaldi, N., Calzas, C., Gottschalk, M., 2017. Critical *Streptococcus suis* virulence factors: are they all really critical? Trends Microbiol 25, 585–599. https://doi.org/10.1016/j.tim.2017.02.005.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9, e112963. https://doi.org/10.1371/journal.pone.0112963.

Weinert, L.A., Chaudhuri, R.R., Wang, J., Peters, S.E., Corander, J., Jombart, T., Baig, A., Howell, K.J., Vehkala, M., Välimäki, N., Harris, D., Chieu, T.T.B., Van Vinh Chau, N., Campbell, J., Schultsz, C., Parkhill, J., Bentley, S.D., Langford, P.R., Rycroft, A.N., Wren, B.W., Farrar, J., Baker, S., Hoa, N.T., Holden, M.T.G., Tucker, A.W., Maskell, D.J., 2015. Genomic signatures of human and animal disease in the zoonotic pathogen *Streptococcus suis*. Nat. Commun. 6. https://doi.org/10.1038/ncomms7740.

Werinder, A., Aspán, A., Backhans, A., Sjölund, M., Guss, B., Jacobson, M., 2020. *Streptococcus suis* in Swedish grower pigs: occurrence, serotypes, and antimicrobial susceptibility. Acta Vet. Scand. 62, 1–9. https://doi.org/10.1186/s13028-020-00533-3.

Werinder, A., Aspán, A., Söderlund, R., Backhans, A., Sjölund, M., Guss, B., Jacobson, M., 2021. Whole-genome sequencing evaluation of MALDI-TOF MS as a species identification tool for *Streptococcus suis*. J. Clin. Microbiol. https://doi.org/10.1128/JCM.01297-21.

Wileman, T.M., Weinert, L.A., Howell, K.J., Wang, J., Peters, S.E., Williamson, S.M., Wells, J.M., Langford, P.R., Rycroft, A.N., Wren, B.W., Maskell, D.J., Tucker, A.W., 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. https://doi.org/10.1128/JCM.01712-18.

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 Genom. 12, 523. https://doi.org/10.1186/1471-2164-12-523.

Zhou, Z., Alikhan, N.-F., Sergeant, M.J., Luhmann, N., Vaz, C., Francisco, A.P., Carriço, J. A., Achtman, M., 2018. GrapeTree: visualization of core genomic relationships among 100,000 bacterial pathogens. Genome Res 28, 1395–1404. https://doi.org/10.1101/gr.232397.117.

Zhou, Z., Charlesworth, J., Achtman, M., 2021. HierCC: a multi-level clustering scheme for population assignments based on core genome MLST. Bioinformatics. https://doi.org/10.1093/bioinformatics/btab234.