Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Accessory-genome diversity and contrasting genotype–phenotype relationships in ten porcine Streptococcus suis isolates representing five serotypes

Hongqiao Lai
College of Animal Science, Xizang Agricultural and Animal Husbandry University
Hengzhi Zhu
Institute of Veterinary Medicine, Jiangsu Academy of Agricultural Sciences
Zenghao Xiang
Institute of Veterinary Medicine, Jiangsu Academy of Agricultural Sciences
Na Li
Institute of Veterinary Medicine, Jiangsu Academy of Agricultural Sciences
Jingze Song
Institute of Veterinary Medicine, Jiangsu Academy of Agricultural Sciences
Yonghong Qu
College of Animal Science, Xizang Agricultural and Animal Husbandry University
Kongwang He
Institute of Veterinary Medicine, Jiangsu Academy of Agricultural Sciences
Jiaqiang Niu
College of Animal Science, Xizang Agricultural and Animal Husbandry University
Libin Wen
Institute of Veterinary Medicine, Jiangsu Academy of Agricultural Sciences
Qi Xiao
20030014@jaas.ac.cn

Institute of Veterinary Medicine, Jiangsu Academy of Agricultural Sciences

Research Article

Keywords: Streptococcus suis, comparative genomics, public genomes, accessory genome, antimicrobial susceptibility, virulence-associated loci

Posted Date: October 28th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11193735/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Genomic markers differ in their ability to predict antimicrobial susceptibility and pathogenicity in *Streptococcus suis*. This study combined comparative genomics and experimental phenotyping of ten porcine isolates representing five serotypes with a public-genome phylogeny to assess genotype–phenotype relationships.

Results

The ten isolates represented eight sequence types and were placed within a phylogeny of 358 public genomes. The pan-genome comprised 3402 clusters, including 1832 accessory clusters (53.85%). All isolates carried *tet(M)* or *tet(O)*. Lincomycin MICs showed complete separation by *erm(B)* carriage: >32 μg/mL in seven carriers versus ≤ 0.015625 μg/mL in three non-carriers. Macrolide MICs varied among carriers, which showed five *erm(B)*-associated genomic arrangements. GyrA S81 and E85 were conserved despite differences in enrofloxacin MICs. Forty-seven of 63 virulence-associated loci were shared; three capsule-specific loci were restricted to serotype 2. The complete *mrp–epf–sly* homology profile occurred only in the three serotype 2 isolates, whereas SS4 ZJ, lacking a qualifying *mrp* match, had the lowest mouse LD_{50}. Mouse LD_{50} values ranged from 1.18×10^7 to 2.78×10^8 CFU/mouse across the ten isolates. Whole-blood recovery differed between host blood sources in an isolate-dependent manner.

Conclusions

Several resistance determinants closely corresponded to MICs, but the classical virulence-marker profile did not rank mouse lethality among these isolates. These findings support combining genomic characterization with experimental phenotypes and extending pathogenicity assessment beyond serotype 2 and its classical markers.

Introduction

Streptococcus suis is a porcine pathogen and zoonotic agent associated with meningitis, septicemia and other invasive infections. Its population contains diverse capsular serotypes and genetic lineages, and disease-associated isolates occur beyond serotype 2 [1, 2, 3]. Serotype is therefore informative for capsule classification but cannot substitute for whole-genome assessment of strain relatedness.

Comparative genomics has revealed extensive recombination, accessory-gene variation and antimicrobial resistance diversity in *S. suis* [4, 5, 6]. However, the presence of a resistance determinant or a homologue of a virulence-associated gene does not establish its expression, functional integrity or

phenotypic effect. Classical markers such as *mrp*, *epf* and *sly* have particularly limited value as universal predictors of pathogenicity across genetically diverse isolates [7, 8].

This study placed ten porcine *S. suis* isolates representing five serotypes and eight sequence types in a 368-genome phylogenetic framework to examine relationships between accessory-genome variation, antimicrobial susceptibility and pathogenicity-related phenotypes. The analyses focused on how closely acquired resistance genes tracked MICs and whether classical virulence-marker profiles reflected mouse lethality. Whole-blood recovery provided a complementary host-dependent phenotype.

Methods

Study isolates and experimental scope

Ten porcine isolates preserved at the Institute of Veterinary Medicine, Jiangsu Academy of Agricultural Sciences, were examined. SS2 NJ2025 was newly characterized; the other isolates have been described previously [9–12]. All ten isolates underwent comparative genomic analysis and MIC testing. Mouse challenge experiments involved six isolates, with published mouse pathogenicity data incorporated for SS3 ZJ, SS4 ZJ, SS7 ZJ and SS9 ZJ [12]. The isolates were purified and stored at − 80°C in Todd–Hewitt broth (Oxoid, UK) containing 20% glycerol. PCR-based serotypes were compared with genome-based calls.

Genome sequencing and characterization

Genomic DNA was extracted from logarithmic-phase cultures using a bacterial genomic DNA extraction kit (TIANGEN, China) and submitted to Majorbio Bio-Pharm Technology Co., Ltd. (Shanghai, China) for sequencing. Short- and long-read sequencing was performed using the Illumina NovaSeq 6000 platform with 2 × 150-bp paired-end libraries and the PacBio platform, respectively. Raw reads were quality-filtered using fastp v0.20.0. Short reads were assembled using Unicycler v0.4.8, whereas long reads were assembled using NextDenovo v2.5.2 and polished with the corresponding short-read data. Assembly quality was evaluated using QUAST v5.3.0, CheckM v1.2.2 and BUSCO v5.4.5; assemblies with > 95% completeness and < 5% contamination were considered acceptable. Genome annotation was performed using Prokka v1.15.6.

Species identity was assessed with fastANI v1.34 against the *S. suis* type-material assembly GCA_900475585.1, using ≥ 95% average nucleotide identity (ANI) as the species threshold [13]. *recN* sequences provided an additional species check. Capsular polysaccharide synthesis loci were compared with serotype reference sequences and markers [14], and multilocus sequence types were assigned using mlst v2.33.1 and the PubMLST *S. suis* scheme [15]. Pairwise ANI was calculated among the ten isolates and against P1/7.

Public-genome phylogenetic context

Public *S. suis* genomes were retrieved using NCBI Datasets v2 in September 2026. Of 13,401 records, 9,482 passed basic quality assessment (fastANI v1.34, ANI ≥ 95%; CheckM2 v1.1.0, completeness ≥ 95%, contamination ≤ 5%). After deduplication, candidates were restricted to serotypes 2, 3, 4, 7 and 9 using SsuisChara, with > 95% cps locus coverage, yielding 2,455 genomes in 28 PopPUNK v2.7.8 clusters. Representative selection was stratified by genomic cluster, serotype, MLST profile and country, while retaining serotype 9 genomes and clusters with five or fewer candidates [16], producing 358 public representatives from 27 countries. The 28 clusters refer to the public-genome candidate pool, whereas the 30 clusters shown in Fig. 1 refer to the final 368-genome dataset including the ten study isolates. The 358 representatives and ten study isolates were uniformly annotated with Prokka v1.15.6. Core genes present in ≥ 95% of genomes were identified using Panaroo v1.7.0 in strict mode and aligned with MAFFT [17, 18]. A maximum-likelihood phylogeny was inferred from the Panaroo-filtered alignment using IQ-TREE v3.1.2 under GTR + F+G4, with 1,000 ultrafast bootstrap and 1,000 SH-aLRT replicates [19]. No additional recombination masking was performed; the topology was interpreted as unrooted and is displayed with a midpoint root for orientation only.

Ten-isolate pan-genome and mobile-element predictions

The ten study genomes were analyzed with Panaroo v1.7.0 in strict mode [17]. Core clusters were defined as those present in all ten isolates (≥ 95% occupancy for $n = 10$), shared accessory clusters as those present in two to nine isolates, and isolate-specific clusters as those present in one isolate. Pan- and core-genome accumulation was summarized across 1000 randomized genome orders. COG functional categories were assigned using eggNOG annotations and their proportions compared descriptively. Genomic islands and prophages were predicted with IslandPath-DIMOB v1.0.0 and Phigaro v2.4.0, respectively [20, 21].

Resistance gene screening and MIC determination

Acquired resistance genes were screened with ABRicate v1.4.0 against ResFinder, requiring ≥ 90% nucleotide identity and ≥ 80% reference coverage [22]. Gene names and associated antimicrobial classes were harmonized with CARD [23].

The genomic regions flanking *erm(B)* were compared among *erm(B)*-positive isolates to characterize variation in local gene organization.

Broth microdilution MICs were determined for 12 antimicrobials following CLSI M07, using cation-adjusted Mueller–Hinton broth supplemented with 5% lysed horse blood, twofold antimicrobial dilutions and a final inoculum of approximately 5×10^5 CFU/mL. The initial antimicrobial concentration ranges were 0.125–8 μg/mL for enrofloxacin, 1–64 μg/mL for tetracycline, 0.25–16 μg/mL for doxycycline, 0.5–

32 μg/mL for erythromycin, azithromycin and lincomycin, 2–128 μg/mL for gentamicin, 8–512 μg/mL for streptomycin and kanamycin, 1–64 μg/mL for chloramphenicol, and 0.03125–2 μg/mL for ampicillin and cefotaxime. Plates were incubated at 37 ± 1°C for 24 h. The MIC was defined as the lowest concentration without visible bacterial growth. *Streptococcus pneumoniae* ATCC 49619 served as the quality-control strain; each plate included inoculated antimicrobial-free and sterility controls. Results were accepted only when controls were valid and quality-control MICs fell within the applicable ranges; the quality-control MICs obtained are listed in Supplementary tables, Table S4d. Each strain–antimicrobial combination was tested once. For the 45 of 120 endpoints at or below the lower limit of the initial dilution range, MICs were retested using seven additional twofold dilutions extending the lower-concentration range. Endpoints outside the tested range retained their ≤ or > signs. MICs were compared directly without assigning clinical susceptibility categories.

The *gyrA* quinolone resistance-determining region (QRDR) was amplified using primers 5'-CTCTGCCTGATGTACGTGATGGTC-3' and 5'-GCACGGTGAATTCCAGATTTCCCC-3', yielding a 598-bp product. Bidirectional Sanger sequences (Sangon Biotech, Shanghai, China) were compared with genome-derived calls at GyrA residues S81 and E85.

Virulence-associated sequence homology

A panel of 63 virulence-associated reference loci was screened against the study genomes with BLASTN [24]. The best hit for each reference locus was retained; qualifying matches required ≥ 80% nucleotide identity and ≥ 80% query coverage. Partial or low-coverage hits were distinguished from qualifying matches. Sequence integrity and coverage were examined in greater detail for *mrp*, *epf* and *sly*.

Whole-blood assay

Pig blood obtained as a by-product of slaughter and fresh BALB/c mouse blood were anticoagulated with heparin (15 U/mL). Mouse blood was pooled from ten donors per independent experiment, using 30 donors overall. Bacterial suspensions (50 μL; approximately 1×10^6 CFU/mL) were mixed with 450 μL of blood or phosphate-buffered saline (PBS) and incubated at 37°C for 1 h. Viable bacteria were enumerated by plating. Recovery was calculated as blood CFU/mL divided by the corresponding PBS-control CFU/mL; ratios > 1 indicate greater viable recovery than in PBS. Three independent experiments were performed per isolate and blood source. In each experiment, the same batch of pig blood obtained as a slaughter by-product was used for all ten isolates.

Mouse challenge and LD estimation

SS2 NJ2025, SS2 NJ, SS2-1, SS3 JSHZ, SS3 GD7 and SS3 H30 were tested in female BALB/c mice aged 6–8 weeks and weighing 18–22 g (Shanghai BK Kay Wing Biotechnology Co., Ltd.; production licence SCXK (Hu) 2023-0009). Each isolate was tested at three dose levels with five mice per dose; five

additional mice served as saline controls. Mice received 1 mL of bacterial suspension intraperitoneally and were monitored daily for clinical signs and mortality for 10 days. Animals reaching humane endpoints, including severe depression, persistent lateral recumbency or a marked decrease in body temperature, were euthanized and counted as mortality endpoints.

LD_{50} was estimated using the Reed–Muench method [25] and expressed as CFU/mouse. Saline controls were excluded from estimation. Animal use comprised 90 challenged mice, five saline controls and 30 blood donors.

Statistical analysis

Genotype–MIC comparisons were descriptive, with censored endpoints retained rather than treated as exact concentrations.

Whole-blood recovery was analyzed using $\log_{10}$-transformed blood/PBS CFU ratios in a linear mixed-effects model with isolate, blood source and their interaction as fixed effects and experiment as a random intercept. Ten within-isolate pig-versus-mouse contrasts were tested as paired t-tests with Holm adjustment; adjusted P values < 0.05 were considered significant. The experiment-to-experiment variance component was negligible (ICC < 0.001). Analyses used Python v3.11 (statsmodels v0.14, SciPy v1.14). No inferential comparisons were conducted among serotypes or sequence types because the panel was small and unbalanced. The per-experiment ratios underlying these analyses are listed in Supplementary tables, Table S6d.

Results

Genomic characteristics and placement among public genomes

The ten isolates represented five serotypes and eight sequence types (Tables 1 and 2). The serotype 2 isolates comprised ST1 and ST7, whereas the four serotype 3 isolates comprised ST27, ST1004 and ST117. Serotypes 4, 7 and 9 were represented by ST17, ST373 and ST243, respectively. Genome sizes ranged from 2068760 to 2416981 bp, with GC contents of 40.85–41.38% and 1951–2357 predicted coding sequences. Eight assemblies contained one contig; SS2 NJ and SS3 H30 each contained two. Species comparisons yielded ANI ≥ 98% against the type-material reference, with concordant *recN* assignments.

Table 1
Sources of the ten study isolates

Isolate	Year	Location in China	Isolation site	Clinical status
SS2 NJ2025	2025	Nanjing, Jiangsu	Synovial fluid	Lameness
SS2 NJ	-	Nanjing, Jiangsu	Not reported	-
SS2-1	1998	Nantong, Jiangsu	Organs	Pig with streptococcosis
SS3 JSHZ	2020	Hongze, Jiangsu	Lung	Acute mortality
SS3 GD7	-	Guangdong	Lung	Acute mortality
SS3 H30	-	Jiangsu	Tonsil	Healthy
SS3 ZJ	2024	Zhenjiang, Jiangsu	Lung	Acute mortality; pneumonia
SS4 ZJ	2024	Zhenjiang, Jiangsu	Blood	Acute mortality; septicemia/pneumonia
SS7 ZJ	2024	Zhenjiang, Jiangsu	Lung	Acute mortality; pneumonia
SS9 ZJ	2024	Zhenjiang, Jiangsu	Lung	Acute mortality; pneumonia

Table 2
Genome assembly and typing characteristics

Isolate	Serotype	ST	Genome size (bp)	Contigs	GC (%)	CDSs
SS2 NJ2025	2	ST1	2068760	1	41.17	1951
SS2 NJ	2	ST7	2103245	2	41.09	1988
SS2-1	2	ST7	2095973	1	41.11	1978
SS3 JSHZ	3	ST27	2203332	1	41.11	2145
SS3 GD7	3	ST1004	2282512	1	41.07	2223
SS3 H30	3	ST117	2263307	2	41.11	2179
SS3 ZJ	3	ST117	2240342	1	41.06	2170
SS4 ZJ	4	ST17	2240003	1	40.99	2128
SS7 ZJ	7	ST373	2416981	1	40.85	2357
SS9 ZJ	9	ST243	2130830	1	41.38	2019

The study isolates occupied several positions across the 368-genome phylogeny (Fig. 1). The three serotype 2 isolates occurred within the same broad sector, while the serotype 3 isolates occupied another; SS4 ZJ, SS7 ZJ and SS9 ZJ occupied separate positions. These placements show the genetic backgrounds represented in the experimental panel.

Pairwise ANI within the study panel ranged from 96.38% to approximately 100.00% (Supplementary tables, Table S2). ANI was 99.999733% between SS2 NJ and SS2-1 and 99.963501% between SS3 H30 and SS3 ZJ, consistent with their shared ST7 and ST117 assignments, respectively. The four serotype 3 isolates shared ANI of 99.57–99.96% despite representing three sequence types.

Fig. 1. Public-genome phylogenetic context of the ten study isolates.

Maximum-likelihood core-genome phylogeny of 358 public *S. suis* genomes and the ten study isolates. Two annotation rings surround the tree: the inner ring indicates serotype, and the outer ring indicates PopPUNK cluster, with the seven clusters containing ten or more genomes shown in alternating dark and light gray and labeled PC1–PC6 and PC8, and the remaining 23 clusters shown in pale gray. Study isolates are marked by red circles at the corresponding tips and are labeled with isolate name and ST. The tree is drawn with a midpoint display root, which sets orientation only and preserves every split and branch length; no biological outgroup is implied and the topology is interpreted as unrooted. The phylogeny was inferred without additional recombination masking. The scale bar represents 0.01 substitutions per site.

A large accessory genome differentiates the study isolates

Accessory genes accounted for more than half of the ten-isolate pan-genome: 1832 of 3402 clusters (53.85%), comprising 963 shared accessory clusters (28.31%) and 869 isolate-specific clusters (25.54%). The remaining 1570 clusters (46.15%) were core (Fig. 2; Supplementary tables, Table S2). The pan-genome continued to expand as isolates were added, while the core-genome curve approached 1570 clusters, revealing substantial gene-content diversity within this small panel.

The number of shared accessory clusters per isolate ranged from 214 to 611. SS7 ZJ had the most isolate-specific clusters (336), followed by SS9 ZJ (223) and SS4 ZJ (172); none were identified in SS2 NJ, SS2-1 or SS3 ZJ. Among these strains, 141 accessory clusters were shared exclusively by the four serotype 3 isolates and 91 by the three serotype 2 isolates. These intersections describe the sampled strains and cannot distinguish serotype effects from lineage-associated variation.

Core clusters included many genes involved in translation and nutrient metabolism functions. Accessory fractions contained relatively larger proportions of defense, mobilome and replication/recombination categories, together with many unannotated clusters (Fig. 2D). Predicted genomic islands ranged from four to eleven per isolate and prophages from zero to four (Fig. 2E; Supplementary tables, Table S3). SS4 ZJ had the largest island count, and SS3 GD7 had the largest prophage count. No prophages were predicted in the three serotype 2 isolates.

Fig. 2. Pan-genome composition and predicted mobile elements in the ten study isolates.

(A) Pan- and core-genome accumulation across 1000 randomized genome orders; lines show means and bands show the 2.5th–97.5th percentiles. (B) Shared accessory clusters occur in two to nine isolates; isolate-specific clusters occur in one. (C) The 20 largest accessory-cluster intersections. (D) Descriptive COG category proportions. J, translation; G, carbohydrate metabolism; E, amino acid metabolism; K, transcription; L, replication and repair; M, cell envelope biogenesis; V, defense; X, mobilome; S, function unknown. (E) Genomic islands and prophages predicted by IslandPath-DIMOB and Phigaro.

Resistance determinants track MICs but do not fully resolve their magnitude

All ten isolates carried a tetracycline resistance determinant: *tet(M)* occurred in SS2 NJ, SS2-1 and SS4 ZJ, and *tet(O)* in the other seven isolates (Fig. 3A; Supplementary tables, Table S4). *erm(B)* occurred in all seven non-serotype 2 isolates. *ant(6)-Ia* was detected in SS2 NJ, SS2-1 and SS4 ZJ; *aac(6')-aph(2'')* occurred only in SS4 ZJ, and *ant(9)-Ia* and *cat(pC194)* only in SS2 NJ.

The 120 isolate–antimicrobial measurements included 74 exact MIC endpoints, 27 lower-censored endpoints and 19 upper-censored endpoints (Fig. 3B). Tetracycline and doxycycline MICs ranged from 8 to 64 μg/mL and 0.5 to 16 μg/mL, respectively. All seven *erm(B)*-positive isolates had lincomycin MICs >

32 μg/mL, compared with ≤ 0.015625 μg/mL in the three *erm(B)*-negative isolates. Erythromycin and azithromycin MICs were > 32 μg/mL in five *erm(B)*-positive isolates, 16 and 32 μg/mL in SS3 H30, and 2 and 8 μg/mL in SS7 ZJ. Thus, *erm(B)* carriage corresponded to higher MICs, but MIC magnitude varied among carriers.

Additional gene–MIC correspondences were equally distinctive. SS4 ZJ, the only *aac(6')-aph(2")* carrier, had gentamicin and kanamycin MICs > 128 and > 512 μg/mL, respectively. SS2 NJ, the only *cat(pC194)*-positive isolate, had a chloramphenicol MIC of 32 μg/mL, compared with ≤ 4 μg/mL in all other isolates. All three *ant(6)-Ia* carriers had streptomycin MICs of 128 μg/mL, whereas non-carriers ranged from ≤ 0.25 to 32 μg/mL (Fig. 3C). These consistent separations show that resistance-gene carriage tracked antimicrobial profiles closely here.

Enrofloxacin MICs ranged from 0.25 to 2 μg/mL. All ten isolates retained the AGT and GAA codons encoding GyrA S81 and E85, respectively, with Sanger results concordant with the genome-derived calls (Supplementary tables, Table S4). Variation at these two positions therefore did not account for the observed MIC differences; other QRDR positions, target genes and non-target mechanisms were not resolved by this assay.

The seven *erm(B)*-positive isolates showed five genomic-context types (Fig. 3D). Type 1 was shared by SS3 GD7, SS3 H30 and SS3 ZJ; types 2–5 occurred in SS3 JSHZ, SS4 ZJ, SS7 ZJ and SS9 ZJ, respectively. Thus, a common resistance determinant occurred in multiple local gene arrangements.

Fig. 3. Acquired resistance genes and antimicrobial MICs in the ten study isolates.

(A) Acquired resistance-gene calls at ≥ 90% identity and ≥ 80% reference coverage. (B) Isolate-level antimicrobial MICs; censored values retain their ≤ or > boundaries. (C) Selected genotype–MIC comparisons. (D) Five *erm(B)*-associated genomic-context types; arrows indicate annotated genes and their orientation. AMP, ampicillin; CTX, cefotaxime; TET, tetracycline; DOX, doxycycline; ERY, erythromycin; AZM, azithromycin; LIN, lincomycin; STR, streptomycin; GEN, gentamicin; KAN, kanamycin; CHL, chloramphenicol; ENR, enrofloxacin.

Classical virulence-marker carriage does not rank mouse lethality

Detectable BLASTN homology was recovered at 62 of the 63 screened loci in each of the three serotype 2 isolates; however, three of the 63 reference loci (*cps2E*, *cps2F* and *neuB*) are serotype-2-specific capsule genes that are absent from other serotype backgrounds, so the comparable non-capsule panel comprises 60 loci and the serotype 2 isolates returned homology at 59 of these. Among the same 60 non-capsule loci, counts were 50, 50, 49 and 50 for SS3 JSHZ, SS3 GD7, SS3 H30 and SS3 ZJ, respectively, and 58, 52 and 49 for SS4 ZJ, SS7 ZJ and SS9 ZJ. Forty-seven loci returned homology in

every isolate and 35 had nucleotide identity ≥ 95% in all ten (Fig. 4; Supplementary tables, Tables S5A and S5C). These tallies correspond to numerical identity entries in Table S5A and are not counts of qualifying matches. Alignment statistics for the mrp–epf–sly marker set and listed below-threshold hits are given in Supplementary tables, Table S5B. The counts summarize the selected reference panel rather than the complete virulence repertoire.

Qualifying *mrp*, *epf* and *sly* matches co-occurred only in the three serotype 2 isolates. SS4 ZJ retained *epf* and *sly* but had only a partial *mrp* match (33.6% query coverage, 97.2% identity over the C-terminal third; the N-terminal repeat-containing region was absent). SS7 ZJ retained a qualifying *sly* match and a partial *mrp* match (54.1% coverage), while the serotype 3 isolates and SS9 ZJ lacked the complete profile (SS9 ZJ had no *mrp* hit at any coverage). The variable-locus panel further resolved sequence heterogeneity in *ofs*, *ideS*, *zmpC* and other loci (Fig. 4B). These differences distinguish both marker carriage and allelic similarity; a below-threshold match alone does not establish loss of function. Per-locus alignment statistics for *mrp*, *epf* and *sly* in all ten isolates are listed in Supplementary tables, Table S5B.

Fig. 4. Nucleotide homology of 63 virulence-associated loci across the ten study isolates.

(A) Nucleotide homology across 63 virulence-associated reference loci. Qualifying matches require ≥ 80% identity and ≥ 80% query coverage. The color scale represents nucleotide identity; light gray indicates no qualifying match, and dark gray with # indicates a partial or low-quality hit. (B) Selected variable loci across the ten isolates. The dendrogram summarizes locus-profile similarity rather than isolate phylogeny. Reference accessions and sequence lengths are listed in Supplementary tables, Table S5C.

Mouse LD_{50} values varied approximately 24-fold, from 1.18×10^7 CFU/mouse for SS4 ZJ to 2.78×10^8 CFU/mouse for SS2 NJ2025 (Table 3). SS3 JSHZ had the second-lowest value. The isolate with the lowest LD_{50} (SS4 ZJ) lacked the complete classical marker profile, whereas the most marker-rich serotype 2 isolate (SS2 NJ2025) had the highest. This observation is consistent with the classical markers being insufficient to rank lethality within this small panel, but ten isolates cannot support broader generalization.

Table 3
Mouse LD_{50} estimates for the ten study isolates

Isolate	Serotype	LD_{50} (CFU/mouse)	$\log_{10} LD_{50}$	Data source
SS2 NJ2025	2	2.78×10^8	8.444	Current study
SS2 NJ	2	1.48×10^8	8.170	Current study
SS2-1	2	4.3×10^7	7.633	Current study
SS3 JSHZ	3	2.24×10^7	7.349	Current study
SS3 GD7	3	2.55×10^7	7.406	Current study
SS3 H30	3	2.5×10^8	8.398	Current study
SS3 ZJ	3	2.46×10^8	8.391	Published [12]
SS4 ZJ	4	1.18×10^7	7.072	Published [12]
SS7 ZJ	7	3.04×10^7	7.483	Published [12]
SS9 ZJ	9	2.63×10^7	7.420	Published [12]

LD_{50} values are rounded to three significant figures; logarithms were calculated from unrounded values. Lower LD_{50} indicates greater lethality in this mouse model. Values for SS3 ZJ, SS4 ZJ, SS7 ZJ and SS9 ZJ were incorporated from the previous study [12].

Host-dependent whole-blood recovery

Mean recovery ratios ranged from 0.740 to 2.981 in pig blood and from 0.424 to 2.977 in mouse blood (Supplementary tables, Table S6; Fig. 5). SS2-1 had the highest mean recovery in pig blood, whereas SS7 ZJ had the highest in mouse blood. The linear mixed-effects model identified significant isolate, blood-source and interaction effects (all $P < 0.001$), with negligible experiment-to-experiment variance (ICC < 0.001). Holm-adjusted within-isolate contrasts indicated significantly higher recovery in pig blood for SS2 NJ2025 (adjusted $P = 0.041$), SS3 H30 (adjusted $P = 0.025$) and SS3 JSHZ (adjusted $P = 0.023$). The remaining seven isolates did not show significant pig-versus-mouse contrasts after Holm correction, although the point estimates for SS2 NJ, SS2-1, SS3 GD7, SS3 ZJ and SS4 ZJ were numerically higher in pig blood and those for SS7 ZJ and SS9 ZJ were numerically higher in mouse blood.

Fig. 5. Whole-blood recovery of the ten *S. suis* study isolates.

Each point represents a single experiment (n = 3 per isolate per blood source). Horizontal bars show mean ± SD. The dashed line at 1.0 marks equal recovery in blood and PBS. Asterisks denote significant

pig-versus-mouse differences within an isolate (Holm-adjusted paired t-test: *P < 0.05, **P < 0.01). Isolates are ordered by descending mean pig-blood recovery.

Discussion

Acquired resistance determinants closely tracked several antimicrobial MICs, whereas the classical *mrp–epf–sly* profile did not rank mouse lethality. These findings span five serotypes and eight sequence types and show that genomic markers differ in how well they predict phenotype. Here, specific resistance determinants were informative indicators of antimicrobial response, but the classical virulence-marker combination was insufficient to rank pathogenicity in the mouse model.

Accessory genes constituted 53.85% of the observed pan-genome, making gene-content variation a major axis of diversity among these isolates. SS7 ZJ, SS9 ZJ and SS4 ZJ contributed the largest isolate-specific repertoires, and island and prophage predictions further differentiated the isolates. These findings support the view of *S. suis* as a genomically plastic species [2, 4, 26] and show the information gained by including non-serotype 2 backgrounds. Pan-genome analyses of serotype 2 alone have identified accessory loci associated with virulence phenotypes [27], yet the accessory fractions unique to serotypes 4, 7 and 9 in our data show that serotype 2-based panels underrepresent the species-level gene pool. Recent work on prophage-mediated lysogenic conversion in serotype 9 further illustrates how mobile elements reshape the accessory genome and contribute to virulence evolution [28]. The expanding accumulation curve is compatible with an unsaturated pan-genome in this sample, although ten genomes cannot establish species-wide openness. Likewise, accessory intersections shared by a serotype remain inseparable from ancestry in this sampling design.

The strongest resistance signal was the complete separation of lincomycin MICs by *erm(B)* carriage: >32 μg/mL in all seven carriers versus ≤ 0.015625 μg/mL in all three non-carriers, a separation exceeding 2048-fold even at the assay boundaries. Correspondences involving *aac(6')-aph(2")*, *ant(6)-Ia* and *cat(pC194)* reinforced the link between acquired determinants and antimicrobial profiles, consistent with established *S. suis* resistance biology [5, 6]. Yet *erm(B)* carriers differed in macrolide MICs and occupied five genomic-context types. The combination of a shared determinant and variable phenotypic magnitude makes genomic context a relevant candidate explanation, alongside regulation and strain background. Context diversity supports structural heterogeneity around *erm(B)*; horizontal transfer of different *erm(B)*-carrying mobile elements, including integrative conjugative elements, has been documented across serotypes [29], providing a mechanistic basis for the multiple genomic arrangements observed here. The combination of a shared determinant in structurally diverse contexts and variable macrolide MICs underscores the need to pair genomic detection with phenotypic measurement.

Enrofloxacin MICs varied eightfold despite conservation of GyrA S81 and E85. These two positions therefore did not explain the variation observed here. Changes elsewhere in *gyrA*, other targets and non-

target mechanisms remain possible. Together with the variable macrolide MICs among *erm(B)* carriers, these results support retaining direct susceptibility testing alongside genomic screening [5, 22].

The virulence results challenge a simple marker-count interpretation. The three serotype 2 isolates had the largest detected repertoire and the complete classical *mrp–epf–sly* profile, but neither feature identified the isolate with the lowest mouse LD_{50}. That isolate was SS4 ZJ, which lacked a qualifying *mrp* match; SS3 JSHZ also showed a low estimate despite an incomplete classical profile. These observations place non-serotype 2 isolates firmly within the pathogenicity spectrum represented here and are consistent with the context-dependent value of classical markers [7, 8]. The implication is not that these loci lack biological roles, but that their carriage alone did not rank lethality here. The emerging clinical significance of non-serotype 2 isolates reinforces this conclusion. Serotype 4 clonal complex 94 strains recovered from human meningitis and swine cases carried virulence gene repertoires comparable to those of epidemic serotype 2 lineages [30], and serotype 7 ST373 strains caused 100% mortality in mouse infection models despite lacking the classical marker profile [31]. These reports, together with the low LD_{50} of SS4 ZJ observed here, support extending pathogenicity evaluation beyond serotype 2. Allelic variation, expression and interactions with the wider genomic background may contribute and warrant functional testing.

Whole-blood recovery showed that survival depended on host blood source. The direction of pig-versus-mouse differences varied among isolates: three showed significantly higher pig-blood recovery after Holm correction, while SS7 ZJ and SS9 ZJ had numerically higher mouse-blood recovery. A single host-blood assay therefore cannot establish a single survival ranking. Antibodies, complement and phagocyte responses offer plausible explanations. Antibody- and complement-dependent granulocyte activity has been implicated in porcine *S. suis* clearance [32]. These mechanisms were not tested here. Blood recovery and intraperitoneal mouse lethality probe different biological conditions and measure different things and should not be extrapolated directly to disease severity in pigs or zoonotic risk.

The study panel comprised ten isolates from a single institution, so the observations describe this convenience sample and should be tested in larger, population-stratified collections. Censored MIC endpoints and homology thresholds limited quantitative precision.

Conclusions

In these ten *S. suis* isolates, the five serotypes exhibited extensive accessory-genome diversity and contrasting genotype–phenotype relationships. Several acquired resistance genes closely tracked MIC profiles, whereas the classical *mrp–epf–sly* profile did not rank mouse lethality. The findings for SS4 ZJ support evaluating non-serotype 2 isolates beyond classical marker profiles. These observations describe a small convenience panel; whether they extend to the wider *S. suis* population requires testing in larger, population-stratified collections. Genomic context and directly measured phenotypes together characterize *S. suis* strains more fully than marker profiles alone.

Abbreviations

ANI
average nucleotide identity
CFU
colony-forming units
CLSI
Clinical and Laboratory Standards Institute
COG
Clusters of Orthologous Groups
LD_{50}
median lethal dose
MIC
minimum inhibitory concentration
MLST
multilocus sequence typing
PBS
phosphate-buffered saline
QRDR
quinolone resistance-determining region
ST
sequence type.

Declarations

Ethics approval and consent to participate

All animal experiments were approved by the Animal Ethics Committee of Jiangsu Academy of Agricultural Sciences (approval number IACUC-RE-2026-09-058) and were performed in accordance with the institutional guidelines for the care and use of laboratory animals and with the relevant national regulations of the People's Republic of China. The study is reported in accordance with the ARRIVE guidelines 2.0. Pig blood was obtained as a by-product of slaughter; no blood was collected from live pigs and no pigs were killed specifically for this study. Mice were monitored at least twice daily and animals that reached predefined humane endpoints were euthanised by carbon dioxide inhalation. This study did not involve human participants, human data or human tissue.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are available in the NCBI BioProject repository under accession PRJNA1525114 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1525114). The BioSample and SRA accessions for each of the ten study isolates are listed in Supplementary tables, Table S1. The accessions of the 358 public genomes used for the phylogeny, the core-gene alignment and the resulting Newick tree file are provided in Supplementary tables. All other data generated or analysed during this study, including genome and pan-genome summaries, mobile-element predictions, resistance-gene calls, MICs, GyrA QRDR sequencing results, the virulence-associated homology matrices and the whole-blood and mouse-challenge phenotype data, are included in Supplementary tables.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the Science and Technology Projects of Xizang Autonomous Region, China (XZ202601ZY0136) and the National Key Research and Development Program of China (2022YFD1800902-2).

Authors' contributions

HL contributed to methodology, investigation, formal analysis and drafting. HZ contributed to investigation, data curation and validation. ZX contributed to formal analysis, software and visualization. NL contributed to investigation and validation. JS contributed to investigation and data curation. YQ contributed to investigation and resources. KH contributed to resources, supervision and manuscript revision. JN contributed to conceptualization, project administration, funding acquisition and manuscript revision. LW contributed to conceptualization and supervision. QX contributed to conceptualization, project administration, funding acquisition and manuscript revision. All authors have read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Goyette-Desjardins G, Auger JP, Xu J, Segura M, Gottschalk M. *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. 2014;3:e45. doi:10.1038/emi.2014.45.
2. Weinert LA, Chaudhuri RR, Wang J, Peters SE, Corander J, Jombart T, et al. Genomic signatures of human and animal disease in the zoonotic pathogen *Streptococcus suis*. Nat. Commun. 2015;6:6740. doi:10.1038/ncomms7740.

3. Dong X, Chao Y, Zhou Y, Zhou R, Zhang W, Fischetti VA, et al. The global emergence of a novel *Streptococcus suis* clade associated with human infections. EMBO Mol. Med. 2021;13:e13810. doi:10.15252/emmm.202013810.
4. Lee IPA, Andam CP. Frequencies and characteristics of genome-wide recombination in *Streptococcus agalactiae*, *Streptococcus pyogenes*, and *Streptococcus suis*. Sci. Rep. 2022;12:1515. doi:10.1038/s41598-022-04995-5.
5. Hadjirin NF, Miller EL, Murray GGR, Yen PLK, Phuc HD, Wileman TM, et al. Large-scale genomic analysis of antimicrobial resistance in the zoonotic pathogen *Streptococcus suis*. BMC Biol. 2021;19:191. doi:10.1186/s12915-021-01094-1.
6. Dechêne-Tempier M, Marois-Créhan C, Libante V, Jouy E, Leblond-Bourget N, Payot S. Update on the mechanisms of antibiotic resistance and the mobile resistome in the emerging zoonotic pathogen *Streptococcus suis*. Microorganisms 2021;9:1765. doi:10.3390/microorganisms9081765.
7. Segura M, Fittipaldi N, Calzas C, Gottschalk M. Critical *Streptococcus suis* virulence factors: are they all really critical? Trends Microbiol. 2017;25:585–599. doi:10.1016/j.tim.2017.02.005.
8. Roodsant TJ, van der Putten BCL, Tamminga SM, Schultsz C, van der Ark KCH. Identification of *Streptococcus suis* putative zoonotic virulence factors: a systematic review and genomic meta-analysis. Virulence 2021;12:2787–2797. doi:10.1080/21505594.2021.1985760.
9. Zhou JM, Wang DD, Ni YX, Zhu HD, Sun K, Li B, et al. Isolation, identification and pathogenic characteristics of *Streptococcus suis* from a pig farm in Jiangsu Province. Animal Husbandry & Veterinary Medicine 2020;52:85–89.
10. Wang J, He K, He J, Ni Y, Hou J, Lv L, et al. Animal pathogenicity test of *Streptococcus suis* type 2. Chinese Journal of Veterinary Science and Technology 2001;31:27–29.
11. Xiao Q, Zhao XL, Qian WX, Guo JH, Tan JM, Yu ZY, et al. Biological characteristics and pathogenicity of *Streptococcus suis* serotype 3 isolates. Jiangsu J. Agric. Sci. 2020;36:1475–1481. doi:10.3969/j.issn.1000-4440.2020.06.017.
12. Lai HQ, Zhu HZ, Li N, et al. Isolation, identification and characterization of *Streptococcus suis* serotypes 3, 4, 7 and 9 from a mixed-serotype infection in a pig farm in Jiangsu Province. Chinese Journal of Animal Infectious Diseases 2025;33:11–20.
13. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 2018;9:5114. doi:10.1038/s41467-018-07641-9.
14. Athey TBT, Teatero S, Lacouture S, Takamatsu D, Gottschalk M, Fittipaldi N. Determining *Streptococcus suis* serotype from short-read whole-genome sequencing data. BMC Microbiol. 2016;16:162. doi:10.1186/s12866-016-0782-8.
15. King SJ, Leigh JA, Heath PJ, Luque I, Tarradas C, Dowson CG, et al. 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. 2002;40:3671–3680. doi:10.1128/JCM.40.10.3671-3680.2002.

16. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, Weiser JN, et al. Fast and flexible bacterial genomic epidemiology with PopPUNK. Genome Res. 2019;29:304–316. doi:10.1101/gr.241455.118.
17. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, et al. Producing polished prokaryotic pangenomes with the Panaroo pipeline. Genome Biol. 2020;21:180. doi:10.1186/s13059-020-02090-4.
18. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–780. doi:10.1093/molbev/mst010.
19. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 2020;37:1530–1534. doi:10.1093/molbev/msaa015.
20. Bertelli C, Brinkman FSL. Improved genomic island predictions with IslandPath-DIMOB. Bioinformatics 2018;34:2161–2167. doi:10.1093/bioinformatics/bty095.
21. Starikova EV, Tikhonova PO, Prianichnikov NA, Rands CM, Zdobnov EM, Ilina EN, et al. Phigaro: high-throughput prophage sequence annotation. Bioinformatics 2020;36:3882–3884. doi:10.1093/bioinformatics/btaa250.
22. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, et al. ResFinder 4.0 for predictions of phenotypes from genotypes. J. Antimicrob. Chemother. 2020;75:3491–3500. doi:10.1093/jac/dkaa345.
23. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, et al. CARD 2020: antibiotic resistome surveillance with the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res. 2020;48:D517–D525. doi:10.1093/nar/gkz935.
24. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009;10:421. doi:10.1186/1471-2105-10-421.
25. Reed LJ, Muench H. A simple method of estimating fifty per cent endpoints. Am. J. Epidemiol. 1938;27:493–497. doi:10.1093/oxfordjournals.aje.a118408.
26. Dechêne-Tempier M, Bougeard S, Loux V, Chiapello H, Libante V, Marois-Créhan C, et al. Pan-immune system, mobilome and resistome in *Streptococcus suis*. Microb. Genom. 2025;11:001521. doi:10.1099/mgen.0.001521.
27. Guo G, Du D, Yu Y, Zhang Y, Qian Y, Zhang W. Pan-genome analysis of *Streptococcus suis* serotype 2 revealed genomic diversity among strains of different virulence. Transbound. Emerg. Dis. 2021;68:637–647. doi:10.1111/tbed.13725.
28. Wen Z, Guo Y, Liang Y, Li Y, Lian K, Tang R, et al. Prophage-mediated lysogenic conversion drives virulence evolution and genomic plasticity in *Streptococcus suis* serotype 9. Microbiol. Spectr. 2026;14:e00061-26. doi:10.1128/spectrum.00061-26.
29. Chen L, Huang J, Huang X, He Y, Sun J, Dai X, et al. Horizontal transfer of different *erm*(B)-carrying mobile elements among *Streptococcus suis* strains with different serotypes. Front. Microbiol. 2021;12:628740. doi:10.3389/fmicb.2021.628740.

30. Hatrongjit R, Boueroy P, Jenjaroenpun P, Wongsurawat T, Meekhanon N, Chopjitt P, et al. Genomic characterization and virulence of *Streptococcus suis* serotype 4 clonal complex 94 recovered from human and swine samples. PLoS One 2023;18:e0288840. doi:10.1371/journal.pone.0288840.
31. Hatrongjit R, Boueroy P, Chopjitt P, Wongsurawat T, Jenjaroenpun P, Wankaew N, et al. Genomic characterization and virulence of *Streptococcus suis* serotype 7 sequence type 373 of clonal complex 94. Vet. Res. 2025;56:125. doi:10.1186/s13567-025-01562-4.
32. Rungelrath V, Ohlmann S, Alber G, Schrödl W, von Köckritz-Blickwede M, de Buhr N, et al. Survival of *Streptococcus suis* in porcine blood is limited by the antibody- and complement-dependent oxidative burst response of granulocytes. Infect. Immun. 2020;88:e00598-19. doi:10.1128/IAI.00598-19.

Figures

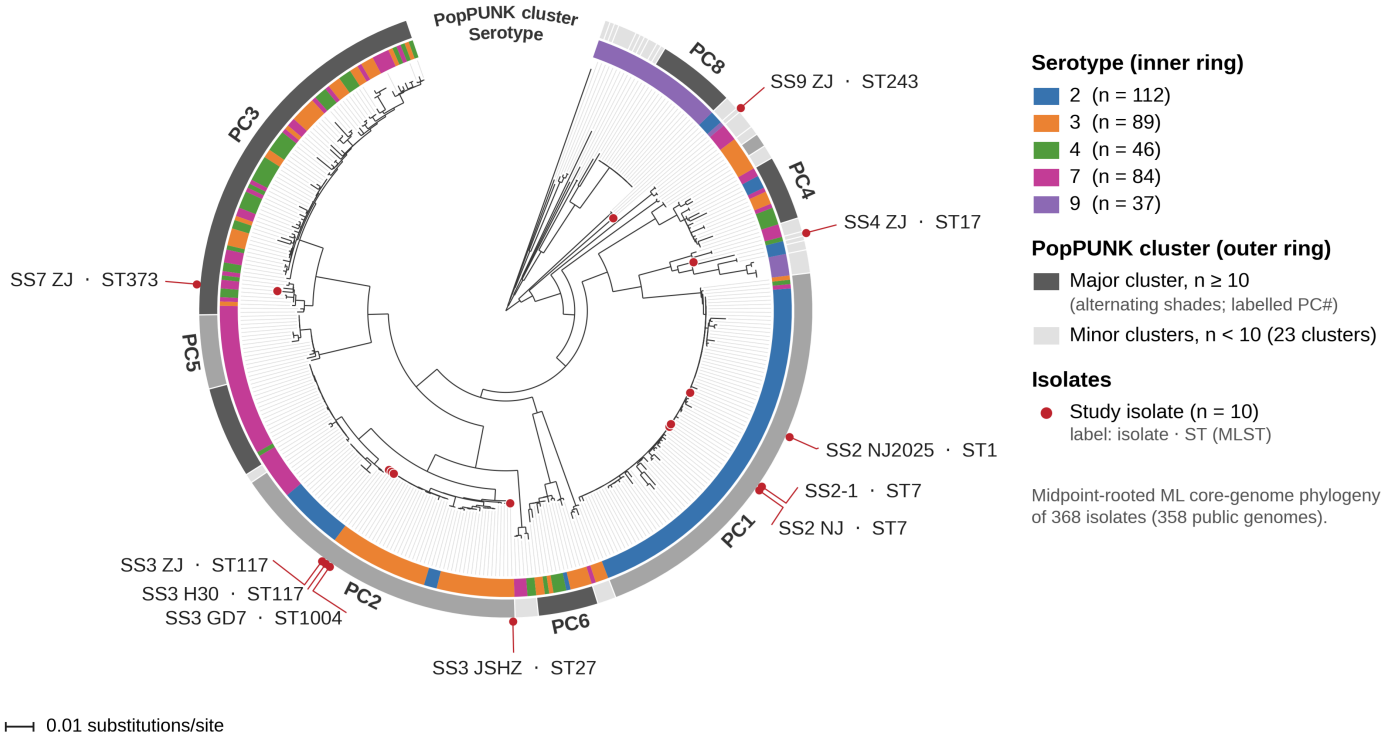

Figure 1

Public-genome phylogenetic context of the ten study isolates.

Maximum-likelihood core-genome phylogeny of 358 public *S. suis*genomes and the ten study isolates. Two annotation rings surround the tree: the inner ring indicates serotype, and the outer ring indicates PopPUNK cluster, with the seven clusters containing ten or more genomes shown in alternating dark and light gray and labeled PC1–PC6 and PC8, and the remaining 23 clusters shown in pale gray. Study

isolates are marked by red circles at the corresponding tips and are labeled with isolate name and ST. The tree is drawn with a midpoint display root, which sets orientation only and preserves every split and branch length; no biological outgroup is implied and the topology is interpreted as unrooted. The phylogeny was inferred without additional recombination masking. The scale bar represents 0.01 substitutions per site.

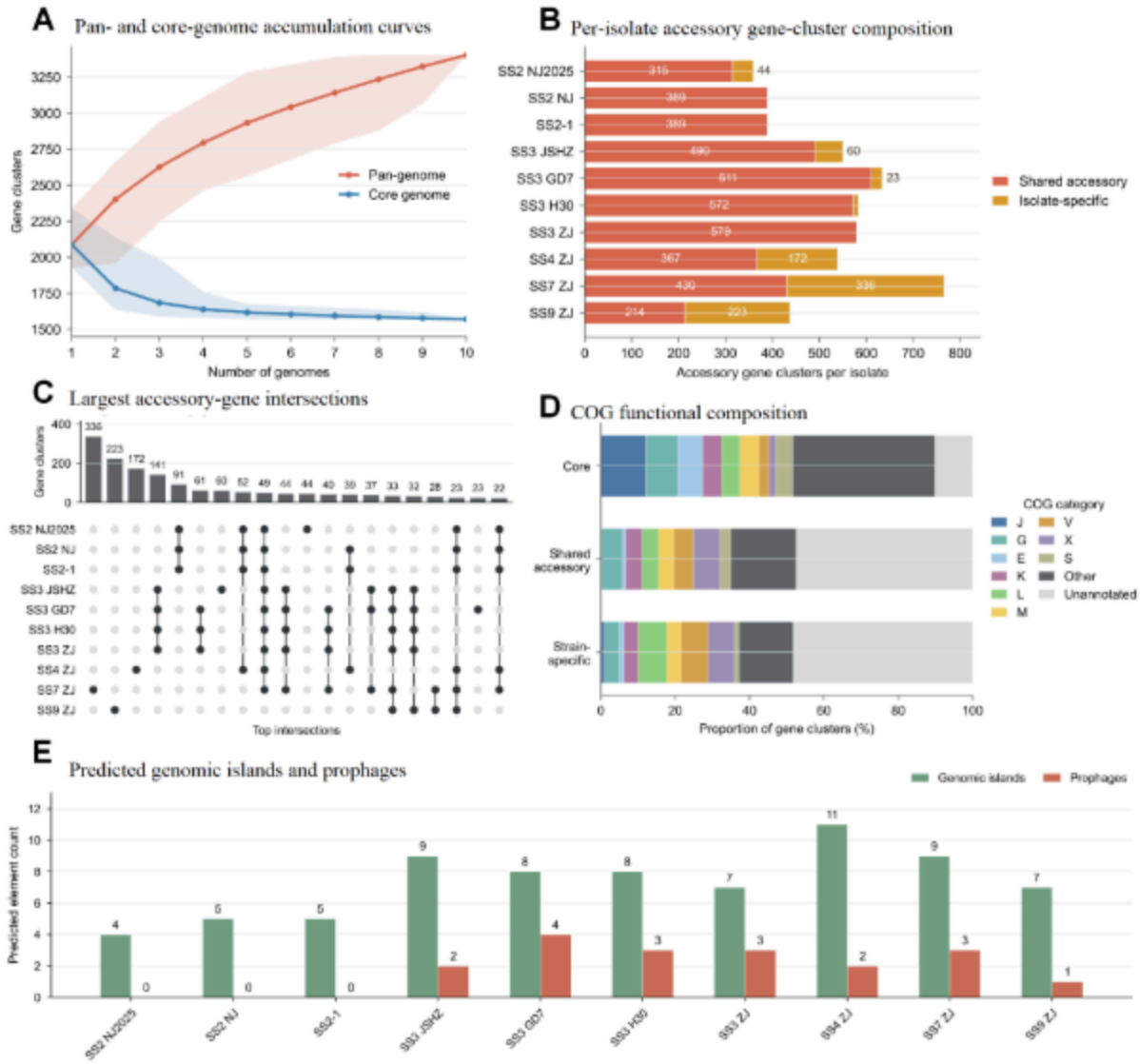

Figure 2

Pan-genome composition and predicted mobile elements in the ten study isolates.

(A) Pan- and core-genome accumulation across 1000 randomized genome orders; lines show means and bands show the 2.5th–97.5th percentiles. (B) Shared accessory clusters occur in two to nine isolates; isolate-specific clusters occur in one. (C) The 20 largest accessory-cluster intersections. (D) Descriptive COG category proportions. J, translation; G, carbohydrate metabolism; E, amino acid metabolism; K, transcription; L, replication and repair; M, cell envelope biogenesis; V, defense; X,

mobilome; S, function unknown. (E) Genomic islands and prophages predicted by IslandPath-DIMOB and Phigaro.

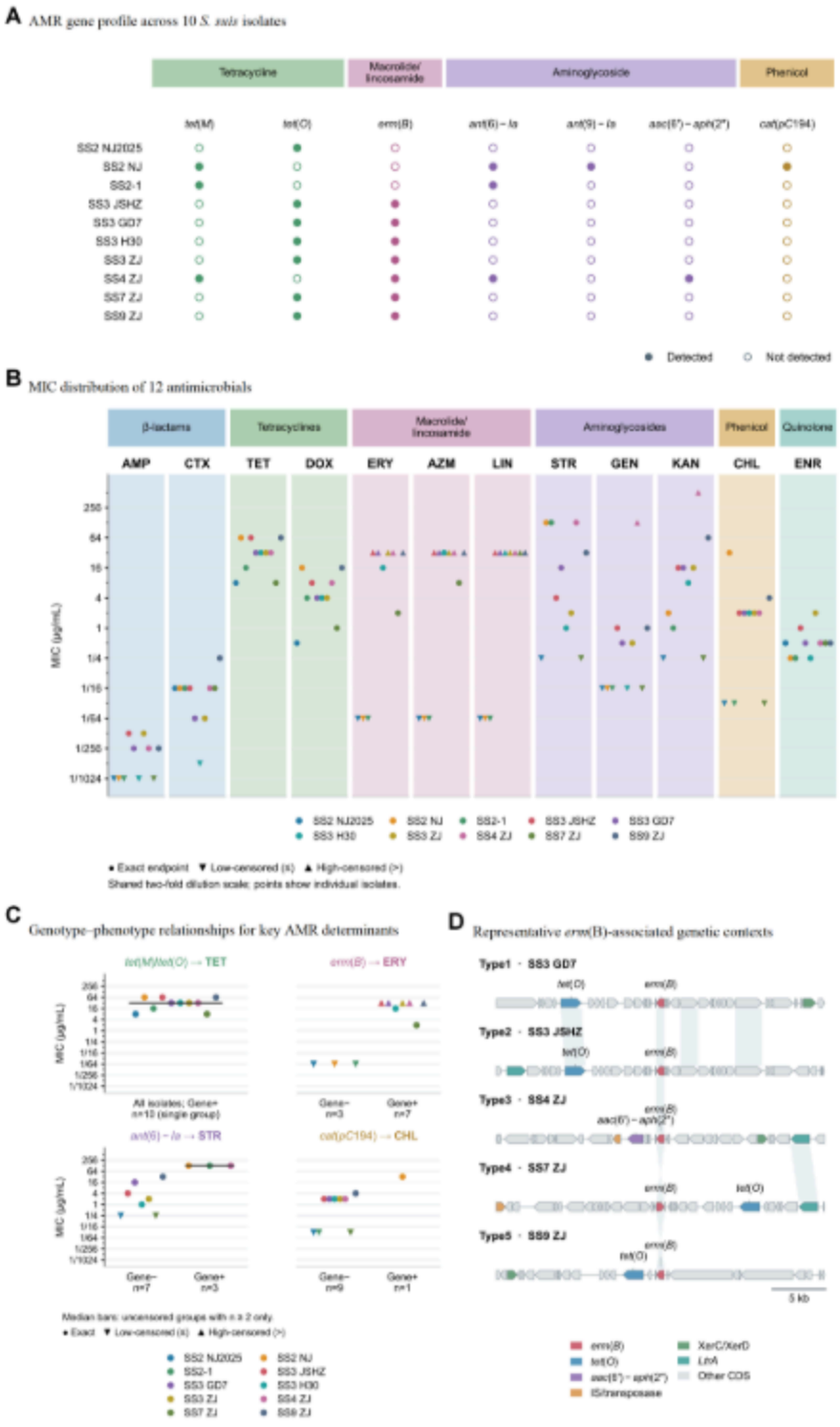

Figure 3

Acquired resistance genes and antimicrobial MICs in the ten study isolates.

(A) Acquired resistance-gene calls at ≥90% identity and ≥80% reference coverage. (B) Isolate-level antimicrobial MICs; censored values retain their ≤ or > boundaries. (C) Selected genotype–MIC comparisons. (D) Five *erm(B)*-associated genomic-context types; arrows indicate annotated genes and their orientation. AMP, ampicillin; CTX, cefotaxime; TET, tetracycline; DOX, doxycycline; ERY, erythromycin; AZM, azithromycin; LIN, lincomycin; STR, streptomycin; GEN, gentamicin; KAN, kanamycin; CHL, chloramphenicol; ENR, enrofloxacin.

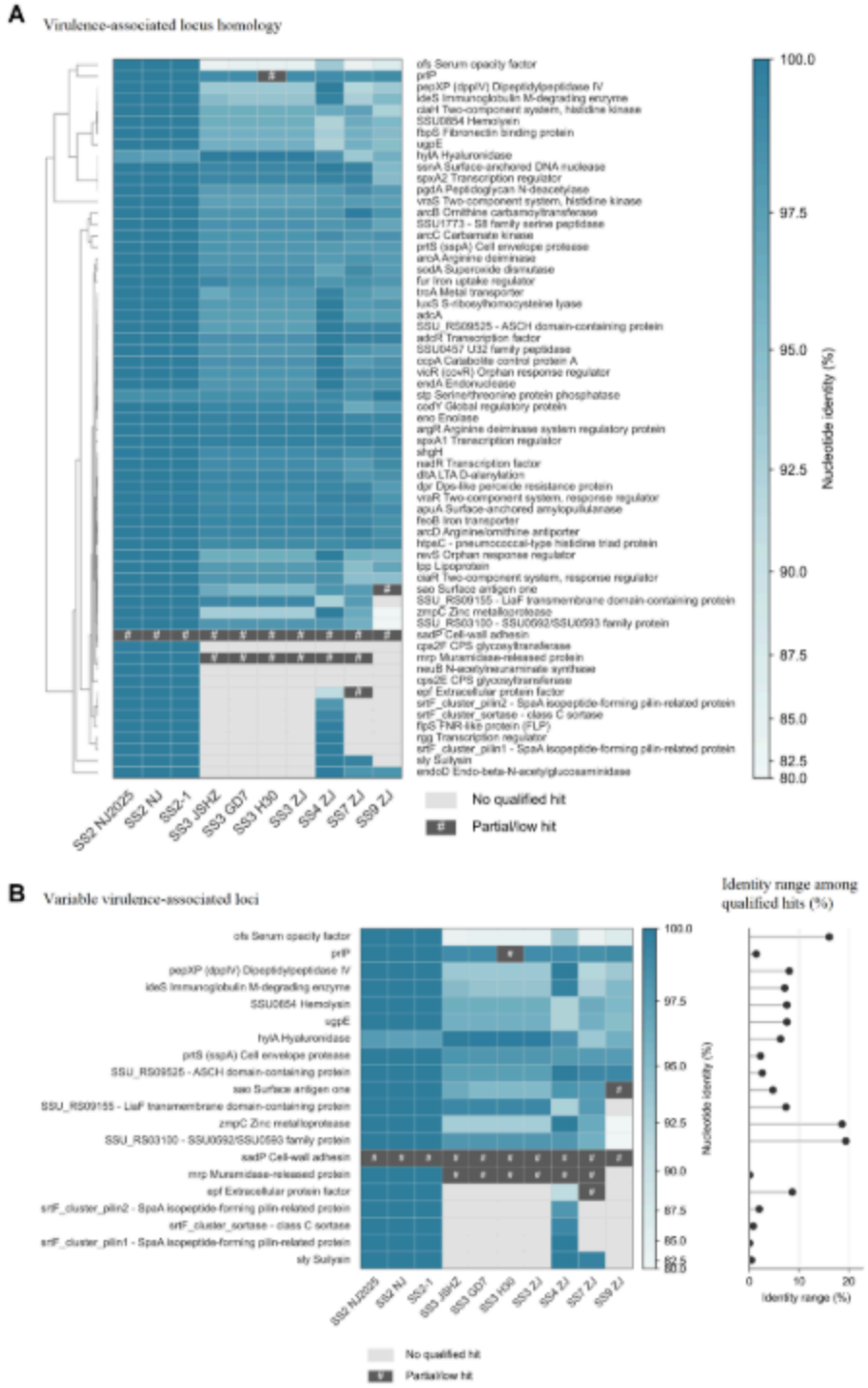

Figure 4

Nucleotide homology of 63 virulence-associated loci across the ten study isolates.

(A) Nucleotide homology across 63 virulence-associated reference loci. Qualifying matches require ≥80% identity and ≥80% query coverage. The color scale represents nucleotide identity; light gray indicates no qualifying match, and dark gray with # indicates a partial or low-quality hit. (B) Selected variable loci across the ten isolates. The dendrogram summarizes locus-profile similarity rather than isolate phylogeny. Reference accessions and sequence lengths are listed in Supplementary tables, Table S5C.

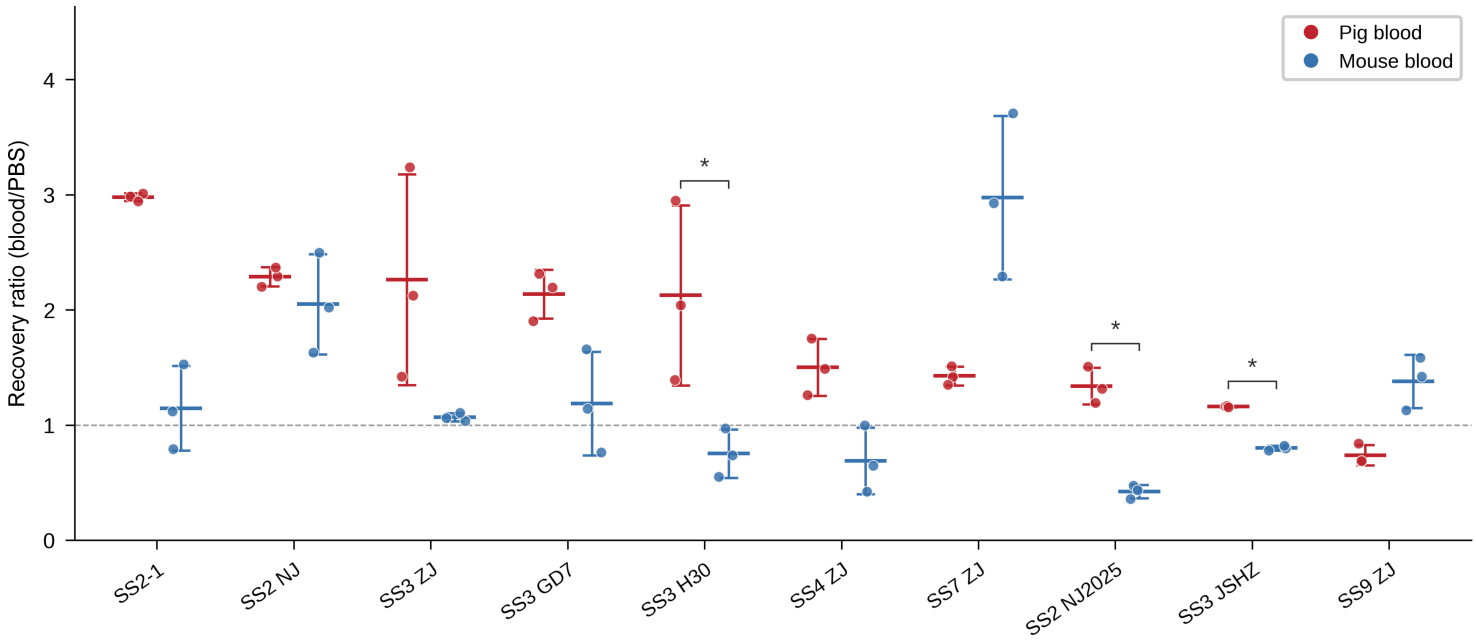

Figure 5

Whole-blood recovery of the ten *S. suis* study isolates.

Each point represents a single experiment (n = 3 per isolate per blood source). Horizontal bars show mean ± SD. The dashed line at 1.0 marks equal recovery in blood and PBS. Asterisks denote significant pig-versus-mouse differences within an isolate (Holm-adjusted paired t-test: *P < 0.05, **P < 0.01). Isolates are ordered by descending mean pig-blood recovery.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarytables.xlsx