MICROBIAL GENOMICS

RESEARCH ARTICLE

Bah *et al.*, *Microbial Genomics* 2026;12:001696

DOI 10.1099/mgen.0.001696

MICROBIOLOGY SOCIETY

OPEN DATA OPEN ACCESS

Molecular analysis of Lancefield group C/G streptococci causing human infections in Sheffield, UK

Saikou Y. Bah[1,2], Henna Khalid[1,2], Sona Jabang[1,2], Roy R. Chaudhuri[1,2], Lisa Tilley[3], Luke R. Green[2,4], David Partridge[2,3,5], Thushan I. de Silva[2,6] and Claire E. Turner[1,2,*]

Abstract

Lancefield group C/G streptococci (GCS/GGS) are increasingly recognized as significant human pathogens that cause a disease spectrum similar to *Streptococcus pyogenes*. Despite their high clinical burden in the UK, their genomic diversity remains poorly understood. We performed whole-genome sequencing on a prospective collection of 109 consecutive GCS/GGS isolates from any infection types in Sheffield, UK, over 5 months in 2020. *Streptococcus dysgalactiae* subsp. *equisimilis* (SDSE) accounted for 104 isolates, while 5 were identified as *Streptococcus canis*. The SDSE population was highly diverse, comprising 15 genomic clusters and 38 unique *emm*-ST combinations. We identified the presence of the ST20/stG62647 international lineage (24% of isolates), a cluster globally associated with severe invasive disease. Antimicrobial resistance genes were prevalent (49%), predominantly linked to mobile genetic elements carrying tetracycline and macrolide resistance. Furthermore, a variation in the penicillin-binding protein PBP2X (P601L) was linked to reduced penicillin sensitivity (minimum inhibitory concentration 0.03 mg l^{-1}). There were few or no genetic changes in isolates obtained from the same patient, even when they were collected 8–10 weeks apart, indicating long-term persistence within a host. The unexpected detection of *S. canis* in human infections and the high diversity of SDSE, persistence and virulence-associated regions underscore the need for enhanced national genomic surveillance to track emerging virulent and antibiotic-resistant SDSE lineages.

Impact Statement

Lancefield group C and G streptococci, most often the species *Streptococcus dysgalactiae* subsp. *equisimilis* (SDSE), is an increasingly significant human pathogen, often mirroring the severity of infections caused by Lancefield group A *Streptococcus* (*Streptococcus pyogenes*). Despite its clinical importance, we know little about the population of SDSE circulating in the UK. This study provides the first comprehensive genomic analysis of SDSE isolates from a single UK region, identifying a highly diverse population comprising 15 distinct genomic clusters but with evidence of long-term persistence within a single host. Notably, we confirm the presence of the international *stG*62647/ST20 lineage in the UK, which is globally associated with severe invasive disease. Our findings also reveal a high prevalence of antimicrobial resistance genes (~49%), primarily linked to mobile genetic elements, and the presence of a specific variation in the penicillin-binding protein PBP2X that reduces penicillin sensitivity.

[*Continued on next page*]

Received 02 February 2026; Accepted 28 March 2026; Published 11 May 2026

Author affiliations: [1]School of Biosciences, University of Sheffield, Sheffield, UK; [2]The Florey Institute of Infection, University of Sheffield, Sheffield, UK; [3]Laboratory Medicine, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK; [4]School of Clinical Dentistry, University of Sheffield, Sheffield, UK; [5]Sheffield NIHR Biomedical Research Centre, Sheffield, UK; [6]School of Medicine and Population Health, University of Sheffield, Sheffield, UK.
***Correspondence:** Claire E. Turner, c.e.turner@sheffield.ac.uk

Keywords: antimicrobial resistance ; clinical isolates; stG62647; *Streptococcus dysgalactiae* subsp *equisimilis*; *Streptococcus canis*; whole-genome sequencing.

Abbreviations: ADI, arginine deiminase; COVID-19, coronavirus disease 2019; FCT, fibronectin-collagen binding-T antigen; GC-1, genome cluster 1; GC-2, genome cluster 2; GCS/GGS, Lancefield group C/G streptococci; MIC, minimum inhibitory concentration; MLST, multi-locus sequence types; ONT, Oxford Nanopore Technologies; PBP, Penicillin Binding Protein; PopPUNK, Population Partitioning Using Nucleotide K-mers; SDSE, *Streptococcus dysgalactiae* subsp. *equisimilis*; *sil*, *Streptococcus* invasion locus; SLO, streptolysin O; ST, sequence type; TF, trigger factor; THB, Todd–Hewitt broth; UKAS, United Kindgom Accreditation Service; UKHSA, UK Health Security Agency; WHO, World Health Organization.

All supporting data, code and protocols have been provided within the article or through supplementary data files. Five supplementary figures and two supplementary datasets are available with the online version of this article.

Additionally, the unexpected detection of *Streptococcus canis* in human infections rather than animals highlights a need for monitoring. By defining the UK's SDSE population structure and its resistance landscape, this research underscores the critical need for enhanced national genomic surveillance to track emerging high-virulence and antibiotic-resistant lineages

DATA SUMMARY

Sequence files for isolates from Sheffield used for this study have been uploaded to the sequence read archive with project accession number PRJNA1333937 and accession numbers provided in Dataset S1. The completed genome for SDE096 has been deposited on GenBank with the accession number JBSXMJ000000000.

INTRODUCTION

Streptococcus dysgalactiae subsp. *equisimilis* (SDSE) is a beta-haemolytic *Streptococcus* that is closely related to the human pathogen *Streptococcus pyogenes*, sharing many virulence factors and mobile genetic elements with this species [1]. Although historically described to be the cause of infections in animals, over the past decade, increased incidence of SDSE in humans has been documented in many parts of the world and in some geographical regions, even overtaking *S. pyogenes* as a cause of invasive diseases [2, 3]. Like *S. pyogenes*, SDSE can cause a wide range of infections in humans, including mild and self-limiting throat, skin and soft-tissue infections but also much more severe and potentially lethal invasive infections, such as bacteraemia, necrotizing fasciitis and toxic shock syndrome [1, 4, 5].

SDSE can carry the Lancefield group carbohydrate C or G, but is now recognized to occasionally carry group A, confounding species identification using standard microbial methods, as group A was thought to be restricted to *S. pyogenes*. SDSE also carries an *emm* gene, encoding for the surface protein M. The hyper-variable region of the *emm* gene is used for *emm*-typing in the same manner as *S. pyogenes*, and to date, more than 100 *emm*-types have been identified for SDSE, with a few of these also identified in *S. pyogenes,* possibly due to homologous recombination [6, 7]. SDSE also has a fibronectin-collagen binding-T antigen (FCT) genomic region, which, similar to *S. pyogenes*, contains genes encoding for pilin structural proteins, pilus biosynthesis enzymes and other adhesins [1]. This region can be variable in gene content, and some SDSE isolates have also been shown to carry two FCT regions [8, 9].

Although group C and group G streptococci combined cause almost as many cases of bacteraemia in England and Wales as group A *Streptococcus* [10], nothing is known about the genetic diversity of SDSE in the UK. Recent data from other European countries and elsewhere have consistently identified a lineage with the *emm*-type *stG*62647 and multi-locus sequence type (ST) 20 as a predominant cause of disease, including severe and invasive diseases [6, 11–20]. It is not currently known if this lineage is circulating in the UK. To try and address the lack of knowledge regarding the SDSE population in the UK, we collected all Lancefield group C and G streptococci identified at the microbiology diagnostic laboratory at the Northern General Hospital, Sheffield, UK, in the summer (June–October) of 2020. This hospital diagnostic laboratory serves the community as well as all of the city's hospitals. We performed a comprehensive genetic characterization of these 109 group G/C streptococci, confirming the presence of the *stG*62647/ST20 lineage as a major cause of infection in addition to other diverse lineages.

METHODS

Isolates

All Lancefield group C/G streptococci that were routinely diagnosed from any infection type using Lancefield typing latex agglutination and MALDI-TOF by the Department of Laboratory Medicine, Northern General Hospital, Sheffield, UK, between June and October 2020, were stored (n=109). Laboratory Medicine is UKAS-accredited and the single regional microbiology diagnostic laboratory for Sheffield NHS trusts, as well as primary and community care, servicing around 600,000 adults and children. Isolates were fully anonymized, and no clinical or patient data was obtained apart from a general description of the original site of isolation (either skin, throat, blood, genital or others).

Whole-genome sequencing

Stored isolates were sub-cultured on Columbia horse blood agar (Oxoid) and then in Todd–Hewitt broth (THB) (Oxoid) for DNA extraction. DNA was extracted using a previously described method [21] and sent to MicrobesNG (https://microbesng.com) for short-read whole-genome sequencing. Sequencing libraries were prepared using the Nextera-XT library preparation kit (Illumina) and sequenced using the NovaSeq 6,000 platform, generating 250-bp paired-end reads. One isolate, SDE096, was subjected to Oxford Nanopore long-read sequencing. DNA was prepared for sequencing using the Oxford Nanopore Technologies (ONT) ligation sequencing kit (SQK-LSK109), according to the manufacturer's protocols. Libraries were sequenced with the ONT MinION Mk1C using rev C 9.4 flow cells and run for 24 h. Resulting trace files were analysed with the ONT Guppy

basecaller (version 3.2.10) using high-accuracy basecalling. A complete single-contig genome was assembled by combining the long-read sequence data with the short-read Illumina data using Unicycler (v0.5.1) [22]. The completed genome was submitted to GenBank for annotation (accession number: JBSXMJ010000000).

Bioinformatics analysis

Illumina reads were trimmed using Trimmomatic (v0.38) [23] with the settings: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36. Draft assemblies were generated using Unicycler (v0.5.1) [22] and assembly statistics checked with Quast (v5.3.0) [24]. All assemblies passed the quality control (all fewer than 120 contigs and with an overall length of 2.0–2.3 Mb). Sequence coverage and assembly statistics are provided in Dataset S1 (available in the online Supplementary Material). Mean coverage ranged from 32- to 348-fold. Prokka (v1.14.6) [25] was used to annotate the draft assemblies and core genomes determined using Panaroo (v1.5.2) [26] with strict mode. SNPs to infer genetic distances between assemblies were determined from the Panaroo-generated core genome alignment using snp-dists (v0.7.0) (https://github.com/tseemann/snp-dists). The sequences in the pan_genome_reference.fa generated by Panaroo were submitted to EggNOG (v2) [27] for functional annotation and grouped into cellular process, metabolism, information processing or uncharacterized. This output was visualized with a Sankey plot generated with Python Plotly. Maximum-likelihood phylogenetic trees were generated with RAxML (8.2.12) [28] using a general time-reversal substitution model with 100 bootstraps and evaluated and optimized with a GAMMA model of rate heterogeneity and then visualized using iTol (v7.2.1) [29].

Typing and cluster analysis

The *emm*-types were determined using the emm_typer.pl script available at https://github.com/BenJamesMetcalf/GAS_Scripts_Reference and the multi-locus sequence types (MLST) determined using the mlst_check from the Sanger Institute available at https://github.com/sanger-pathogens/mlst_check. All novel alleles and STs were submitted to the pubMLST database (https://pubmlst.org/organisms/streptococcus-dysgalactiae). Population Partitioning Using Nucleotide K-mers (PopPUNK) [30] was used to cluster the SDSE draft assemblies using a reference database available at https://www.bacpop.org/poppunk-databases/. For further analysis of genome clusters, reads for the isolates within this cluster were mapped to a reference genome or the *de novo* assembly of one of the cluster members, core SNPs were determined using Snippy (v4.6.0) (https://github.com/tseemann/snippy) and regions of recombination determined using Gubbins (3.1.0) [31] and visualized using plot_gubbins.R. Further genome sequence data for the *stG*62647/ST20 lineage were obtained from National Center for Biotechnology Information (NCBI) using the list of isolates identified to be within this lineage by Xie *et al.* [13] (Dataset S2).

FCT regions

To determine the FCT region, the location of some FCT genes was identified from the annotated *de novo* assemblies, and the surrounding full FCT region was then extracted. Extracted regions were subjected to a BLAST search of the entire protein NCBI database to identify all the genes within the region, and then the order of these genes was used to assign FCT gene arrangement pattern types compared with previously determined FCT regions [1]. Some samples had contig breaks in the extracted region and, therefore, a pattern could not be determined.

Virulence factor, regulator and antimicrobial resistance typing

The presence of 13 *S. pyogenes* superantigen genes [32], the SDSE version of *speG* (*speG*$_{dys}$) gene [33] and the DNase genes *sda1, sda2*, *sdn*, *spd1*, *spd3* and *spd4* were determined by BLAST (100% coverage, 80% identity) against *de novo* assemblies with representative gene sequences and manual checks where necessary. The CovR/S genes were extracted from the draft assemblies using *in silico* PCR (https://github.com/simonrharris/in_silico_pcr), translated to amino acids, and an in-house Python script was used to identify variants. Antimicrobial resistance (AMR) genes were identified using ABRicate (https://github.com/tseemann/abricate) with the NCBI AMRFinder plus database [34]. The penicillin-binding protein genes encoding PBP1A, PBP1B, PBP2A and PBP2X were extracted from the *de novo* assemblies and variants determined as for CovR/S. The presence of potential vaccine antigen genes encoding for C5a peptidase (ScpC/ScpG), streptolysin O (SLO), arginine deiminase (ADI), SpyAD, SpyCEP and trigger factor (TF) was determined by BLAST (80% coverage and identity) against draft assemblies using the antigen sequence database in Davies *et al.* [35].

Penicillin susceptibility testing

SDSE isolates were grown overnight in THB (Oxoid) at 37 °C with 5% CO_2. Cultures were adjusted to OD600nm=0.1 in fresh THB, before diluting further 1/100 in THB. Seventy-five microlitres of cultures were added to the wells of a 96-well plate (Corning), along with 75 µl of penicillin G (Merck) equating to final antibiotic concentrations ranging from 0.004 to 0.05 mg l^{-1} in 0.001 mg l^{-1} increments up to 0.01 mg l^{-1}, then 0.01 mg l^{-1} increments. Plates were incubated for 24 h at 37 °C with 5% CO_2. The minimum inhibitory concentration (MIC) was the lowest concentration of penicillin G that was found to inhibit the growth of each SDSE isolate.

RESULTS

Diversity within the population

After quality control filtering, we had *de novo* assemblies for 109 whole-genome-sequenced Lancefield group C/G isolates from 101 participants. Although all were obtained on different dates, up to 10 weeks apart, three isolates were from one patient, and two isolates were obtained from each of a further six patients. 104/109 isolates were determined to be SDSE, predominantly isolated from skin (*n*=60, 57.7%) or from genital swabs (*n*=20, 19.2%) with low numbers of blood isolates (*n*=7, 6.7%) or throat isolates (*n*=4, 3.8%) (Dataset S1). Unexpectedly, five isolates were determined to be *Streptococcus canis*, a species thought to be a rare cause of disease in humans [36], isolated from skin (*n*=2) or other sites (*n*=3) and all from different individuals.

Pangenome analysis of the 104 SDSE genomes identified 4,172 genes in total, with a core genome size of 1,602 genes (Fig. 1). Orthologous grouping indicated that 47.1% of accessory genes were uncharacterized compared to 23.5% of the core genome. A large proportion of the core genome (38.7%) was predicted to be involved in metabolism, compared to just 11.0% of the accessory genes, but an almost equal proportion of core and accessory genes (14.9% and 14.5%, respectively) was predicted to be involved in cellular processes (Fig. S1).

Twenty different *emm*-types were identified in the 104 SDSE isolates; 1 isolate was deemed 'non-typable' due to a deletion in and around the *emm* gene locus. Those isolates obtained from the same patient were the same *emm*-type and MLST, so, therefore, only a single isolate was used as representative per patient for further analysis. The *emm*-type *stG*62647 was the most common, accounting for 24.0% (*n*=23/96) of the isolates, followed by *stC*74A, accounting for 18.8% (*n*=18/96) of the isolates. MLST was also determined for the isolates, and 27 different STs were identified. ST17 was the most common (34.4%), and eight different *emm*-types were found with this ST. ST20 was the second most prevalent type, accounting for 24.0% of the isolates, but only one *emm*-type was associated with this ST (*stG*62647). Overall, there were 38 different *emm*-MLST combinations. Core gene phylogeny identified that, while the *stG*62647/ST20 isolates formed a single discrete lineage, multiple *emm*-types and multiple STs were found in other lineages (Fig. 2). This finding is consistent with that of others [5], that *emm*-type and MLST are not good markers for the population structure of SDSE. We, therefore, additionally assigned genomic clusters to the population using PopPUNK. This produced 15 genomic clusters, 8 of which consisted of just 1 isolate. For the 12 *emm*-types that were found in 2 or more isolates, only 4 (*stG*10, *stC*839, *stC*5345 and *stG*62647) were uniquely assigned to a single cluster, confirming the diversity within *emm*-types. There was no association between infection type and genome cluster, but two genome clusters (1 and 2) were the most dominant.

Antimicrobial resistance gene and penicillin-binding protein gene variations

Tetracycline resistance genes were the most prevalent AMR genes identified within SDSE, with 28/96 (29.2%) isolates carrying *tetM*, a further five carrying *tetO* and one additional isolate with *tetW* (Fig. 2). Macrolide resistance genes *ermA* and *ermB* were

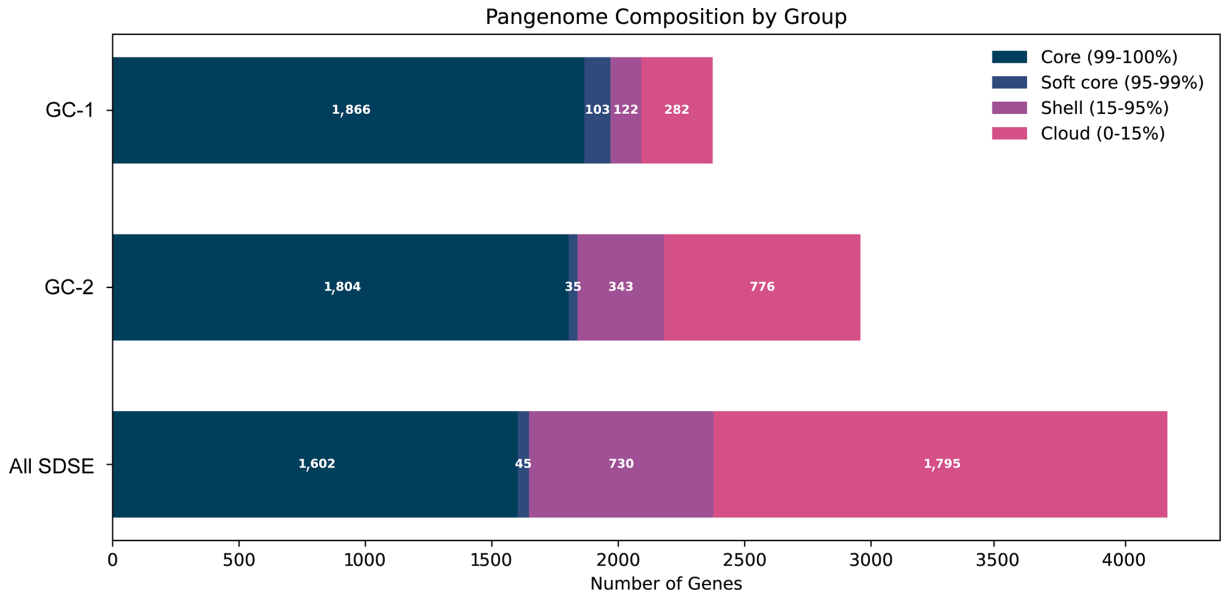

Fig. 1. Pangenome analysis. Core and accessory genes were determined by Panaroo for all 104 isolates, as well as the genomic cluster 1 (GC-1) and genomic cluster 2 (GC-2) isolates. The total number of genes for all the SDSE isolates, GC-1 and GC-2 was 4,172, 2,373 and 2,994 genes, respectively.

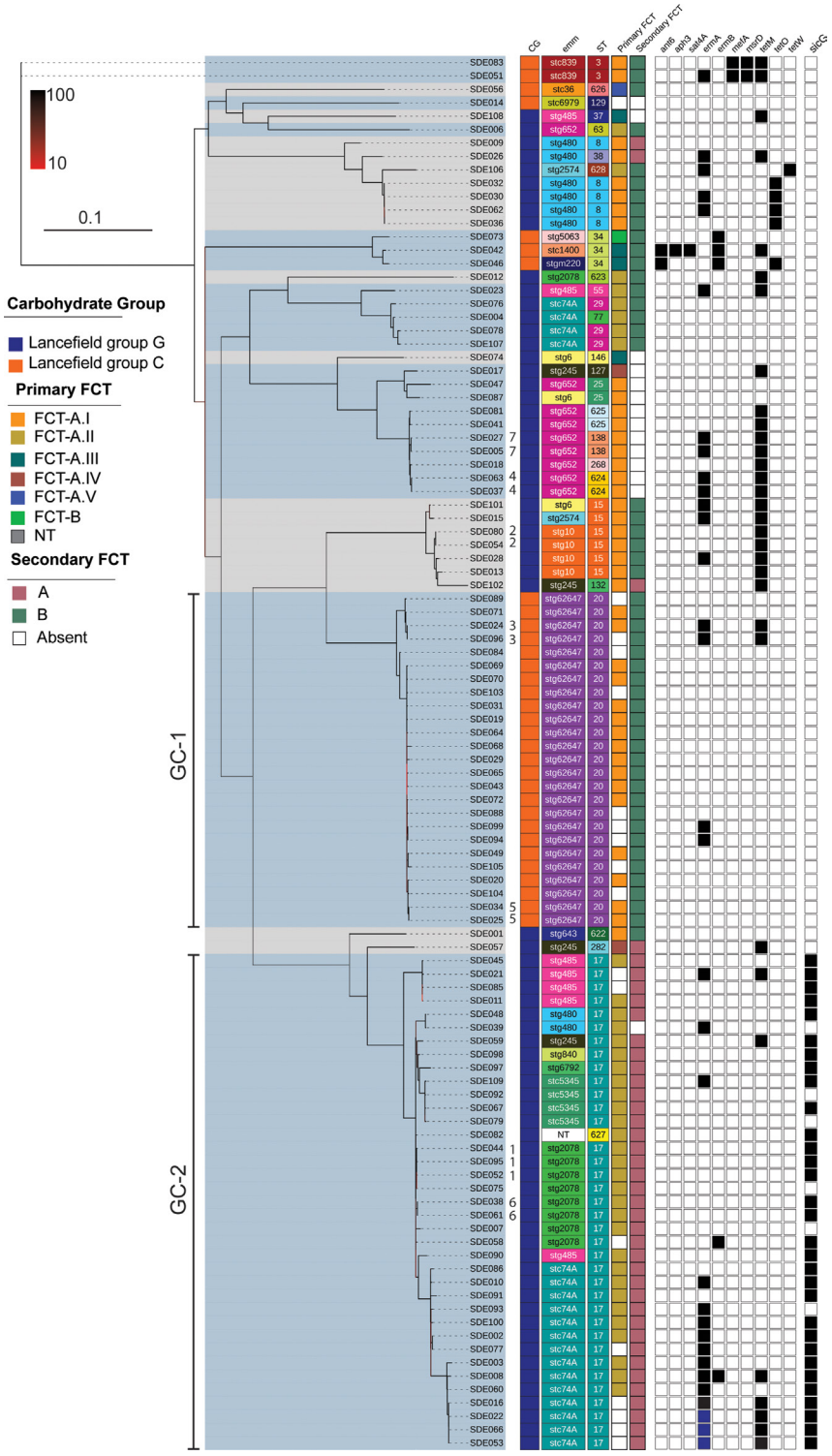

Fig. 2. Phylogeny of the 104 Sheffield SDSE isolates. A maximum-likelihood phylogenetic tree was constructed from the core genome SNPs using RAxML with 100 bootstraps. While Lancefield grouping of C (orange) or G (blue) was associated with PopPUNK genomic clusters (alternating grey/blue shading), *emm* and ST overall were not. GC-1 and GC-2 are labelled. The different primary and secondary FCT types are indicated, as is the presence (black) of AMR-associated genes and the gene *sicG*. Isolates from the same patient are indicated by numbers 1–7, each number representing a different patient. Three isolates had two copies of *ermA* (blue). NT, non-typable. The scale bar (0.1) represents the number of nucleotide substitutions per site. Bootstrap scale (10–100) for branch colours is also provided.

found in 29 (30.2%) and 5 (5.2%) of isolates, respectively. These genes were found across the population and not restricted to single genome clusters, although three isolates in the same genomic cluster had two copies of *ermA* (Fig. 2). The genes *ant6*, *aph3* and *sat4A* were together only in one isolate, although an additional isolate had *ant6* alone, and *mefA* and *msrD* were found together in two isolates of the same genome cluster (Fig. 2).

Variations within penicillin-binding proteins have been previously identified in *S. pyogenes* that conferred reduced sensitivity to beta-lactam antibiotics [37–40]. To determine if any variants exist in SDSE, we extracted and compared the penicillin-binding protein sequences (PBP1A, PBP1B, PBP2A and PBP2X) to identify any amino acid variations, focusing on the transpeptidase domain (Dataset S1).

A number of different variations were found within PBP1A, including a variable number of tri-amino acid repeats towards the end of the protein. However, none of the 11 different variants identified were within the transpeptidase domain (residues 335–632, pfam: PF00905), and most of the combinations of variations were associated with genome clusters. This was also the case for PBP1B, including a common pattern of 16–17 amino acid variants found in clusters of isolates, but none of these variants were within the transpeptidase domain (399–635, pfam: PF00905). However, one isolate, SDE056, did have a variation within this domain (I413V).

Within the transpeptidase domain of PBP2A (422–685, pfam: PF00905), we found a common variation of T488M in all but one of the *stG*62647/ST20 genome cluster isolates, as well as ten other isolates from other clusters. One of these *stG*62647/ST20 genome cluster isolates also had A592G, and another one (of a pair from the same patient) had T561I. All but two ST17-containing genome cluster isolates had both T464A and I481V. These variations were also found in other isolates (Dataset S1).

PBP2X was by far the most conserved with only six variations: V90L, D262N, A397V, R424Q, Q462K and P601L; the latter four are within the transpeptidase domain (292–603, pfam: PF00905). P601L has been previously reported in *S. pyogenes* and is associated with a reduction in penicillin sensitivity [40]. This variation was found in the pair of *stG62647*/ST20 isolates from the same patient. The Q462V variation was found in all but 5 of the 34 ST17/ST627 isolates (genome cluster 2). This change has not been previously reported, although the variant Q462H was not associated with a change in sensitivity to penicillin in *S. pyogenes* [40].

To determine if any of the variations we identified in any of the *pbp* genes altered the sensitivity to penicillin, we measured the MIC to penicillin G. All isolates had an MIC of 0.02 mg l^{-1} or lower, apart from two isolates that had an MIC of 0.03 mg l^{-1}; these two isolates, from the same patient, had the P601L variation in PBP2X, suggesting this variation slightly reduces the penicillin sensitivity of SDSE.

FCT pattern

The FCT genomic region, which contains pilin genes and other adhesin genes, can be highly variable and has been implicated in determining tissue tropism in *S. pyogenes* [41]. The arrangement of the genes within the FCT region determines the FCT type. While only one FCT region has been identified in *S. pyogenes,* two FCT regions have been found in SDSE isolates [1, 8, 9]. In our SDSE isolates, the primary FCT loci had a gene arrangement similar to *S. pyogenes* FCT, all starting with the transcriptional regulator *rofA* (Fig. 3). Six different patterns were identified, with five variants (I–V) based on the previously defined FCT-A [1, 9] depending on the variable presence of a fibronectin binding protein gene (*gfbA*) downstream of the transcriptional regulator, sortase genes and the pilin adhesin gene *pilA*. All FCT-A variants carried the pilin linker gene *pilL*, the pilin backbone gene *pilB* and a second fibronectin binding protein (*fbpZ*). FCT-B was found in a single isolate and was previously defined by the lack of *fbpZ* and a different order of the *pil* genes [1, 9]. The most common, found in the largest number of isolates, were FCT-A.I and FCT-A.II, without and with *gfbA,* respectively. Broadly, FCT types were associated with genome clusters (Fig. 2).

A secondary FCT region was found in 90/104 isolates in two different variations, depending on the presence of one copy of *srtB* and three pilin genes (variant A) or two *srtB* and four pilin genes (variant B) (Fig. 4).

CovR/S

In *S. pyogenes,* the CovR/S two-component system controls the expression of ~15% of the genome, and mutations in CovR or CovS are associated with hyper-virulence [42]. A similar role for CovR/S has been identified in SDSE, repressing major virulence factors and potentially contributing to more severe disease [43–45]. We found six isolates with unique variations in CovR: E9D, I50F, V88M, P103L, Y160F and S164G. These six isolates all belonged to five different genomic clusters, suggesting they were spontaneous variations rather than lineage-associated and were either from skin infections (n=5) or other infections (n=1). Fourteen different amino acid change variants were identified in CovS in 17/96 (17.7%) of the isolate genomes, although combinations of some appeared to be associated with genomic clusters and we did not find any premature stop codons.

DNases and superantigens

S. pyogenes can carry a combination of 13 different superantigens [32], some associated with bacteriophages. The only superantigen we identified in our SDSE isolates was $speG_{dys}$, which is the variant *speG* superantigen gene found in SDSE and differs from *speG*

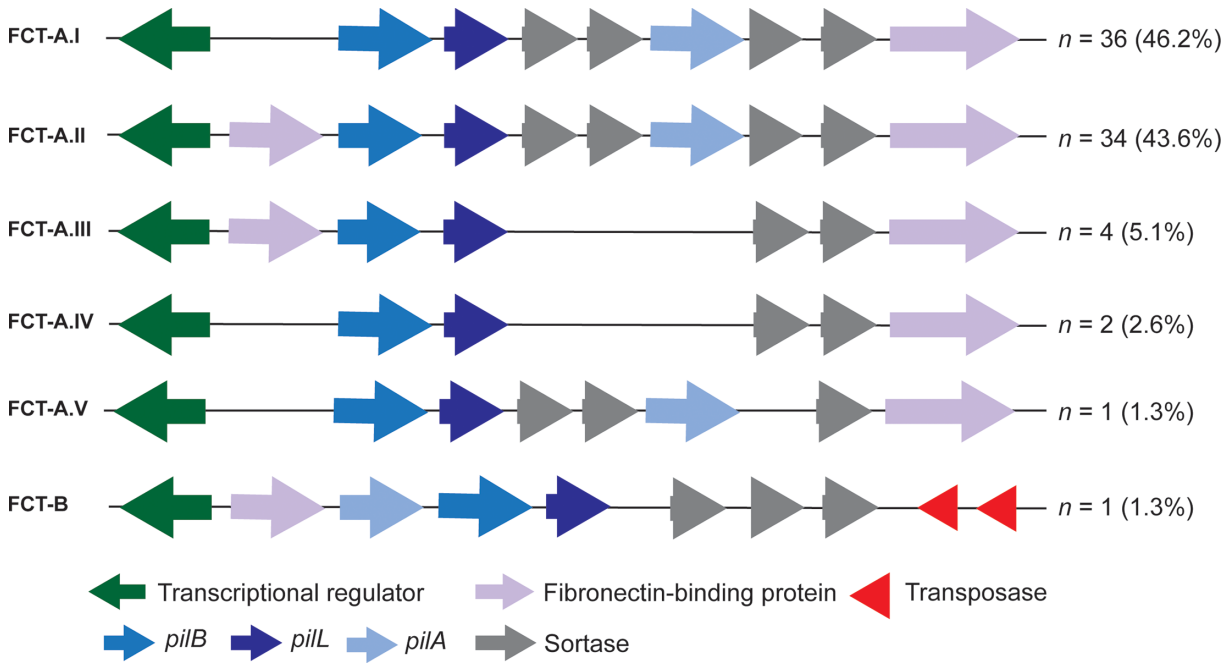

Fig. 3. Gene arrangement of the primary FCT locus in SDSE. The FCT region was extracted within single contigs for 78/96 *de novo* genome assemblies. Five different variants of FCT-A (I–IV) and one variant of FCT-B were identified, and the number of isolates (%) with these variants is indicated on the right.

found in *S. pyogenes* (~85% identity). It was present in 48 isolates, restricted to all isolates within ten genomic clusters. *S. pyogenes* can also variably carry prophage-associated DNases, and so we tested for the presence of *sda*, *spd1*, *spd3*, *spd4* and *sdn*. We found *sda*, *sdn* and *spd3* in three, one and three isolates, respectively, although one isolate carried both *sda* and *spd3*.

Diversity of the genome cluster 1 (*stG62647/ST20*)

Although it was the second most common cause of infections in our sampling, the genome cluster 1 (GC-1), which consists exclusively of ST20 and *stG*62647, has in other countries been a highly prevalent cause of infections, associated with severe disease, including streptococcal toxic shock syndrome and necrotizing soft tissue infections [6, 14, 15, 46]. Here, we found that the majority were associated with skin infections (13/25), including two isolate pairs from the same patients, all skin, and none were blood isolates.

Consistent with the conserved ST and *emm* within GC-1, there was a fairly low level of recombination (Fig. S2) and a low level of accessory genome content (Fig. 1). They also all had, where we could determine the type, primary FCT-A.1V and the secondary FCT region variant B.

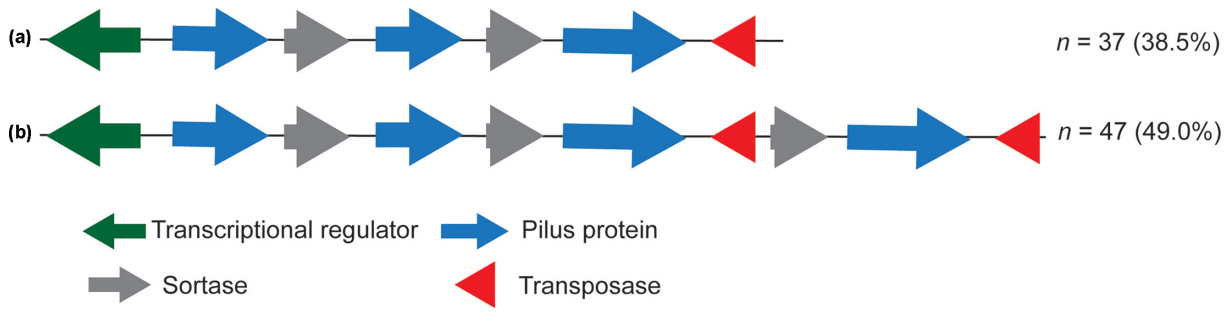

Fig. 4. Gene arrangement of the secondary FCT locus in SDSE. Two variants of the secondary FCT region were identified in 84/96 isolates and the number of isolates (%) with these variants is indicated on the right.

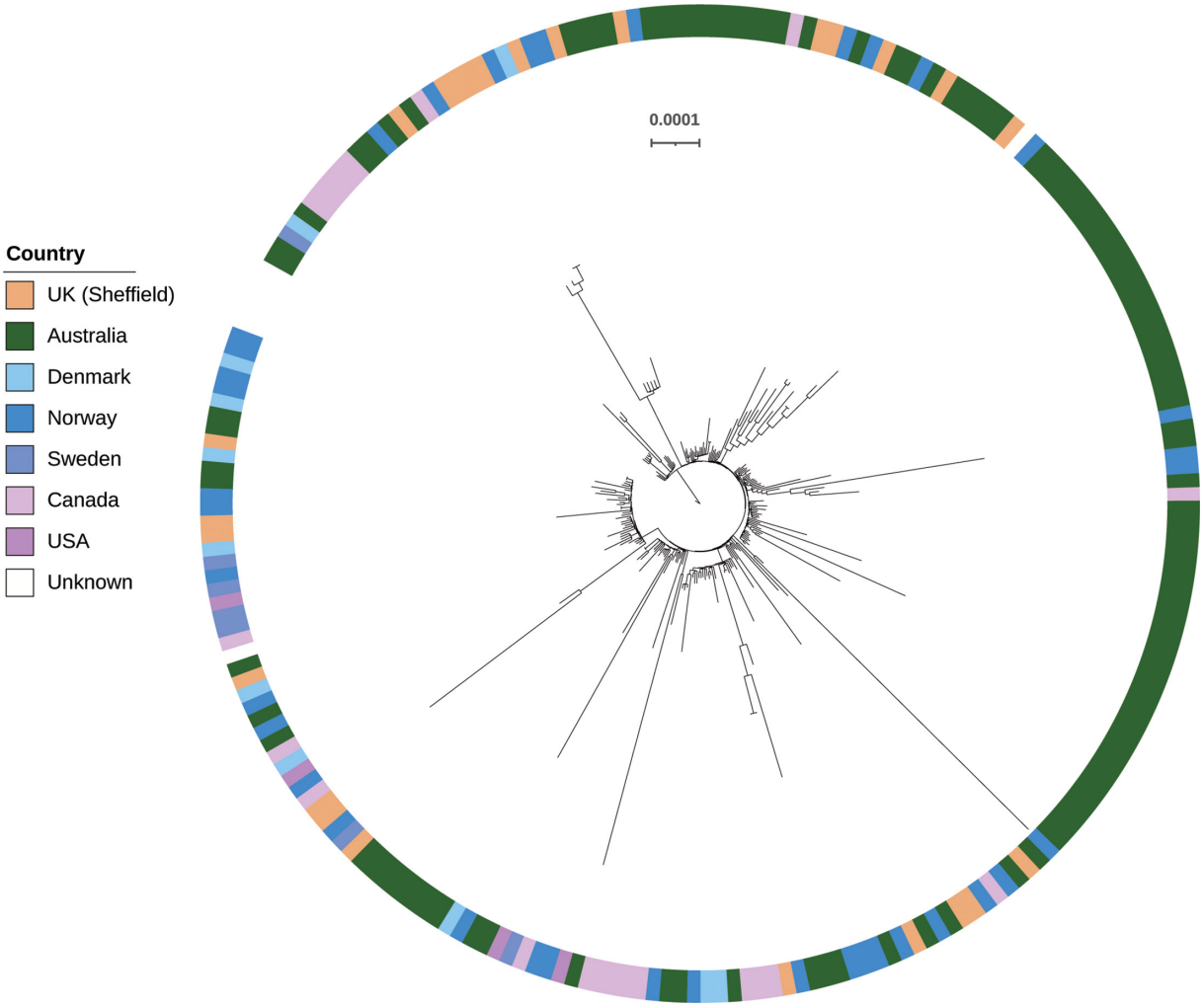

Fig. 5. A phylogenetic tree of genome cluster 1 isolates. A maximum-likelihood phylogenetic tree was constructed using RAxML [28] from core genome SNPs after mapping to the reference genome SDE096. Short-read sequence data were obtained from NCBI SRA using the list of international isolates that belong to this lineage, defined by Xie *et al*. [13]. The ring indicates the country that the isolates were collected from: Australia (*n*=115) [13, 50], France (*n*=6) [19], Norway (*n*=74) [14, 20, 59], Denmark (*n*=11) [60], Canada (*n*=21) [16], the USA (*n*=4) [50, 61], Sweden (*n*=7) [62] and Unknown (*n*=2) (Rostock University Medical Centre, PRJEB34961) (Dataset S2). It also includes all 25 Sheffield isolates. The scale bar represents substitutions per site.

Comparison with other publicly available *stG*62647/ST20 isolates identified that our Sheffield *stG*62647/ST20 GC-1 isolates also belonged to this international lineage and were distributed throughout the tree, indicating no regional isolation (Fig. 5). We sequenced one of our isolates, SDE096, with long-read sequencing to complete the genome. As previously identified in *stG*62647/ST20 isolates, SDE096 has a transposon inserted into the *silB* gene, part of the *S*treptococcus *i*nvasion *l*ocus (*sil*), mutations in which have been associated with more severe clinical outcomes [20]. All other Sheffield GC-1 isolates also had this insertion in *silB*. The *sil* locus was found in 58/96 (60.4%) isolates in total, but only GC-1 isolates had an insertion in *silB*, and other isolates had full-length *silB*, apart from three which had premature stop codons truncating SilB after 82, 99 or 139 aa, one of which was a blood isolate.

The SDE096 isolate was determined to carry the P601L variation in *pbp2X*, conferring reduced penicillin sensitivity. This isolate also carried *ermA* and *tetM*, which were found within a region between the conserved chromosomal genes *metQ* (D-methionine-binding lipoprotein, ACXAHW_06730) and *rlmD* (putative RNA methyltransferase, ACXAHW_07415), containing genes associated with plasmids and transposons (Fig. 6) as well as a predicted incomplete phage. This incomplete phage region, along with the *tetM* and surrounding genes, was also found in the published *S. pyogenes emm*230 genome [35] (Fig. 6). Analysis of this region between *metQ* and *rlmD* in other isolates determined it to be highly variable and frequently associated with resistance genes (Fig. S3). Of the 37 for which we could extract this region over a single contig, 3 had *ermA* and *tetM* genes within this region, 3 had

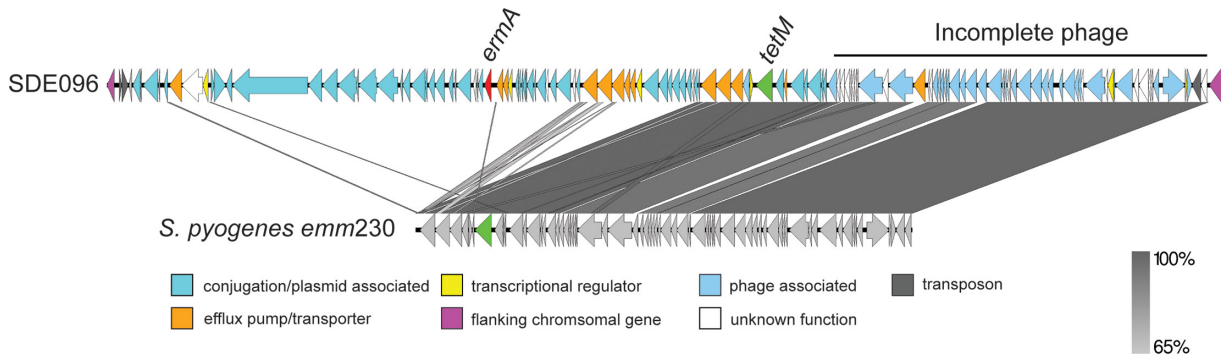

Fig. 6. Schematic representation of the *ermA* and *tetM* region in SDE096. The *ermA* (red, ACXAHW_06910) and *tetM* (green, ACXAHW_07100) genes of SDE096 were found within a region located between the conserved chromosomal genes SDE096 ACXAHW_06730 (*metQ*) and ACXAHW_07415 (*rlmD*). The region downstream of *tetM* contained a predicted incomplete phage genome, which shared homology with a region in *S. pyogenes emm*230 (NZ_CP035451.1). Figure generated using EasyFig. Percentage identity is indicated on the right.

tetM alone and 3 had *ermA* alone. Within the GC-1 (*stG*62647/ST20) lineage, only four isolates carried any resistance genes at all, although two were from the same patient, and all carried them within this region (Fig. S4).

Diversity of the genomic cluster 2 isolates

The genomic cluster 2 (GC-2) lineage, predominantly consisting of ST17 isolates, with one single locus variant ST627 and ST282, and all Lancefield group G, was the most common cause of infections (*n*=34/96 isolates, 35.4%, counting three isolates from patient 1 and two isolates from patient 6 as one each). Eight different *emm*-types were found in this cluster, along with one non-typable due to a deletion within the *emm* locus. Gubbins analysis for recombination identified 692 genes across the genome that showed evidence of recombination from a total of 2,958 genes and a total of 275 blocks of recombination (Fig. 7). As expected, given the diversity of *emm* genes in this cluster, *emm* was one of the genes. There was also a greater proportion of accessory genes in this cluster compared to GC-1 (Fig. 1).

All isolates for which we could determine the primary FCT type (26/34) were FCT-A.II (26/30) similar to *S. pyogenes* FCT-6 [1, 9] with two fibronectin-binding genes, *gfbA* and *fbpZ*, and the *pilB*, *pilA* and *pilL* genes. All but one isolate also had a secondary FCT region with another fibronectin-binding protein gene and other pilin-related genes. The *drsG*/*sicG* gene, which encodes for *d*istantly *r*elated to *S*ic (DRS) and confers resistance to antimicrobial peptides, was found in 28/34 isolates within this cluster and was not found in any isolates outside of this cluster (Fig. 2).

The prevalence of antibiotic resistance genes was higher in this lineage than for GC-1, with 17 (44.7%) of isolates carrying at least 1 gene. Fifteen isolates carried *ermA*, five of these also carried *tetM*, one isolate carried *ermB* in addition to *ermA* and *tetM*, one carried *tetM* alone and one carried *ermB* alone. Three isolates (SDE022, SDE066 and SDE053) carried two copies of the *ermA* gene, and both were located in the same region as *tetM*, within the *metQ*-*rlmD* variable region (Fig. S5). This region was the same for these three isolates, and the region surrounding one *ermA* was found in seven others and two more with some differences, but differed substantially from another that had both *ermA* and *tetM*.

Intrapatient variation

For six patients, we had two separate isolates, obtained between 3 and 74 days apart, and for an additional patient, we had three isolates, taken 4 and 36 days apart (Table 1). The three isolates from patient 1 and the two isolates from patient 6 were *stG*2078 and were present in GC-2 (Fig. 7). Mapping to the *de novo* assembly of the earliest isolate for each patient identified that the isolates from patient 6 (12 days apart) were identical, as were the first two isolates from patient 1 (3 days apart); however, the third isolate from patient 1 (36 days later) differed by 5 SNPs. For all isolates within GC-2, excluding regions of recombination (Fig. 7), there was a median and mean of 81 SNPs (range 0–144) between isolates, indicating that, as there was a low level of variation in the later patient 1 isolate compared to the earlier ones, it was likely to be persistence of the same isolate rather than reinfection with a different isolate. However, SDE075 from another patient was identical to the early patient 1 isolates and collected around a similar time, so it could potentially have been part of a transmission event. However, we do not have any epidemiological information to determine this.

The isolates from patient 3 and patient 5 were *stG*62647 in GC-1 (Fig. S2), but both pairs were identical to each other, despite the 57 days between the patient 3 isolates. The isolates from patient 4 and patient 7 were from another cluster, *stG*652 but different

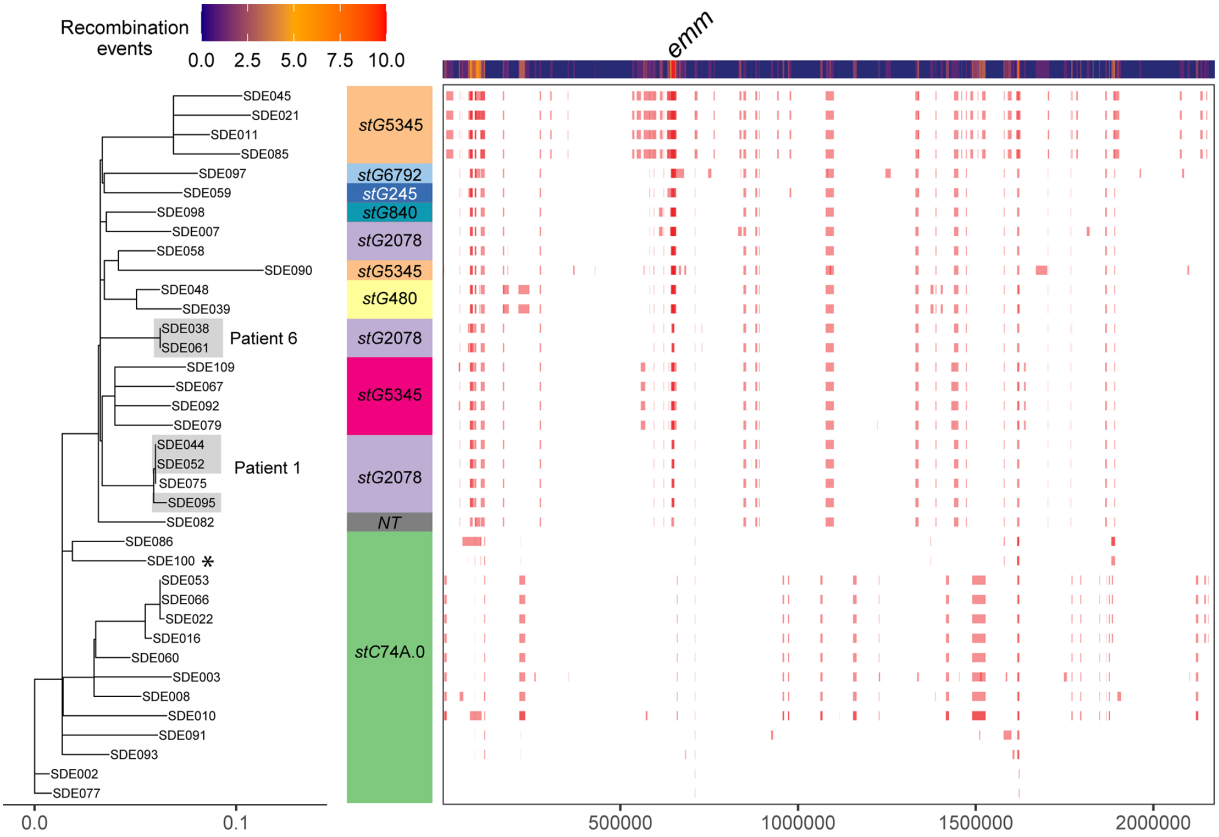

Fig. 7. Regions of recombination in the SDSE genome cluster 2. Regions of recombination were determined using Gubbins with the *de novo* assembly for SDE100 as a reference (*). The tree on the left is a maximum-likelihood tree generated from the Gubbins output (excluding recombination), the *emm*-type is given and on the right side are blocks showing recombination events with a heatmap of events above. The *emm*-type is given in coloured blocks. Three isolates from patient 1 and two isolates from patient 6 are shaded. SDE075 was not from patient 1 but was closely related to the isolates from this patient.

ST (Fig. 2), and although the pair from patient 4 were identical despite the 74-day separation, the two isolates from patient 7, 28 days apart, were two SNPs different, one non-synonymous (C123R) in *mntR* (a HTH transcriptional regulator) (Table 1).

S. canis isolates

The five S. *canis* isolates we identified were Lancefield group G, and all five were of a different ST, suggesting these were independent infections. One isolate carried two antimicrobial resistance genes (*ermA* and *tetO*), and another carried *tetO* alone.

Table 1. Genomic differences between isolates from the same patient

Patient	*stG*/GC	Days apart	Changes	Impact
Patient 1	*stG*2078/GC-2	3	None	None
		36	5 SNPs	Y586C in BtuD (vitamin B12 import ATP-binding protein) E150A in a prophage tail protein E140G in ResA (TlpA disulphide reductase family protein)
Patient 2	*stG*10/GC-4	22	2 SNPs	R166H in SagC (part of streptolysin S production)
Patient 3	*stG*62647/GC-1	57	None	None
Patient 4	*stG*652/GC-7	74	None	None
Patient 5	*stG*62647/GC-1	3	None	None
Patient 6	*stG2078*/GC-2	12	None	None
Patient 7	*stG*652/GC-7	28	2 SNPs	C123R in MntR (HTH transcriptional regulator)

Consistent with the different STs, core genome comparison indicated a large genetic distance between the isolates with a mean SNP distance of 8,063 (range 57–80,625).

The penicillin-binding proteins were also extracted from the *S. canis* assemblies and regions of variation identified. PBP1A had only two loci with variations A694T and G709S (Dataset S1). PBP1B had seven loci with variable amino acids N272K, V409I, A419T, V430V, S697P, A702T and N703P (Dataset S1). There were three loci, P68A, D127N and N764K, in PBP2A that had variable amino acids, all of which were only found in one isolate compared to the other four. Finally, PBP2X had three loci that differed at position A108T, A281T and T535I (Dataset S1). MIC testing of these isolates, however, did not indicate decreased sensitivity to penicillin.

Potential vaccine coverage

There has been a recent acceleration in efforts to generate a vaccine against *S. pyogenes* [47]. The non-M protein targets currently in the pipeline include antigens that may also be present in SDSE and *S. canis* and, therefore, could provide cross-species protection. We found all our SDSE isolates carried the genes encoding for potential vaccine antigens C5a peptidase (ScpC/ScpG), SLO, ADI, the surface-exposed adhesin SpyAD and TF, with greater than 97% identity to the *S. pyogenes* genes, except TF, which was only ~87% identical. However, none of the isolates carried the complete gene encoding for another vaccine antigen, the chemokine-cleaving cell envelope proteinase SpyCEP. All five *S. canis* isolates carried the genes encoding for ADI (~93% identity) and TF (~87% identity), but none of the other antigens. This suggests fairly good coverage against SDSE could be achieved by the various combination vaccines, excluding SpyCEP, but they may not protect against *S. canis*.

DISCUSSION

In this study, we investigated the genomic characteristics of SDSE isolates collected from any infection type in Sheffield, UK, as no genomic data on SDSE isolates from the UK have yet been published. We found that, similar to other countries in Europe, North America and Australia, the GC-1 *stg*62647/ST20 lineage was common, and there was a high level of diversity within other lineages.

Predominantly, our isolates came from skin infections, followed by genital swabs, with a relatively low level of throat swab isolates. However, our collection period was June to October 2020, which was immediately following the Coronavirus disease 2019 (COVID-19) pandemic lockdown period in the UK, and, therefore, transmission patterns may have been different, and throat swabs taken for microbiological identification of bacterial infections are likely to have been reduced. However, the number of bacteraemia cases notified to the UK Health Security Agency (UKHSA) for Lancefield group C and G during 2020 was no different from previous and post-years [10]. This is in contrast to group A streptococci, where notifications of bacteraemia were substantially lower during 2020 compared to previous years [10]. This was also the case for other countries as well [13, 15, 48, 49]. We do, however, recognize that this collection period is likely to be different from a time without an active global viral pandemic.

Our SDSE population was diverse with 15 different genomic clusters, 20 different *emm*-types and 27 STs indicative of high levels of recombination driving diversity. The most common genomic cluster (GC-2) comprised 8 *emm*-types and had a high level of recombination with 275 recombination events, with an average length of ~128 kbp. This was higher than GC-1, with 50 recombination events, and higher than that observed for major lineages of *S. pyogenes* [35, 50].

GC-1 (*stG*62647/ST20) was first reported in Argentina [18] but described in detail using whole-genome sequencing by Oppegaard *et al.* in Norway [20]. Although commonly reported as causing invasive disease in patients [3, 6, 12, 14, 20], other studies have shown no association with more severe outcomes in patients over other lineages [13, 16, 51] and varying mortality in mouse models of infection [19]. Our *stG*62647/ST20 isolates were part of the international lineage genetically predicted to have emerged in 1956 before rapidly increasing in the population [13]. Characteristically of this lineage, all our isolates carried the disrupted *silB*, which may influence virulence, although there are many other mechanisms influencing virulence in this lineage [52]. In our sampling period, we had very few isolates from blood infections (n=7), and none of these were GC-1 isolates; however, it is not known what the prevalence of this lineage is on a national scale and over a much wider time period.

Tetracycline resistance genes (*tetM*/*tetO*/*tetW*) were found in 36.5% (n=38) of isolates, which is a similar proportion to what we previously found in *S. pyogenes* from skin infections in Sheffield (37.3%) [53]. Other reports of tetracycline resistance in SDSE have been between 28% and 62% [6, 15, 17, 18, 51, 54]. The proportion of our SDSE isolates carrying *ermA* or *ermB*, potentially conferring macrolide resistance, was also high (35.6%) and ~5 times higher than in the Sheffield *S. pyogenes* population (6.5%) [53]. Similar levels have also been found in other studies [6, 18, 51]. Whilst group A streptococci (*S. pyogenes*) are now included on the World Health Organization (WHO) priority pathogen list because of the growing concern of macrolide resistance [55], there should also be concern about macrolide resistance in group G/C streptococci.

A high proportion of the *tet* and *erm* genes were present in a variable region between the conserved chromosomal genes *metQ* and *rlmD*. In SDE096 and other isolates, there were several genes in this region predicted to function as conjugal transfer and mobilization proteins, indicative of a plasmid or extrachromosomal element. A previous study did find an extrachromosomal circular element in an *stG*62647 strain that had an attachment site homologous to *rlmD*; however, it could not occupy that region

as there was already a different mobile genetic element in that site [19]. It is not clear if this region in any of our isolates could mobilize and replicate extrachromosomally or if there are additional extrachromosomal elements. For 37/104 of our isolates, the region assembled over a single contig, but for other isolates, it would require long-read sequencing to accurately determine the composition of this region, as well as any extrachromosomal elements. Some isolates carried resistance genes outside of these regions, and these were associated with highly variable ICEs that we did not describe in detail here. However, it does indicate that the acquisition of resistance genes by SDSE can be via many routes and with several other bacterial species acting as potential donors, as well as the potential to act as donors for other bacterial species.

We found a number of variations within the penicillin-binding proteins, although only P601L in PBP2X showed a reduction in sensitivity to penicillin, as had previously been found with *S. pyogenes* [40]. The MIC for the two isolates with the P601L variation was the highest at 0.03 mg l^{-1}, which was similar to that found for *S. pyogenes* carrying this variant (0.032 mg l^{-1}) [38], and for reference, it is the breakpoint as defined by EUCAST (https://www.eucast.org/). It is possible that, although the PBP2X variation in itself does not confer high-level resistance nor has reduced sensitivity to a level that might have clinical impact, it is possible that these mutations are the starting point for such resistance to develop. These two isolates were obtained from the same patient, 57 days apart, and had remained unchanged genetically. Whether the presence of the P601L mutation impacted the lack of infection clearance in this patient is unclear, as we do not have the clinical information available to know what treatment was given and for how long. We do not know the history of this patient and, therefore, can only hypothesize that the variation in PBP2X arose during infection, although, as no other isolates were identified carrying this variation, it seems likely that this was the case.

The lack of genetic change from 4/7 pairs of isolates from the same patient indicated long-term persistence within the host. The earliest isolate from patient 4 was from blood culture, and this was followed by a skin swab isolate 74 days later, yet these were identical, indicating long-term persistence within a patient, even with the potential to cause invasive disease. When genetic changes did occur, they did not seem to be related to the length of time or genomic cluster, indicating a host factor. It is not clear what the impact of any of the genetic changes observed would be on the isolate, but there was no evidence for changes that would alter antimicrobial sensitivity. Although it is possible that the variation in *mntR* that arose in patient 7 could alter metal transport, which in turn could alter resistance to oxidative stress [56]. As no other clinical information can be obtained for any of these isolates, it is not known if there were other infections prior to or post-to those we collected, nor do we know what kind of therapy the patients underwent and whether there was any evidence of treatment failure. However, long-term persistence and the potential for repeated treatment failure are a concern.

An additional public health concern was the finding that five isolates were *S. canis*, a pathogen thought to rarely infect humans. Originally identified in dogs, *S. canis* can infect a number of different animals, including cats, and, therefore, household pets may be a source of transmission to humans [36]. Prevalence in humans may be underestimated as *S. canis* is often diagnosed as a Lancefield group G *Streptococcus* rather than confirmed at the species level, as was the case in our study prior to genome sequencing. Given the fairly high proportion (4.5%) we found, coupled with the fact that two isolates carried tetracycline resistance genes and one of these also carried a macrolide resistance gene, it is important to note that *S. canis* can cause infections in humans and should be highlighted as a public health concern. It also demonstrates the importance of speciating streptococci with tools such as MALDI-TOF [57, 58] or whole-genome sequencing.

Overall, we have identified that, like other countries, genome clusters GC-1 and GC-2 are highly prevalent, accounting for 60.6% of all the isolates. The high level of diversity within the SDSE population makes it difficult to monitor with typing methods, like *emm* or MLST, alone and whole-genome sequencing is required to fully understand the population. It is still unclear as to why GC-1 and GC-2 are so common in the population and what bacterial factors might be driving this. Genome sequencing of SDSE on a national scale is needed to better understand the UK population, in particular to determine any invasive disease potential associated with GC-1 or GC-2. It would also be of interest to confirm the low level of throat infections caused by SDSE in comparison to skin infections and if this differs from *S. pyogenes,* for which we have found more throat infections compared to skin infections [32].

Funding information
This work was funded by seed project funding from the Florey Institute AMR Research Capital Funding awarded by the NIHR (NIHR Antimicrobial Resistance 2018 reference NIHR200636). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. C.E.T. is a Wellcome Trust Career Development Fellow (227240/Z/23/Z).

Author contributions
S.Y.B.: data curation, formal analysis, investigation, software, visualization and writing – original draft. H.K.: resources, methodology and investigation. S.J.: investigation. R.R.C.: software and investigation. L.R.G.: investigation. L.T.: resources and methodology. D.P.: resources. T.I.d.S.: resources, funding acquisition and writing – review and editing. C.E.T.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, supervision, visualization, writing – original draft and writing – review and editing.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Xie O, Davies MR, Tong SYC. *Streptococcus dysgalactiae* subsp. *equisimilis* infection and its intersection with *Streptococcus pyogenes*. *Clin Microbiol Rev* 2024;37:e0017523.
2. Lambertsen LM, Ingels H, Schønheyder HC, Hoffmann S, Danish Streptococcal Surveillance Collaboration Group 2011. Nationwide laboratory-based surveillance of invasive beta-haemolytic streptococci in Denmark from 2005 to 2011. *Clin Microbiol Infect* 2014;20:O216–23.
3. Wajima T, Morozumi M, Hanada S, Sunaoshi K, Chiba N, *et al*. Molecular characterization of invasive *Streptococcus dysgalactiae* subsp. *equisimilis*, Japan. *Emerg Infect Dis* 2016;22:247–254.
4. Kurachi A, Ishida Y, Nakazawa K, Okada T, Kishida T, *et al*. Necrotizing fasciitis and septic shock due to streptococcal toxic shock syndrome in an elderly patient: a case report. *Clin Case Rep* 2023;11:e6846.
5. Xie O, Zachreson C, Tonkin-Hill G, Price DJ, Lacey JA, *et al*. Overlapping *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subspecies *equisimilis* household transmission and mobile genetic element exchange. *Nat Commun* 2024;15:3477.
6. Itzek A, Weißbach V, Meintrup D, Rieß B, van der Linden M, *et al*. Epidemiological and clinical features of *Streptococcus dysgalactiae* ssp. *equisimilis* stG62647 and other emm types in Germany. *Pathogens* 2023;12:589.
7. McNeilly CL, McMillan DJ. Horizontal gene transfer and recombination in *Streptococcus dysgalactiae* subsp. *equisimilis*. *Front Microbiol* 2014;5:676.
8. Shimomura Y, Okumura K, Murayama SY, Yagi J, Ubukata K, *et al*. Complete genome sequencing and analysis of a Lancefield group G *Streptococcus dysgalactiae* subsp. *equisimilis* strain causing streptococcal toxic shock syndrome (STSS). *BMC Genomics* 2011;12:17.
9. Oppegaard O, Mylvaganam H, Skrede S, Kittang BR. Exploring the arthritogenicity of *Streptococcus dysgalactiae* subspecies *equisimilis*. *BMC Microbiol* 2018;18:17.
10. UKHSA. Laboratory surveillance of streptococcal bacteraemia in England: 2023 update; 2024 [accessed 12 December 2024].
11. Bläckberg A, Nilson B, Özenci V, Olaison L, Rasmussen M. Infective endocarditis due to *Streptococcus dysgalactiae*: clinical presentation and microbiological features. *Eur J Clin Microbiol Infect Dis* 2018;37:2261–2272.
12. Ruppen C, Rasmussen M, Casanova C, Sendi P. A 10-year observational study of *Streptococcus dysgalactiae* bacteraemia in adults: frequent occurrence among female intravenous drug users. *Swiss Med Wkly* 2017;147:w14469.
13. Xie O, Featherstone L, Nguyen ANT, Hayes AJ, Pitt ME, *et al*. Invasive *Streptococcus dysgalactiae* subspecies *equisimilis* compared with *Streptococcus pyogenes* in Australia, 2011-23, and the emergence of a multi-continent stG62647 lineage: a retrospective clinical and genomic epidemiology study. *Lancet Microbe* 2025;6:101182.
14. Kaci A, Jonassen CM, Skrede S, Sivertsen A, Steinbakk M, *et al*. Genomic epidemiology of *Streptococcus dysgalactiae* subsp. *equisimilis* strains causing invasive disease in Norway during 2018. *Front Microbiol* 2023;14:1171913.
15. Leitner E, Zollner-Schwetz I, Zarfel G, Masoud-Landgraf L, Gehrer M, *et al*. Prevalence of emm types and antimicrobial susceptibility of *Streptococcus dysgalactiae* subsp. *equisimilis* in Austria. *Int J Med Microbiol* 2015;305:918–924.
16. Lother SA, Demczuk W, Martin I, Mulvey M, Dufault B, *et al*. Clonal clusters and virulence factors of group C and G *Streptococcus* causing severe infections, Manitoba, Canada, 2012–2014. *Emerg Infect Dis* 2017;23:1079–1088.
17. Rojo-Bezares B, Toca L, Azcona-Gutiérrez JM, Ortega-Unanue N, Toledano P, *et al*. *Streptococcus dysgalactiae* subsp. *equisimilis* from invasive and non-invasive infections in Spain: combining epidemiology, molecular characterization, and genetic diversity. *Eur J Clin Microbiol Infect Dis* 2021;40:1013–1021.
18. Traverso F, Blanco A, Villalón P, Beratz N, Sáez Nieto JA, *et al*. Molecular characterization of invasive *Streptococcus dysgalactiae* subsp. *equisimilis*. Multicenter study: Argentina 2011-2012. *Rev Argent Microbiol* 2016;48:279–289.
19. Beres SB, Olsen RJ, Long SW, Eraso JM, Boukthir S, *et al*. Analysis of the genomics and mouse virulence of an emergent clone of *Streptococcus dysgalactiae* subspecies *equisimilis*. *Microbiol Spectr* 2023;11:e0455022.
20. Oppegaard O, Mylvaganam H, Skrede S, Lindemann PC, Kittang BR. Emergence of a *Streptococcus dysgalactiae* subspecies equisimilis stG62647-lineage associated with severe clinical manifestations. *Sci Rep* 2017;7:7589.
21. Pospiech A, Neumann B. A versatile quick-prep of genomic DNA from gram-positive bacteria. *Trends Genet* 1995;11:217–218.
22. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 2017;13:e1005595.
23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.
24. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 2018;34:i142–i150.
25. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
26. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, *et al*. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 2020;21:180.
27. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, *et al*. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res* 2019;47:D309–D314.
28. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–1313.
29. Letunic I, Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res* 2024;52:W78–W82.
30. Lees JA, Harris SR, Tonkin-Hill G, Gladstone RA, Lo SW, *et al*. Fast and flexible bacterial genomic epidemiology with PopPUNK. *Genome Res* 2019;29:304–316.
31. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, *et al*. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. *Nucleic Acids Res* 2015;43:e15.
32. Hall JN, Bah SY, Khalid H, Brailey A, Coleman S, *et al*. Molecular characterization of *Streptococcus pyogenes* (StrepA) non-invasive isolates during the 2022-2023 UK upsurge. *Microb Genom* 2024;10:001277.
33. Sachse S, Seidel P, Gerlach D, Günther E, Rödel J, *et al*. Superantigen-like gene(s) in human pathogenic *Streptococcus dysgalactiae*, subsp *equisimilis*: genomic localisation of the gene encoding streptococcal pyrogenic exotoxin G (speG(dys)). *FEMS Immunol Med Microbiol* 2002;34:159–167.
34. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, *et al*. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype–phenotype correlations in a collection of isolates. *Antimicrob Agents Chemother* 2019;63:e00483-19.
35. Davies MR, McIntyre L, Mutreja A, Lacey JA, Lees JA, *et al*. Atlas of group A streptococcal vaccine candidates compiled using large-scale comparative genomics. *Nat Genet* 2019;51:1035–1043.
36. Pagnossin D, Smith A, Oravcová K, Weir W. *Streptococcus canis*, the underdog of the genus. *Vet Microbiol* 2022;273:109524.
37. Hayes A, Lacey JA, Morris JM, Davies MR, Tong SYC. Restricted sequence variation in *Streptococcus pyogenes* penicillin-binding proteins. *mSphere* 2020;5.
38. Yu D, Guo D, Zheng Y, Yang Y. A review of penicillin binding protein and group A *Streptococcus* with reduced-β-lactam susceptibility. *Front Cell Infect Microbiol* 2023;13:1117160.

39. Olsen RJ, Zhu L, Musser JM. A single amino acid replacement in penicillin-binding protein 2X in *Streptococcus pyogenes* significantly increases fitness on subtherapeutic benzylpenicillin treatment in a mouse model of necrotizing myositis. *Am J Pathol* 2020;190:1625–1631.
40. Musser JM, Beres SB, Zhu L, Olsen RJ, Vuopio J, *et al*. Reduced *in vitro* susceptibility of *Streptococcus pyogenes* to *β*-lactam antibiotics associated with mutations in the *pbp2x* gene is geographically widespread. *J Clin Microbiol* 2020;58:e01993-19.
41. Kratovac Z, Manoharan A, Luo F, Lizano S, Bessen DE. Population genetics and linkage analysis of loci within the FCT region of *Streptococcus pyogenes*. *J Bacteriol* 2007;189:1299–1310.
42. Graham MR, Smoot LM, Migliaccio CAL, Virtaneva K, Sturdevant DE, *et al*. Virulence control in group A *Streptococcus* by a two-component gene regulatory system: global expression profiling and *in vivo* infection modeling. *Proc Natl Acad Sci USA* 2002;99:13855–13860.
43. Hasegawa T, Matsumoto M, Hata N, Yano H, Isaka M, *et al*. Homologous role of CovRS two-component regulatory system in NAD^+-glycohydrolase activity in *Streptococcus dysgalactiae* subsp. *equisimilis* as in *Streptococcus pyogenes*. *APMIS* 2019;127:87–92.
44. Steiner K, Malke H. Dual control of streptokinase and streptolysin S production by the covRS and fasCAX two-component regulators in *Streptococcus dysgalactiae* subsp. *equisimilis*. *Infect Immun* 2002;70:3627–3636.
45. Watanabe S, Shimomura Y, Ubukata K, Kirikae T, Miyoshi-Akiyama T. Concomitant regulation of host tissue-destroying virulence factors and carbohydrate metabolism during invasive diseases induced by group g streptococci. *J Infect Dis* 2013;208:1482–1493.
46. Rößler S, Berner R, Jacobs E, Toepfner N. Prevalence and molecular diversity of invasive *Streptococcus dysgalactiae* and *Streptococcus pyogenes* in a German tertiary care medical centre. *Eur J Clin Microbiol Infect Dis* 2018;37:1325–1332.
47. Walkinshaw DR, Wright MEE, Mullin AE, Excler JL, Kim JH, *et al*. The *Streptococcus pyogenes* vaccine landscape. *NPJ Vaccines* 2023;8:16.
48. Christiansen CH, Søgaard KK, Dam-Dalgeir G, Dessau RB, Dzajic E, *et al*. Surveillance of invasive beta-haemolytic streptococci in Denmark, 2012 to 2023: a nationwide study. *J Infect* 2025;91:106559.
49. Oppegaard O, Glambek M, Skutlaberg DH, Skrede S, Sivertsen A, *et al*. *Streptococcus dysgalactiae* bloodstream infections, Norway, 1999-2021. *Emerg Infect Dis* 2023;29:260–267.
50. Xie O, Morris JM, Hayes AJ, Towers RJ, Jespersen MG, *et al*. Inter-species gene flow drives ongoing evolution of *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subsp. *equisimilis*. *Nat Commun* 2024;15:2286.
51. López de Egea G, González-Díaz A, Olsen RJ, Guédon G, Berbel D, *et al*. Emergence of invasive *Streptococcus dysgalactiae* subsp. *equisimilis* in Spain (2012-2022): genomic insights and clinical correlations. *Int J Infect Dis* 2025;153:107778.
52. Eraso JM, Olsen RJ, Long SW, Gadd R, Boukthir S, *et al*. Integrative genomic, virulence, and transcriptomic analysis of emergent *Streptococcus dysgalactiae* subspecies *equisimilis* (SDSE) emm type stG62647 isolates causing human infections. *mBio* 2024;15.
53. Bah SY, Keeley AJ, Armitage EP, Khalid H, Chaudhuri RR, *et al*. Genomic characterization of skin and soft tissue *Streptococcus pyogenes* isolates from a low-income and a high-income setting. *mSphere* 2023;8:e0046922.
54. López de Egea G, González-Díaz A, Guédon G, Lao J, Berbel D, *et al*. A new integrative and mobilizable element is a major contributor to tetracycline resistance in *Streptococcus dysgalactiae* subsp. *equisimilis*. *Antibiotics* 2023;12:579.
55. Sati H, Carrara E, Savoldi A, Hansen P, Garlasco J, *et al*. The WHO bacterial priority Pathogens List 2024: a prioritisation study to guide research, development, and public health strategies against antimicrobial resistance. *Lancet Infect Dis* 2025;25:1033–1043.
56. Henningham A, Döhrmann S, Nizet V, Cole JN. Mechanisms of group A *Streptococcus* resistance to reactive oxygen species. *FEMS Microbiol Rev* 2015;39:488–508.
57. Jensen CS, Dam-Nielsen C, Arpi M. Matrix-assisted laser desorption/ionization-time of flight mass spectrometry identification of large colony beta-hemolytic streptococci containing Lancefield groups A, C, and G. *Infecti Dis* 2015;47:575–579.
58. Nybakken EJ, Oppegaard O, Gilhuus M, Jensen CS, Mylvaganam H. Identification of *Streptococcus dysgalactiae* using matrix-assisted laser desorption/ionization-time of flight mass spectrometry; refining the database for improved identification. *Diagn Microbiol Infect Dis* 2021;99:115207.
59. Porcellato D, Smistad M, Skeie SB, Jørgensen HJ, Austbø L, *et al*. Whole genome sequencing reveals possible host species adaptation of *Streptococcus dysgalactiae*. *Sci Rep* 2021;11:17350.
60. Rebelo AR, Ibfelt T, Bortolaia V, Leekitcharoenphon P, Hansen DS, *et al*. One day in Denmark: nationwide point-prevalence survey of human bacterial isolates and comparison of classical and whole-genome sequence-based species identification methods. *PLoS One* 2022;17:e0261999.
61. Chochua S, Rivers J, Mathis S, Li Z, Velusamy S, *et al*. Emergent invasive group A *Streptococcus dysgalactiae* subsp. *equisimilis*, United States, 2015–2018. *Emerg Infect Dis* 2019;25:1543–1547.
62. Löwbeer N, Stegger M, Söderquist B. Genomic characterization of beta-haemolytic streptococci isolated from prosthetic joint infections. *APMIS* 2023;131:189–196.