Running Title: comparative genomics analysis of SDSE

Comprehensive comparative genomics analysis for the emerging human pathogen *Streptococcus dysgalactiae* subsp. *equisimilis* (SDSE): A case study and Pan-subspecies genomic analysis

Amr T. M. Saeb* ^{1, 2} and Hamsa T. Tayeb* ^{3,4}

1. Strategic Center for Diabetes Research, College of Medicine, King Saud University, KSA.
2. Zifo RnD Solutions, New Market, United Kingdom (Current address).
3. Genetics Department, King Faisal Specialist Hospital and Research Center, Riyadh, KSA.
4. Saudi Human Genome Project, King Abdulaziz City for Science and Technology (KACST), Riyadh, KSA

***Corresponding Authors:**

Amr T. M. Saeb, Ph.D.
College of Medicine, King Saud University, KSA.
Zifo RnD Solutions, New Market, United Kingdom (Current address).
Tel: +44-7405832804
Email: saeb.1@osu.edu

Hamsa T. Tayeb, Ph.D.
Genetics Department, King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia.
Tel: +966-505502122
hamsa3000@hotmail.com

Abstract: Background: *Streptococcus dysgalactiae* subsp. *equisimilis* (SDSE) is the causal agent of various diseases that include wound infection, erysipelas, cellulitis, life-threatening necrotizing fasciitis, and streptococcal toxic shock syndrome. It is capable of infecting both humans and animals. In this investigation, we present a comprehensive genomic analysis for the SDSE strain SCDR1 that belongs to Lancefield group G, emm type (stG6) and (MLST) sequence type (ST44) that is the first time to be documented in Saudi Arabia and the middle east. Besides, we present the most comprehensive comparative genomics analysis for the emerging human pathogen SDSE. **Methodology:** We utilized next-generation sequencing techniques (NGS), bioinformatics, phylogenetic analysis, and comparative pathogenomics to characterize SCDR1 and all publicly available SDES genomes. **Results:** We found that SCDR1 consisted of a circular genome of 2179136 bp. Comparative analyses among bacterial genomes indicated that SCDR1 was most closely related to AC-2713 and GGS_124. Genome annotation of SCDR-1 strain showed that it contains many genes with homology to known virulence factors, including genes involved in cellular invasion, Antiphagocytosis, immune evasion, invasion of skin and soft tissue, host mortality and tissue damage, toxins, pore-forming proteins, cytotoxins, beta-hemolysis agents. Two CRISPR arrays were identified in SCDR1 that are consist of 35 CRISPR repeats and 33 CRISPR spacers. Two CAS systems were observed in the SCDR-1 genome, namely, CAS-TypeIIA and CAS-TypeIC. SDSE core Resistome is consisting of 22 genes, including *folA*, *gyrA*, *gyrB*, and *FabK*. SDSE core Virulome consisting of 38 genes including, *fba*, *fbp54*, *gidA*, and *lsp*. **Conclusion:** Our study confirmed that the SDSE strains possess different characteristics in producing virulence factors for pathogenicity to humans and based on its genome sequence and close relationship with GAS. Our study shed light on the proposed pathogenic mechanisms of SDSE and may form the basis of molecular epidemiological research on these highly virulent bacteria.

Keywords: *Streptococcus dysgalactiae* subsp. *equisimilis*, Lancefield group G, complete genome sequence, phylogeny, comparative genomics, virulence, antimicrobial resistance, drug target, prophage, CRISPR-Cas.

Introduction:

Streptococcus dysgalactiae subsp. *equisimilis* (SDSE) is an emerging human pathogen that causes severe invasive streptococcal diseases (Matsue et al., 2020). It belongs to β -hemolytic streptococci and possesses Lancefield group C or G antigens and rarely, A antigen (Takahashi et al., 2011). SDSE was suggested as a new taxon among human streptococcal isolates in 1996. It is a common colonizer of the pharynx, skin, gastrointestinal tract, and female genital tract (Yung et al., 2019). However, SDSE is the causal agent of various diseases that include wound infection, erysipelas, cellulitis, streptococcal toxic shock syndrome (STSS), and life-threatening necrotizing fasciitis (Rantala, 2014). Besides, SDSE is associated with pharyngitis, acute post-streptococcal glomerulonephritis, and acute rheumatic fever (Reid et al., 1985; Haidan et al., 2000). These clinical characteristics of diseases produced by SDSE infection are closely similar to *Streptococcus pyogenes* (GAS) infections (Brandt and Spellerberg, 2009). Several studies suggested a unique link between group G streptococcal bacteremia and invasive infection with underlying conditions, such as alcoholism, diabetes mellitus, or cardiovascular (Armstrong et al., 1970). In addition, numerous studies reported the invasive pathogenicity of SDSE, especially with older adults with underlying diseases such as diabetes mellitus. Therefore, it was found that the mean age of adult patients with invasive SDSE infection is older than those infected with GAS (Cohen-Poradosu et al., 2004; Wajima et al., 2016). Furthermore, animal studies showed that an SDSE strain causes severe pathogenicity to diabetic mice compared with GAS strains (Ogura et al., 2018). SDSE causes skin and soft-tissue infections, including pyoderma, cellulitis, wound infections, abscesses, erysipelas, and necrotizing fasciitis (Broyles et al., 2009). Severe infections often affect injection drug users, elderly patients with underlying immunosuppressive comorbidities or conditions predisposing them to skin breakdown, and burn patients in whom the infections may lead to graft failure (Rider and McGregor, 1994). For invasive SDSE infections, sites of colonization or primary focal infections are most likely ports of entry (Auckenthaler et al., 1983). Invasive infections comprise arthritis, osteomyelitis, pleuropulmonary infections, peritonitis, intra-abdominal and epidural abscesses, meningitis, endocarditis, puerperal septicemia, neonatal infections, necrotizing fasciitis, myositis,

and streptococcal toxic-like syndrome (Barnham and Weightman, 2004). SDSE has Lancefield group C or G antigens, exhibits strong β -hemolysis, and exerts streptokinase activity upon human plasminogen and proteolytic activity upon human fibrin. Similarly, to group A streptococci, SDSE possesses virulence factors including M protein, streptolysin O, streptolysin S, streptokinase, hyaluronidase, C5a peptidase, and others (Takahashi et al., 2011). In contrast, SDSE lacks several key virulence factors present in GAS, such as superantigens other than *speG*, the cysteine protease *speB*, the hyaluronic synthesis operon *hasA* and *hasB* and an inhibitor of complement activation *sic* (Wajima et al., 2016; Watanabe et al., 2016). It was found that 59% of clinical isolates of SDSE clinical isolates were able to form biofilm in vitro; however, all of the tested isolates were able to form biofilm in vivo (Genteluci et al., 2015). It was found that in vivo passage in mice enhanced biofilm-forming ability in some SDSE isolates. Epidemiological surveillance has been crucial to detect changes in the geographical and temporal variation of GAS associated disease pattern; for this purpose, the M protein gene (*emm*) gene typing is the most widely used genotyping method than 200 *emm* types recognized (Gherardi et al., 2018). The *emm1* GAS, the strain with enhanced virulence, is the dominant type in Europe, North America, and Japan (Wajima et al., 2008; Gherardi et al., 2018). The enhanced virulence and invasion fitness of *emm1* GAS strain is justified by phage mobilization-based diversification and the capability to recognize and adapt to the diverse host environments (Aziz and Kotb, 2008). Consequently, it was reported that *emm1* type is associated with the distribution of phage-associated superantigen genes (*speA*, *speC* and *ssa*) and the number of necrotizing fasciitis cases (Plainvert et al., 2012). On the contrary, prevalent *emm* types of SDSE differ amongst countries. For instance, in Japan, the most prevalent is stG6792 followed by stG485 and stG245 (Wajima et al., 2016; Ikebe et al., 2018). Whereas in Taiwan, the stG10, stG245, stG840, stG6.1, and stG652 strains predominated (Ma et al., 2017). In Jerusalem and Austria stG485 and stG6.1 were predominant, respectively (Cohen-Poradosu et al., 2004; Leitner et al., 2015). In addition, the Multi-Locus Sequence Typing (MLST) showed that 178 SDSE isolates, representing 37 *emm* types, segregate into 80 distinct sequence types (STs) and that the same ST can harbor different *emm* types (McMillan et al., 2010). In this investigation we present a comprehensive genomic analysis for the SDSE strain SCDR1 that belong to Lancefield group G, *emm* type (stG6) and

Multilocus sequence typing (MLST) sequence type (ST44) that is first time to be documented in Saudi Arabia and middle east. In addition, we present the most comprehensive comparative genomics analysis for the emerging human pathogen SDSE.

Methodology:

Bacterial isolate: *Streptococcus dysgalactiae* subsp. *equisimilis* strain SCDR-SD1 was isolated from a diabetic ulcer patient in the diabetic foot unit at the University Diabetes Center, King Saud University. SDSE-SCDR1 was collected from a 60 years old's neuroischaemic diabetic foot ulcer beneath the left fifth metatarsal head with a wound surface area of one cm² and the depth of the wound was 0.5 cm. The wound was debrided before swabbing, and an appropriate wound swab was collected from the patient and was sent for additional microbiological culture and investigations. SDSE was cultured in 5% sheep blood agar or brain–heart infusion medium at 37 °C under 5% CO₂ as described Miyoshi-Akiyama et al. (Miyoshi-Akiyama et al., 2003). The species identification was conducted by partially sequencing the 16S rRNA gene described previously (Wang et al., 2016). In vitro susceptibility of isolate SDSE-SCDR1 to antibiotics was determined by disk diffusion following the recommendations of CLSI and as described before (Matuschek et al., 2014; Wang et al., 2016).

DNA isolation, purification, sequencing, bioinformatics, and phylogenetic analysis: DNA isolation, purification, genome sequencing, bioinformatics, and phylogenetic analysis were performed as described before (Saeb et al., 2017b; Saeb, 2018). The closest reference and representative genomes were identified using Mash/MinHash (Ondov et al., 2016). PATRIC global protein families (PGFams) (Davis et al., 2016) were selected from these genomes to determine this genome's phylogenetic placement. The protein sequences from these families were aligned with MUSCLE (Edgar, 2004), and the nucleotides for each of those sequences were mapped to the protein alignment. The common set of amino acid and nucleotide alignments were concatenated into a data matrix, and RaxML (Stamatakis, 2014) was used to analyze this matrix, with fast bootstrapping (Stamatakis et al., 2008) was used to generate the support values in the tree.

Furthermore, we performed whole-genome phylogeny based proteomic comparison among *Streptococcus dysgalactiae* subsp. *equisimilis* SCDR-SD1 isolate and other related *Proteus mirabilis* strains using Proteome Comparison service, a protein sequence-based available at (<https://www.patricbrc.org/app/SeqComparison>) as described before (Saeb et al., 2017b, 2017a).

Gene annotation and Pathogenomics analysis: *Streptococcus dysgalactiae* subsp. *equisimilis* SCDR-SD1 genome contigs were annotated using the Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP) available at NCBI (<http://www.ncbi.nlm.nih.gov/>). Besides, contigs were further annotated using the bacterial bioinformatics database and analysis resource (PATRIC) gene annotation service (<https://www.patricbrc.org/app/Annotation>). **PubMLST**, the public databases for molecular typing and microbial genome diversity, was used to confirm the MLST analysis available at (<https://pubmlst.org/>). The PathogenFinder, pathogenicity prediction program available at (<https://cge.cbs.dtu.dk/services/PathogenFinder/>) was used to examine the nature of *Streptococcus dysgalactiae* subsp. *equisimilis* SCDR-SD1 as a human pathogen. Virulence gene sequences and functions, matching to different main bacterial virulence factors of SDSE were claimed from GenBank and confirmed using virulence factors of the pathogenic bacteria database available at (<http://www.mgc.ac.cn/VFs/>), the Victors virulence factors search service (<http://www.phidias.us/victors/>), besides, the PATRIC_VF tool accessible at <https://www.patricbrc.org/> as described before (Saeb et al., 2017b). **RASTA-Bacteria**, the Rapid Automated Scan for Toxins and Antitoxins in Bacteria (Sevin and Barloy-Hubler, 2007), and **TASer**, a web tool that locates Toxins and Antitoxins in a genome using a high-performance Computing Cluster (Akarsu et al., 2019), were used for the Toxins and Antitoxins confirmatory analysis. **DrugBank**, available at (<https://go.drugbank.com/>), was used in the drug target analysis.

Resistome analysis: *Streptococcus dysgalactiae* subsp. *equisimilis* SCDR-SD1 genome contigs were investigated manually for the presence of antibiotic resistance loci using the PGAAP and PATRIC gene annotation services. Antibiotic resistance loci were further investigated using specialized search tools and services, namely: Gene Search, Antibiotic Resistance, and Genome Feature Finder

(<https://www.patricbrc.org>), CARD (The Comprehensive Antibiotic Resistance Database) accessible at (<https://card.mcmaster.ca/>), ARDB (Antibiotic Resistance Genes Database) accessible at (<https://ardb.cbcb.umd.edu/>), and ResFinder 2.1 available at (<https://cge.cbs.dtu.dk/services/ResFinder/>) as described before (Saeb et al., 2017b; Saeb, 2018).

Phage, Insertion sequences, and CRISPR-CAS analysis: **PHASTER** (PHAge Search Tool Enhanced Release) available at (<https://phaster.ca/>), **PHAST** (PHAge Search Tool) available at (<http://phast.wishartlab.com/index.html>), **MARVEL**, a Tool for Prediction of Bacteriophage Sequences in Metagenomic Bins (Amgarten et al., 2018) along with PARTIC's Genome Feature functions were used for prophage search and annotation. Additionally, **ISMMapper** was used in identifying transposase insertion sites in the SDSE-SCDR1 genome (Hawkey et al., 2015). **CRISPRCasFinder**, the program enables the easy detection of CRISPRs and *cas* genes, available at (<https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index>) along with PARTIC's Genome Feature functions were used for CRISPRs-cas systems search and annotation.

Comparative genomics and bioinformatics analysis: All publicly available representative SDSE genome sequences were collected from GenBank and validated using PARTIC deposited genomes. **Supplementary Table 1** represents the taxonomic/genomic information of the 55 publicly available SDSE strains/isolates used in this study. Comparative investigations were performed for all genomes previously described (Saeb et al., 2017a, 2017b): this includes phylogenomic, virulome, resistome, drug targets, prophage, CRISPR-CAS analysis. Additional web recourses used in the analysis are the **Galaxy** (an open-source, web-based platform for data-intensive biomedical research) available at (<https://usegalaxy.org/>). **DDBJ Center** provides sharing and analysis services for data from life science researches and advances science, available at (<https://www.ddbj.nig.ac.jp/index-e.html>).

Results

Case description: Our isolate was collected from 60 years old Saudi male visiting the foot unit at the university diabetes center at King Saud University. He was diagnosed with Type two diabetes mellites for more than 20 years with Neuropathy, Retinopathy, Obesity complications. His HbA1C was 7.7, and his BMI is 28.58. He came with a neuroischaemic diabetic foot ulcer beneath the left fifth metatarsal head with a wound surface area of one cm², and the depth of the wound was 0.5 cm. The case of the DFU was under a callus, and the ankle-brachial pressure index (ABPI) was 0.61. The patient was on Insulin mixtard 30U AM/ 30U BT.

Taxonomic and Genome description: Our *Streptococcus dysgalactiae* subsp. *equisimilis* strain SDSE-SCDR-1 complete genome was deposited in the NCBI-GenBank with GenBank accession number CP033391 under BioProject accession number PRJNA498183, BioSample accession number SAMN10285107 and assembly accession number GCA_009497975.1. The NCBI Taxon ID is 1334 Bacteria; Terrabacteria group; Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Streptococcus; Streptococcus dysgalactiae group; Streptococcus dysgalactiae subsp. equisimilis. The Multilocus sequence type (MLST) is Streptococcus_dysgalactiae_equisimilis.44. In silico analysis showed that SDSE strain SD-SCDR1 belongs to the emm stG6.0 group. The genome size is 2179136, the GC Content is 39.87, the number of RefSeq CDS is 1843, and PATRIC CDS is 2065. The number of contigs is one, the Contig L50 is one, and Contig N50 is 2,179,136. The number of tRNAs is 57, rRNA is 15, misc-RNA is 14, regulatory sequences is 6, repeat regions is 90, CRISPR repeats is 35, and CRISPR spacers is 33. This genome was annotated using genetic code 11. The annotation included 375 hypothetical proteins and 1,671 proteins with functional assignments. The proteins with functional assignments included 561 proteins with Enzyme Commission (EC) numbers (Schomburg et al., 2004), 467 with Gene Ontology (GO) assignments (Ashburner et al., 2000), and 383 proteins that were mapped to KEGG pathways (Kanehisa et al., 2016). PATRIC annotation includes two types of protein families (Davis et al., 2016), and this genome has 1,910 proteins that belong to the genus-specific protein families (PLFams) and 1,914 proteins that belong to the

cross-genus protein families (PGFams). **Figure 1** designates the graphical demonstration of the genome circulation and annotations. The Genome-genome comparison analysis showed that SDSE strain SD-SCDR1 is closely related to SDSE strain SK1250, strain WCHSDSE-1, and strain UT-5345, with distances values of 0.00399629, 0.00467416, and 0.00687462, respectively (**Table 1**). Whole Genome Alignment of strain SD-SCDR1 against reference genomes of strain NCTC7136, strain WCHSDSE-1 and *Streptococcus pyogenes* strain MGAS15252 is presented in **Figure 2**.

Figure 1: Genome representation shows from outer to inner rings, the contigs, CDS on the forward strand, CDS on the reverse strand, RNA genes, CDS with homology to known antimicrobial resistance genes, CDS with homology to known virulence factors, GC content, and GC skew.

Figure 2: Mauve Multiple Whole Genome Alignment of SDSE strain SD-SCDR1 against reference genomes SDSE strain NCTC7136, SDSE strain WCHSDSE-1 and *Streptococcus pyogenes* strain MGAS15252.

Biological Subsystems Analysis: A subsystem is a set of proteins that together implement a specific biological process or structural complex (Overbeek et al., 2005), and PATRIC annotation includes an analysis of the subsystems unique to each genome. An overview of the subsystems for this genome is provided in **Figure 3**. The highest number of genes (342) in our genome is dedicated to metabolism, followed by protein processing (204), then energy production (117). It is noteworthy to mention that 85 genes are dedicated to stress response, defense, and virulence.

Figure 3: An overview of the biological subsystems and genes for *Streptococcus dysgalactiae* subsp. *equisimilis* strain SD-SCDR1 genome

Phylogenetic analysis:

Figure 4: (a) Phylogenetic tree of SDSE based on whole-genome sequences (reference and outgroup genomes) and (b) selected 33 SDSE genome sequences, *Streptococcus dysgalactiae* subsp. *equisimilis* strain SD-SCDR1, *Streptococcus pyogenes* MGAS15252, and *Staphylococcus aureus* MR1 genomes

Whole-genome phylogenetic analysis showed that our strain SD-SCDR1 belongs to a clade that contains *Streptococcus dysgalactiae* subsp. *equisimilis* strains, namely the reference strain AC-2713 and GGS 124 (Figure 4a). Besides, within *Streptococcus dysgalactiae* subsp. *equisimilis* taxa, our strain SD-SCDR1 is closely related to both WCHSDSE-1 and strain SK1250 (Figure 4b).

Specialty Genes investigation:

Genome annotation of Our SD-SCDR-1 strain showed that it contains many genes with homology to known virulence factors, antibiotic resistance genes, and drug targets.

Virulence factors genes: Table 2 represents the observed virulence factors genetic determinants in the SDSE-SCDR-1 genome. Our strains contain genetic determinates involved in adherence to the host cell, cellular invasion, Antiphagocytosis, immune evasion, and skin and soft tissue invasion. In addition to genes responsible for host mortality and tissue damage, toxins, pore-forming proteins, cytotoxins, beta-hemolysis

agents, stress resistance thermotolerance, oxidative stress resistance, growth in blood and lung, and many other metabolic, genetic determinants that are crucial for the pathogenesis and virulence.

Antimicrobial Resistance Genes: A summary of the antimicrobial resistance genes (AMR) genes annotated in our genome and the corresponding AMR mechanism is provided in **Table 3**. It contains resistant genes the confer resistance against 4-quinolones (Fluoroquinolones), Aminocoumarin, Aminoglycoside (Streptomycin), Cycloserine, Daptomycin, Elfamycin (Pulvomycin), Fosfomycin, Fusidic acid, Isoniazid, Mupirocin (Pseudomonic acid), Rifampicin, Vancomycin, Daptomycin, Streptomycin Sulfonamide antibiotics, Tigecycline, Triclosan, and Trimethoprim. It is noteworthy to mention that our strain contains a gene that encodes for the multidrug resistance protein ErmB that confer resistance against Lincosamides, Macrolides, Streptogramins, and Tetracyclines.

Drug target genes: **Table 4** represents the potential novel drug target determinants in the SDSE-SCDR-1 genome. SDSE-SCDR-1 genome contains 12 different potential novel drug targets (genes) namely, *metG*, *glpO*, *fusA*, *rplC*, *parE* (*gyrB*), *guaB*, *rmlC*, *murI*, *proC*, *relA*, *def*, and *rmlB*. Against these targets, six experimental drugs were suggested, such as Inosinic Acid, Merimepodib, and Guanosine-5'-Diphosphate. Two were in the investigational stage, namely, Gatifloxacin and REP8839. Moreover, five approved drugs target four-drug targets: Procaine benzylpenicillin, Retapamulin, Spiramycin, Sulfaphenazole, and Curcumin.

CRISPR and Cas genes/systems information: **Tables 5, 6, and 7** describe CRISPR and Cas genes/systems identified in the SDSE-SCDR-1 genome. Two CRISPR arrays were identified that are consist of 35 CRISPR repeats and 33 CRISPR spacers. Although four potential CRISPR repeats were identified, namely SCDR1-CRISPR1-4, only two gained the highest confidence (evidence) level of 4, which are SCDR1-CRISPR 2 and 3. The length of SCDR1-CRISPR2 is 1286 bases, the repeat length is 36, the spacer length is 19 bases, and the conservation of repeats is 100%. Whereas the length of SCDR1-CRISPR 3 is 962 bases, the repeat length is 32, the spacer length is 14 bases, and the conservation of repeats

is 90.625 %. Two CAS systems were observed in the SDSE-SCDR-1 genome, namely, CAS-TypeIIA and CAS-TypeIC. CAS-TypeIIA contains Csn2-0-IIA as a mandatory gene and Cas2-0-I-II-III, Cas1-0-II, and Cas9-1-II (endonuclease Cas9) as accessory genes. While, CAS-TypeIC contains Cas1-0-IC, Cas7-0-IC, Cas8c-0-IC (Csd2/Csh2 family), Cas5-0-IC as mandatory genes, and Cas2-0-I-II-III-V, Cas4-0-I-II (RecB family exonuclease Cas4) and Cas3-0-I as accessory genes.

Prophage and transposons insertion sequences: Figures 5 and 8 present the 36.4Kb incomplete prophage containing 19 CDS that encodes for phage proteins. The GC% of the prophage is 36.25 and starts with the left attachment site (*attL*) at CDS position 2119960 and ends with the right attachment site (*attR*) at CDS position 2156433. Our prophage contains protein-coding sequences like CDSs from Lactococcus phage, Javan S145 prophage, Listeria phage, Bacillus virus G, Bacillus phage TsarBomba, and Streptococcus phage 315.1. Observed CDSs encodes for phage antirepressor protein, repressor protein, helix-turn-helix transcriptional regulators, site-specific integrase, transposase, DNA helicase, and phosphodiesterase.

Figure 5: incomplete prophage region identified in *Streptococcus dysgalactiae* subsp. *equisimilis* strain SD-SCDR1 genome

Furthermore, 40 Insertion sequences were observed in the SDSE-SCDR1 genome (Table 9). Amongst these, 20 belong to the insertion sequence (IS) family IS3, eight belongs to IS family IS30, five belongs to

the insertion sequence IS family IS1182, four belongs to IS family IS982, two belong to IS family ISAs1, and one belongs to IS family IS481. 36 IS have a *Streptococcus* species origin, two have a *Lactococcus* species origin, and two have an *Actinobacillus* species origin.

***Streptococcus dysgalactiae* subsp. *equisimilis* pan-taxonomic Comparative genomics section:**

Supplementary table 1 represents the taxonomic/genomic information of the 55 publicly available SDSE strains/isolates. These include genome name, NCBI taxon ID, genome status, strain MLST, other available typing serovar information, BioProject accession number, BioSample accession number, genome assembly accession number, GenBank accessions number, number of contigs per genome, genome length, GC% content, isolation Source, and host-name. These included both SDSE GGS_124 and AC-2713 strains as reference/representative genomes. We found that both SDSE strain NCTC11565 and strain NCTC5969 genomes are of low quality and should be revisited or removed from the public databases by the submitting authors. For the NCTC11565 genome, the genome quality was inferior; the length was too long, the CDS count was very high (6041 RefSeq CDS), and with a low **PLfam** CDS ratio. We believe that this genome is mixed with a *Pseudomonas aeruginosa* DNA similar to the PAO1 strain. The NCTC5969 genome's length was too long; the CDS count was high (2850 RefSeq CDS). We believe that this genome is mixed with *Staphylococcus aureus* DNA. Thus, both strains were excluded from our comparative genomics analysis. Nonetheless, the majority of isolates were of human origin; two isolates were collected from horses (genital tract), one isolate was collected from a monkey (throat), and one is from a dog (pleural exudate). Sites of isolation included blood cultures, sterile sites of invasive Diseases, oral, throat (Pharyngitis) or upper respiratory samples, and foot ulcers and Joint fluid samples.

SDSE strains protein families based phylogenetic analysis:

Figure 6 represents protein families based Maximum likelihood phylogenetic analysis for 60 genomes comprising the publicly available SDSE strains/isolates and related outgroup taxa showed that our isolate (SD-SCDR1) is located in a clade that contains 10 SDSE strains, namely, SK1250, WCHSDSE-1, KNZ01,

(Figure 7b). Generally, a high degree of conservation ($\geq 90\%$) was observed in the four tested strains in percent of gene/protein identity and sequence coverage compared to the reference strain SDSE-AC2713. For SDSE-SD-SCDR1 strain only three gene/proteins showed less conservation namely, Pyrroline-5-carboxylate reductase (EC 1.5.1.2) (*proC*) with a sequence coverage of 87%, Isochorismatase (EC 3.3.2.1) with a sequence coverage of 69% and Aspartyl-tRNA (Asn) amidotransferase subunit C (EC 6.3.5.6) @ Glutamyl-tRNA(Gln) amidotransferase subunit C (EC 6.3.5.7) (*gatC*) with a sequence coverage of 87%. Whereas, SDSE SK1250 showed reduced conservation in only two gene/proteins, namely, Undecaprenyl-diphosphatase (EC 3.6.1.27) (*uppP*) with a sequence coverage of 87% and Heptaprenyl diphosphate synthase component I (EC 2.5.1.30) with a sequence coverage of 89%. While in strain KNZ01, four gene/proteins showed less conservation, namely, Threonine synthase (EC 4.2.3.1) (*thrC*) with a percent of gene/protein identity of 70%, Pyrroline-5-carboxylate reductase (EC 1.5.1.2) (*proC*) with a sequence coverage of 87%, Isochorismatase (EC 3.3.2.1) with a sequence coverage of 86% and Heptaprenyl diphosphate synthase component I (EC 2.5.1.30) with a sequence coverage of 89%. For SDSE WCHSDSE-1, only three gene/proteins showed less conservation, namely, Pyrroline-5-carboxylate reductase (EC 1.5.1.2) (*proC*) with a sequence coverage of 87%, a hypothetical protein (Isochorismatase (EC 3.3.2.1) with a sequence coverage of 86%, and Heptaprenyl diphosphate synthase component I (EC 2.5.1.30) with a sequence coverage of 89%. However, the major four regions showed less conservation ($\leq 60\%$ identity and/or coverage) in the non-essential regions of the genome/proteome (**Figure 7b**), namely, 1-4 regions. Region one (starts 265508-ends 287829) contains mobile elements, hypothetical proteins (*int* and *manR*), Integrase gene (*int*), Lantibiotic salivaricin A gene and its associated genes/proteins (*salA*, *salB*, *salX*, *salY*, *salK*, *salR*), Antitoxin HigA (*pheT*), Toxin HigB, Transposase, PTS system, galactitol-specific IIA component, galactitol-specific IIC component and galactitol-specific IIB component (EC 2.7.1.200), and Triosephosphate isomerase (EC 5.3.1.1) (*tpiA*). Competence-stimulating peptide ABC transporter ATP-binding protein ComA was found in SDSE SD-SCDR1, WCHSDSE-1, and SK1250. Region two (starts 1240732– ends 1345833) contains Transposase TnpA (*tnpA*), mobile element proteins (*tra*), site-specific recombinase (*ccrB*), DNA topoisomerase I (EC 5.99.1.2), TrsK-like protein, DNA polymerase IV (EC

Streptococcus dysgalactiae subsp. *equisimilis* pan-taxonomic Antimicrobial Resistance Genes

(Resistome): Supplementary table 3 presents antimicrobial resistance genes (AMR) genes annotated in SDSE genomes and their corresponding AMR mechanism. All tested SDSE genomes contains *folA*, *Dfr*, *Ddl*, *LiaS*, *LiaR*, *pgsA*, *Alr*, *kasA*, *gidB*, *EF-G*, *EF-Tu*, *folP*, *gyrA*, *gyrB*, Iso-tRNA (*ileS*), *LiaF*, *S12p*, *S10p*, *rpoB*, *rpoC*, *MurA*, and *FabK*. Besides, all tested SDSE genomes contain *lmrP* except for two strains, namely, T642 and C161L1. All the strains hold *aadK* except for five strains, namely, NCTC6179, NCTC6181, NCTC9413, NCTC6407, NCTC10321. *rplF* gene was missing from 22 SDSE strains, namely, NCTC8543, NCTC11564, SD-SCDR1, NCTC6181, NCTC8546, NCTC9413, NCTC11554, NCTC 11556, NCTC11565, NCTC6407, NCTC10321, NCTC5969, NCTC9603, NCTC11555, NCTC11557, KNZ10, KNZ04, KNZ07, KNZ06, KNZ12, KNZ15, and KNZ16. Some SDSE strains showed some unique AMR phenotypes and genotypes. It was found that only SDSE strains AKSDE4288, strain ASDSE99, strain NCTC11557, and strain UT_4255RC have *ermA*. Moreover, only strains SDSE KNZ04 and WCHSDSE-1 have *ermB*. Also, only two *Streptococcus dysgalactiae* subsp. *equisimilis* strain Sde558 has *erm(T)*. SDSE strain AC-2713 has ABC-F type ribosomal protection protein **Msr(D)** that confer resistance to the antibiotic telithromycin and MFS efflux pump, **Mef(A)**, which is responsible for resistance against Macrolides, Lincosamides, Streptogramins, Ketolides, Oxazolidinones (MLSKO). Moreover, SDSE strains AKSDE4288, and SK1250 have the tetracycline resistance gene, *tetO*, that encodes for the ribosomal protection type (O) Tetracycline resistance. Whereas, SDSE strains NCTC11557, UT_4242_AB, UT_4234_DH, KNZ04, KNZ06, ASDSE_99, and ASDSE_96 have Tetracycline resistance gene *tetM*, that encodes for the ribosomal protection type (M) Tetracycline resistance. Furthermore, SDSE strain KNZ06 has a *parC* gene that confers resistance against fluoroquinolone antibiotics. The SDSE strain T642 a point mutation in *ParE*, a subunit of topoisomerase IV that decatenates and relaxes DNA to allow access to genes for transcription or translation that prevent Anticoumarin antibiotics from inhibiting DNA synthesis, and thus conferring resistance. Besides, it has the PhoB, phosphate regulon transcriptional regulatory protein (*SphR*), a direct regulator of the macAB efflux pump antibiotic out of a cell to confer resistance. Similarly,

SDSE strain C161L1 has **PhoB** resistance against peptide antibiotics. Moreover, strain C161L1 has the *gyrB* (parE) gene that confers resistance against aminocoumarin (novobiocin).

Streptococcus dysgalactiae subsp. equisimilis pan-taxonomic virulence factors Genes (Virulome):

Supplementary table 4 presents virulence factors (VF) genes annotated in SDSE genomes and related products. The core Virulome of SDSE consists of 38 genetic determinants. All tested SDSE genomes contains the following virulence factors *idnO*, *purB*, *asnA*, *clpP*, *bcaT* (*ilvE*), *ccpA*, *SpyM3-0013*, *pyrG*, *cydA*, *dltA*, *BCHE*, *hasC* (*galU*), *vicK*, *gidA*, *covS*, *lepA*, *luxS*, *SPy_1633*, *ykqC* (*rnjA*), *SP_0095*, *pflD*, *lgt*, *deoB*, *metQ* (*atmB*), *lsp* (*lspA*), *leuS*, *purH*, *guaA*, *glnA*, *fba*, *pstA*, *pstB*, *perR*, *oppA*, *cpsY*, *fbp54* (*fbp*, *flpA*), *perR*. Moreover, all examined strains contain *rpoE* gene, encoding for DNA-directed RNA polymerase delta subunit (EC 2.7.7.6), except for strain 167. Furthermore, all tested strains contain the *gldA* gene, encoding for Glycerol dehydrogenase (EC 1.1.1.6), except for two strains: T642 and C161L1. Similarly, all examined SDSE strains contain *slo* gene, encoding for Thiol-activated cytolysin/streptolysin O, except for two, namely, NCTC6407 and NCTC10321. Additionally, all examined strains contain the *lmb* (*psaA*) gene, which encodes for Laminin-binding surface protein, except for two strains: strain NCTC6407 and strain NCTC10321. All SDSE strains contain both *scpA* and *scpB*, encoding for C5a peptidase (EC 3.4.21.-), except for four strains, namely, NCTC6179, NCTC6181, NCTC6407, and NCTC10321. Besides, all SDSE strains contain the *ska* (*ske*) gene, encoding for Streptokinase, except for four strains, namely, NCTC6407, NCTC10321, NCTC6181, and NCTC6179. Additionally, all SDSE strains contain the *nga* gene, encoding for Nicotine adenine dinucleotide glycohydrolase (NADGH) (EC 3.2.2.5), except for four strains NCTC6407, NCTC10321, NCTC6181, and NCTC6179. Moreover, 87% of SDSE strains contain *sagA*, which encodes for Streptolysin S precursor. Additionally, only 68% of SDSE strains contained *gbs0630* and *gbs0631* genes. Similarly, only 53% of SDSE strains contain *soSIA* that encodes for Ethanolamine utilization protein. Besides, only 51% of SDSE strains contain *csoSIA* that encodes for Ethanolamine utilization polyhedral-body-like protein (EutM). Only 47% of SDSE strains contain *spg* (*spgg4*) gene encodes for Exotoxin, phage associated streptococcal pyrogenic exotoxin G

(SpeG). Only 32% of SDSE strains contain *gbs0632* that encodes for the adherence sortase-assembled pili hypothetical protein. Furthermore, only 43% of SDSE strains contain the *ccrB* gene that encodes for site-specific recombinase. Likewise, only 28% of SDSE strains contain *SP_0939* genetic determinant that encodes for DUF1706 domain-containing protein. Interestingly, some SDSE strains showed some VF unique resistance phenotypes and genotypes. Only eight SDSE strains have the *sda* gene that encodes for Streptodornase D, the exoenzyme and spreading factor, namely, 167, KNZ16, NCTC11554, UT_4031CC, UT_4242_AB, UT_4277_BB, UT_4966_RC, and WCHSDSE-1. Six strains contain *SP_1018* gene which encodes for Thymidine kinase (EC 2.7.1.21) namely, NCTC6181, NCTC10321, NCTC6407, UT_4031CC, ASDSE_96 and NCTC6179. Moreover, six SDSE strains contain *mf3* that encodes for the Streptococcal extracellular nuclease 3 (Mitogenic factor 3), namely, NCTC8543, NCTC9414, UT_4231_KK, UT_4241_XS, UT-SS1069, and SK1249. Four SDSE strains have *emm1/emm* encodes for Antiphagocytic M protein that plays an important role in Antiphagocytosis, Adherence, and Cellular invasion. These strains are ASDSE_96, KNZ06, KNZ16, and NCTC5371. Similarly, Four SDSE strains have *salt/salX* gene that encodes for Lanthionine biosynthesis protein/ Lantibiotic transport ATP-binding protein, namely, KNZ03, KNZ04, KNZ16, and AC-2713. Four SDSE strains have gene *EF1623* encodes for EutM/PduA/PduJ-like protein two clustered with choline trimethylamine-lyase NCTC11554, KNZ03, KNZ04, and KNZ16. Only three isolates have *SP_1143* that encodes for Toxin HigB, namely, strains NCTC11564, NCTC8543, and UT-5354. Similarly, three SDSE strains have the *purl* gene that encodes for Phosphoribosylformylglycinamide synthase, glutamine amidotransferase subunit (EC 6.3.5.3), namely, UT-SS1069, C161L1, and strain T642. Additionally, only two strains have the *mf/spd* gene that encodes for Streptodornase B, the exoenzyme and spreading factor, namely, NCTC6179 and NCTC6181. Additionally, only two strains have the *rumA1* gene encoding RNA methyltransferase, TrmA family, namely, AKSDE4288 and AC-2713. Only SDSE strain RE378 has the *pyrR* gene that encodes for Pyrimidine operon regulatory protein. Only SDSE strain SK1249 has the *pepF* gene that encodes for Oligoendopeptidase F-like protein. Only SDSE strain KNZ16 has the *lacA* gene that encodes for Galactose-6-phosphate isomerase, LacA subunit (EC 5.3.1.26). Only SDSE strain NCTC6179 has the *spot* gene that

encodes for Guanosine-3',5'-bis(diphosphate) 3'-pyrophosphohydrolase (EC 3.1.7.2) / GTP pyrophosphokinase (EC 2.7.6.5), (p)ppGpp synthetase II. Only SEDE strains C161L1 and T642 have the following genes: *pstC* gene that encodes for phosphate ABC transporter, permease protein, *csrR* gene that encodes for Response regulator CsrR, *ciaR* gene that encodes for two-component system response regulator CiaR, the *txrR* gene that encodes for two-component response regulator TrxR, and *trmFO* gene that encodes for Methylenetetrahydrofolate--tRNA-(uracil-5-)-methyltransferase TrmFO (EC 2.1.1.74) virulence factor. ***Streptococcus dysgalactiae* subsp. *equisimilis* pan-taxonomic drug targets:** **Supplementary table 5** presents drug target genes/products annotated in SDSE genomes and proposed drug. All tested SDSE genomes contain the following drug targets, *def*, *guaB*, *gyrB*, *murI*, *relA*, *rmlB*, *rmlC*, and *rplC*. Additionally, the *parE* gene is observed in all SDSE genomes/strains except for three strains: NCTC7136, NCTC9603, and KNZ06. Additionally, the *lacG* gene, which encodes 6-phospho-beta-galactosidase (EC 3.2.1.85) and targeted by beta-D-galactose 6-phosphate, is found in 43% of SDSE genomes/strains. Furthermore, the *spg* gene, which codes for Protein G-related-alpha-2 macroglobulin-binding protein (GRAB) and targeted by N-Formylmethionine, is 30% of SDSE genomes/strains. Furthermore, SDSE strain KNZ06 has the *parC*, which encodes for DNA topoisomerase IV subunit A (EC 5.99.1.3) and is targeted by both Gatifloxacin and Besifloxacin. Finally, SDSE strain SK1249 has Catabolite control protein A and is targeted by the experimental drug 2-Phenylamino-Ethanesulfonic Acid.

***Streptococcus dysgalactiae* subsp. *equisimilis* pan-taxonomic Complete Phage genomes:** **Supplementary table 6** and **supplementary figure 1** presents detailed information about the complete phage genomes annotated in SDSE genomes. To date, 16 complete phages are associated with SDSE strains. The species representative SDSE strain AC-2713 has two phages, namely, *Streptococcus* phage Javan122 and Javan123. In comparison, SDSE UT_4231_KK has two phages: *Streptococcus* phage Javan161 and *Streptococcus* satellite phage Javan162. Similarly, SDSE UT_4966_RC has two phages, namely, *Streptococcus* phage Javan164 and *Streptococcus* satellite phage Javan165. The SDSE strain UT-SS1069 also has two complete prophages: *Streptococcus* phage Javan170 and *Streptococcus* satellite phage

Javan171. Only SDSE strain UT-5345 has three complete phages: Streptococcus phage Javan166, Streptococcus satellite phages Javan167 Javan168. SDSE strains SDSE NS3396, SDSE strain ASDSE_99, SDSE UT_4242_AB, SDSE UT-5354, SDSE WCHSDSE-1 each has one complete phage genome. The SDSE associated phage genome lengths ranged between 8144 and 45241 bp for phage Javan167 and phage Javan166, respectively. Besides, the phages GC% content ranged between 33.3 and 41.6. Furthermore, the number of protein-coding sequences (CDS) ranged between 12 and 85 CDSs for Streptococcus satellite phage Javan167 and Streptococcus phage Javan166, respectively.

Supplementary figure 1: complete phage identified in *Streptococcus dysgalactiae* subsp. *equisimilis* strain/genomes.

Streptococcus dysgalactiae subsp. equisimilis pan-taxonomic CRISPER-CAS systems: **Supplementary table 7** presents detailed information about the CRISPER-CAS systems observed in SDSE genomes. We found that 89% of SDSE strains/genomes have CRISPER-CAS systems associated with genes/products. Two CAS systems were observed in SDSE genomes, namely, CAS-TypeIIA and CAS-

TypeIC. The core SDSE-CAS systems in SDSE contains CRISPR-associated endonuclease Cas9, CRISPR-associated helicase Cas3, CRISPR-associated protein Cas1 (II), CRISPR-associated protein Cas1 (IC), CRISPR-associated protein Cas2 (I-II-III), CRISPR-associated protein Cas2 (I-II-III-V), CRISPR-associated protein Cas5, CRISPR-associated protein, Csd1 family, CRISPR-associated protein, Csd2/Csh2 family, CRISPR-associated protein, Csn2 family and CRISPR-associated RecB family exonuclease Cas4. CAS-TypeIIA contains Csn2-0-IIA as a mandatory gene and Cas2 (I-II-III), Cas1-0-II, and Cas9 (1-II) (endonuclease Cas9) as accessory genes. While CAS-TypeIC contains Cas1(-0-IC), Csd2/Csh2 family (Cas7-0-IC, Cas8c-0-IC), Cas5(-0-IC) as mandatory genes and Cas2 (I-II-III-V), RecB family exonuclease Cas4 (Cas4-0-I-II), and Cas3(-0-I) as accessory genes. We found that 36 (68%) SDSE strains comprises the full core SDSE-CAS systems, namely, SS1575, NCTC7136, NCTC8543, SD-SCDR1, NCTC11554, NCTC11556, NCTC11555, NCTC11557, KNZ03, KNZ01, KNZ10, KNZ04, KNZ06, KNZ03, KNZ12, KNZ15, WCHSDSE-1, UT_4277_BB, UT_4234_DH, ASDSE_99, UT_4242_AB, ASDSE_96, UT_4255RC, AKSDE4288, AC-2713, RE378, SK1250, UT-5345, UT-5354, KNZ07, KNZ16, UT_4031CC, T642, NCTC6179, ATCC 12394, and NCTC6181. Besides, seven SDSE strains (13%) contain only CAS-TypeIIA, namely, NCTC5371, NCTC9414, NCTC6407, NCTC10321, NCTC9603, UT-SS957, GGS_124. We were not able to detect any CRISPER-CAS systems associated genes/products in six SDSE strains/genome (11%) namely, 167, C161L1, NCTC5370, UT_4231_KK, UT_4241_XS, UT_4966_RC, and UT-SS1069. Four strains showed different patterns of missing and/or gaining CAS genes/proteins, such as NCTC11564 that is missing both CRISPR-associated helicase Cas3 and CRISPR-associated protein, Csd1 family.

Supplementary figure 2: *Streptococcus dysgalactiae* subsp. *equisimilis* pan-taxonomic CRISPER- CAS-TypeIC systems.

Supplementary figure 3: *Streptococcus dysgalactiae* subsp. *equisimilis* pan-taxonomic CRISPR-CAS-TypeIIA systems.

Discussion:

In the present work, we present a comprehensive genomic analysis for the SDSE strain SCDR1 that belongs to Lancefield group G, emm type (*stG6*), and (MLST) sequence type (ST44) isolated from neuroischaemic diabetic foot ulcer. To our knowledge, this is the first time for this emerging pathogen to be documented in Saudi Arabia and the middle east. This isolate was characterized by its extreme beta-hemolytic ability (data not shown). Matsue et al. (Matsue et al., 2020) stated that predominant *emm* types of SDSE varied per country. For example, in central Taiwan, the *stG6.1*, *stG10*, *stG245*, *stG840*, and *stG652* strains predominated (Ma et al., 2017), whereas *stG485* and *stG6.1* were predominant in Jerusalem and Austria, respectively (Cohen-Poradosu et al., 2004; Leitner et al., 2015). Moreover, Shimomura et al. reported that *stG10.0* was the most common emm type in Portugal and Japan, *stG643.0* was the most common in western Norway, and *stG6.0* was the most recurrent in the USA (Shimomura et al., 2011). Rantala et al. conducted a retrospective population-based study of 140 episodes of *Streptococcus dysgalactiae* subsp. *equisimilis* bacteremia occurring in Finland during 1995–2004. Skin and soft tissue infections were more common clinical signs amongst cases caused by frequent emm types. They reported 18 *emm* types (including four *stG6*: *stG6.0*, *stG6.1*, *stG6.3*, and *stG6.4*). Besides, *stG480* (27 isolates), *stG6* (23 isolates), and *stG485* (22 isolates) were the three highest frequent *emm* types and represented 51% of all isolates. In all cases, *stG6* was associated with Cellulitis (Rantala et al., 2010). Moreover, Gherardi et al. investigated the genetic relationships, the virulence profile, and cell invasiveness, as well as the susceptibility to macrolide and tetracycline antibiotics of SDSE strains isolated throughout Italy from invasive and non-invasive diseases and carriers (Gherardi et al., 2014). They found that invasiveness properties seemed to be specifically correlated with the emm subtype. For example, emm subtypes *stG6*, *stC36*, and *stG6792* differed between the group G fibronectin-binding protein gene *gfbA*-positive WI/NI (weak invasive/ non-invasive) and HI (high invasive) strains. Similarly, *stG6* and *stG480* emm subtypes encountered within *gfbA*-negative WI/NI and HI strains were different, except for *st480.0*. Moreover, almost all of the most frequent emm types (*stG6*, *stC6979*, *stG480*, *stG62647*, *stG6792*, *stC839*, *stG485*, *stG643*, and *stC36*, in decreasing order) have

also been encountered elsewhere, although with different frequencies, as reported by population genetic studies on SDSE in Australia, Europe and North America (McMillan et al., 2010), reflecting the diffusion of a few successful emm types fit to disseminate in humans. The nominal sample size of the collection does not allow any statistical association between emm type and specific clinical manifestation to be inferred; thus, they cannot confirm some strong associations reported elsewhere (i.e., *stG485*, *stG480*, and *stG6* and invasive infections) (Jensen and Kilian, 2012). For example, while our findings agree with the reported association of *stC839* with non-invasive infections (Pinho et al., 2006) (Pinho et al., 2006), all nine *stG6* strains here examined were isolated from either carriers or non-invasive infections (Gherardi et al., 2014). Lu et al. (Lu et al., 2016) characterized 56 SDSE isolates collected from two tertiary hospitals in Beijing, China. They found that emm types *stG245.0*, *stG652.0*, and *stG485.0*, were prevalent, in disagreement with those reported in Norway (*stG485*, *stG6*, and *stG643* predominated) (Kittang et al., 2011) (Kittang et al., 2011), Japan (*st6792* or subtype *stG6792.3*, *stG485*, and *stG2078* dominated) (Sunaoshi et al., 2010; Takahashi et al., 2010; Takahashi et al., 2011) (Sunaoshi et al., 2010; Takahashi et al., 2010, 2011), and the United States (*stG6*, *stG245*, *stG2078*, and *stG643* dominated) (Broyles et al., 2009) (Broyles et al., 2009), respectively. Wajima et al. (Wajima et al., 2016) gathered β -hemolytic streptococci (1,611 isolates) from invasive streptococcal infections patients in Japan during April 2010–March 2013. SDSE was most frequent ($n = 693$); 99% of patients with SDSE infections were elderly (mean age 75 years, $SD \pm 15$ years). They previously reported that *stG485* and *stG6792* were more prevalent in isolates from iSDSE infections in Japan, whereas *stG10* and *stG6* were more prevalent in noninvasive strains (Takahashi et al., 2010). The most predominant was *stG6792*, which accounted for 27.1% of isolates, followed by *stG485* (13.3%), *stG245* (10.7%), *stG652* (6.8%), *stG10* (6.2%), and *stG6* (5.5%). To explain relationships among emm type and clonal multifaceted CC, the main CC was identified for each emm type. Excluding for *stG652* strains, CCs in almost all strains, *stG653* and *stG6792*, and all of those in *stG2078*, and *stG4974*, were recognized as CC17. Similarly, strains in *stG6*, *stG245*, *stG5420*, and *stG166b*, were allocated to CC25, *stC74a* to CC29, *stG10* to CC15, and *stC6979* to CC129. Numerous emm types, *stG485*, *stC36*, *stC652*, *stG480*, and *stG4222*, belonged to >2 diverse CCs. No significant association was found

between *emm* type and the fatality rate for infected patients. Moreover, Wang et al. (Wang et al., 2016) In a previous study, The MLST ST44 of SDSE was associated with five *emm* types (*stG6*, *stC36*, *stG245*, *stG480*, and *stGL265*) (McMillan et al., 2011). MLST is a nucleotide-based method for characterizing genetic relationships amongst isolates of the same bacterial species. Unlike *emm*-typing, MLST utilizes multiple housekeeping genes that are considered selectively neutral and located in different parts of the genome. MLST, therefore, provides a better tool for determining evolutionary relationships within the SDSE population than typing using the *emm* gene, which is under strong diversifying selective pressure. Recent MLST studies of SDSE isolates from Australia, Portugal, and the USA reported a high degree of genetic diversity in these populations and revealed lateral gene transfer (LGT) of housekeeping alleles was occurring (Ahmad et al., 2009; McMillan et al., 2010). McMillan et al. (McMillan et al., 2011) used MLST to characterize relationships in India's SDSE population, a country where streptococcal disease is endemic. The study revealed that Indian SDSE isolates have sequence types (STs) predominantly different from those reported from other world regions. MLST resolved the 181 isolates into 52 STs. The most common ST, ST84, was found in 43 isolates, *emm*-type *stg4831*. The five next most abundant STs (ST44, ST15, ST89, ST81, and ST107) each individually constitute 5 to 10% of the total collection, and together with ST84, account for 53% of isolates. McMillan et al. [25] investigated the molecular epidemiology, and evolutionary relationships of SDSE isolates collected from Australia, Europe, and North America using MLST analysis in a previous study. They segregated 178 SDSE isolates, representing 37 *emm* types, into 80 distinct sequence types (STs) that form 17 clonal complexes (CCs). They showed that ST44 is associated with *emm* type *stG6* found mainly in Australia and Europe and belongs to the carbohydrate group G, which all in agreement with our isolate. Our genome-genome analysis showed that SDSE strain SD-SCDR1 is closely related to SDSE strain SK1250, strain WCHSDSE-1 and strain UT-5345, with distances values of 0.00399629, 0.00467416, and 0.00687462, respectively. While whole-genome phylogenetic analysis showed that our strain SD-SCDR1 belongs to a clade that contains *Streptococcus dysgalactiae* subsp. *equisimilis* strains, namely the reference strain AC-2713 and GGS 124. In addition, within *Streptococcus dysgalactiae* subsp. *equisimilis* taxa, our strain SDSE-SCDR1 is closely related to both WCHSDSE-1 and

strain SK1250. The SDSE strain WCHSDSE-1 caused an outbreak of tonsillopharyngitis among healthcare workers in China that belongs to the Lancefield group G, *emm* type *stG211.1*, and sequence type 44 (Wang et al., 2016). Whereas strain SK1250 belonged to Lancefield group A and *emm* type *stG840* and was isolated from the human throat in Sweden in 2001. The relative closeness between SDSE-SCDR1 and both WCHSDSE-1 and SK1250 was also confirmed by in silico genome-to-genome comparisons. Both phylogenetic and genome-to-genome analyses confirmed that the SDSE-SCDR1 strain clustered together more closely than the *Streptococcus dysgalactiae subsp. equisimilis* strain AC-2713 and GGS 124. SDSE AC-2713 isolate possesses the group A cell wall carbohydrate antigen. This isolate was recovered from a blood culture of a 78-year-old man who was suffering from an acute onset of headache, 40°C fever, and chills. His medical history included liver cirrhosis and signs of portal hypertension. He presented local erythema and tenderness of his right foot, and the diagnosis of erysipelas was rendered. The site of bacterial entry was assumed to be a minor skin lesion acquired four days before admission. Antimicrobial therapy with a typical dose of intravenous amoxicillin was administered (Brandt et al., 1999). The patient's condition improved, and he was discharged six days later on an oral regimen to complete a 10-day course of amoxicillin. Whereas *Streptococcus dysgalactiae subsp. equisimilis* GGS_124 was isolated from a patient with Streptococcal toxic shock syndrome (STSS) by Shimomura et al. (Shimomura et al., 2011). This isolate belongs to Lancefield group G, ST67, and is generally used for comparative analysis by Japan's International Medical Center. It is associated with endocarditis, meningitis, and septicemia in both humans and swine (*Streptococcus dysgalactiae subsp. equisimilis* GGS_124::Genome Overview). Our SDSE-SCDR-1 isolate's genome annotation showed that it contains many genes with homology to known virulence factors, antibiotic resistance genes, and drug targets. In general, SDSE shares most of the virulence factor genes of GAS, including streptolysin O, streptokinase, CT-like regions, NADase, and DRS (for “distantly related to SIC”)(Shimomura et al., 2011). Our isolate contains genetic determinates involved in adherence to the host cell, such as *dltA*, *lmb*, *gbs0628*, *gbs0632*, *gbs0630*, *gbs0631*, and *fbp54*. Similarly, WCHSDSE-1 contains *lmb*, *pavA*, *gfbA*, and *gapC* for adherence (Wang et al., 2016), whereas GGS_124 possesses genes that encode putative adhesion proteins (Shimomura et al., 2011), including proteins similar

607 to putative fibronectin-binding proteins, pullulanase, phosphopyruvate hydratase, laminin-binding protein,
608 internalin protein, and collagen-binding protein, all of which bind to the extracellular matrix. Besides,
609 GGS_124 possesses genes encoding the multifunctional M protein (*stg480.0*) that governs the
610 antiphagocytic and adhesin activities, though the adhesion function of the GGS_124 M protein might be
611 due to a collagen-binding motif (Nitsche et al., 2006; Barroso et al., 2009). Moreover, our isolate contains
612 genetic determinates involved in cellular and tissue invasion, Antiphagocytosis, and immune evasion, such
613 as *emm1*, *emm*, and *hasC*. Whereas WCHSDSE-1 does not contain Antiphagocytosis or immune evasion
614 genetic determinants (Wang et al., 2016). SDSE-SCDR1 contains *ska*, encoding for Streptokinase, and
615 *SpyM3_0013*, encoding for cationic amino acid transporter-APC Superfamily, gene determinants for the
616 invasion of skin and soft tissue, While WCHSDSE-1 contains *skc* and *skg*, encoding for Streptokinase,
617 aiding in spreading of the pathogen. Moreover, GGS_124 also carries a gene for streptokinase
618 (SDEG_0233), similar to streptokinase A of GAS, with 88% amino acid identity. Furthermore, SDSE-
619 SCDR1 contains the *slo* gene that encodes for Thiol-activated cytolysin streptolysin O that function as a
620 toxin and pore-forming agent. Additionally, SDSE-SCDR1 contains the *sagA* gene that encodes for
621 Streptolysin S precursor (SagA), the oxygen-stable, nonimmunogenic cytotoxin, which causes a zone of
622 beta-hemolysis observed on the surface of blood agar. Similarly, WCHSDSE-1 contains both determinants
623 of streptolysins O and S. Moreover, GGS_124 also has genes encoding streptolysin S (*sagA*) and its
624 biosynthesis proteins (sag BCDEFGHI). Also, GGS_124 harbours a gene for streptolysin O (*slo*)
625 (SDEG_2027), which is necessary for GAS virulence and is essential for the organism to evade from the
626 endosome into the cytosol following the invasion of host cells (Nakagawa et al., 2004). Furthermore, SDSE-
627 SCDR1 contains the *speG* gene that encodes for the exotoxin G (phage associated *Streptococcal pyrogenic*
628 exotoxin G (SpeG)), which functions as the membrane-acting toxin (superantigen). Similarly, WCHSDSE-
629 1 contains exotoxin G encoded by the *speG*, whereas GGS_124 carries the superantigen gene, exotoxin G
630 variant 4 (*spegg4*) is homologous to GAS streptococcal exotoxin G (SpeG), with 79% amino acid identity.
631 Indeed, superantigen G and streptolysin S genes, carried by SDSE-SCDR1, are considered the most
632 important virulence factors triggering invasive diseases (Abdelsalam et al., 2010). It is worth mentioning

that SDSE-SCDR1 possesses *fbp54* gene encoding fibronectin/fibrinogen-binding protein. WCHSDSE-1 contains a similar gene, *gfbA*, the fibronectin-binding protein-encoding gene. Whereas and GGS_124 possesses genes that encode putative adhesion proteins, including proteins similar to putative fibronectin-binding proteins (SDEG_0161, 1263, and 1984). The genetic redundancy and the exact roles of *fbp54*, *gfbA*, *SDEG_0161*, and *fnB*, which causes invasive diseases, still not clear and require additional investigation. In general, we detected the presence of 53 potential virulence factors that serve in host mortality and tissue damage, toxins, pore-forming proteins, cytotoxins, beta-hemolysis agents, stress resistance thermotolerance, oxidative stress resistance, growth in blood and lung, and many other metabolic, genetic determinants that is crucial for the pathogenesis and virulence. Most of these virulence factors are present in the invasive strain GGS_124, though these factors need to be experimentally characterized. The variety of virulence factors of SDSE-SCDR1 is similar to SDSE strains causing invasive diseases. SDSE-SCDR1 contains 53 putative virulent factors that contain all the 48 putative virulent factors observed in the invasive strain GGS_124 (data not shown). Nevertheless, invasive and non-invasive strains cannot be distinguished by the range of virulence factors. As prophages may carry virulence factors, we identified and predicted the presence of incomplete prophage region SCDR1 genome. Our prophage contains protein-coding sequences similar to CDSs from Lactococcus phage, Javan S145 prophage, Listeria phage, Bacillus virus G, Bacillus phage TsarBomba, and Streptococcus phage 315.1. As we mentioned before, the *speG* gene that encodes for the exotoxin G (phage associated *Streptococcal pyrogenic* exotoxin G (SpeG)) was also detected. However, no other virulence factors identified are present on the prophage. In addition to genetic factors that confer resistance to Fluoroquinolones, Aminocoumarin, Aminoglycoside, Streptomycin, Cycloserine, Daptomycin, Pulvomycin, Fosfomycin, Fusidic acid, Isoniazid, Mupirocin, Rifampicin, Vancomycin, Daptomycin, Streptomycin Sulfonamide antibiotics, Tigecycline, Triclosan, and Trimethoprim, SDSE-SCDR1 contains *lmrP* gene that encodes for the multidrug resistance protein ErmB that confer resistance against Lincosamides, Macrolides, Streptogramins and Tetracyclines (Table 3). Wang et al. stated that isolate WCHSDSE-1 harbors only one known antimicrobial resistance gene, erm(B), mediating resistance to macrolides, clindamycin and streptogramins B13 (Wang et al., 2016). However, we

659 detected the presence of an antibiotic-resistant gene variant or mutant that leads to sulfonamide resistance
660 in WCHSDSE-1 (data not shown). On the other hand, our analysis showed that strain GGS_124 contains
661 several putative gene determinants that confer resistance to Fluoroquinolones, Aminoglycoside
662 (Streptomycin), Cycloserine. Besides, it also contains the *lmrP* gene that encodes for the multidrug
663 resistance protein ErmB and the *folP* gene variant or mutant that confer sulfonamide resistance (data not
664 shown). Prokaryotes have a system that confers resistance against infection by exogenous DNA, such as
665 viruses. In phage infection, bacteria integrate short fragments derived from phage DNA into their genomes
666 in regions called clusters of the regularly interspaced short palindromic repeat (CRISPR). Consequently,
667 bacteria utilize CRISPR-RNA transcripts and CRISPR-associated proteins (Cas) complexes to identify and
668 interfere with viral propagation (Brouns et al., 2008; Marraffini and Sontheimer, 2008). This system has
669 been detected in GAS (McShan et al., 2008), *S. equi subsp. zooepidemicus* SESZ (Holden et al., 2009), *S.*
670 *mutans* (van der Ploeg, 2009), and *S. thermophilus* (Horvath and Barrangou, 2010) and SDSE (Shimomura
671 et al., 2011; Wang et al., 2016). Two conserved CRISPR arrays were identified in the SDSE-SCDR-1
672 genome, namely, SCDR1-CRISPR2 and SCDR1-CRISPR 3. The length of SCDR1-CRISPR2 is 1286
673 bases; the repeat length is 36, the spacer length is 19 bases. While the length of SCDR1-CRISPR 3 is 962
674 bases, the repeat length is 32; the spacer length is 14 bases. Additionally, Two CAS systems were observed
675 in the SDSE-SCDR-1 genome, namely, CAS-TypeIIA and CAS-TypeIC. Though it was not mentioned
676 (Wang et al., 2016), WCHSDSE-1 contains one CRISPR array, 19 CRISPR repeats, and 18 CRISPR
677 spacers. Similarly, CAS-TypeIIA and CAS-TypeIC systems were observed in the WCHSDSE-1 genome
678 (data not shown). On the other hand, strain GGS_124 has a CRISPR/Cas system, named CRISPR1/Cas,
679 which consists of CRISPR-associated protein, Csn2 family CRISPR-associated protein Cas2, CRISPR-
680 associated protein Cas1 and, CRISPR-associated endonuclease Cas9 (data not shown). It was suggested
681 that SDSE strains typically have a higher total number of spacers than GAS, suggesting that prophage
682 infection of SDSE was relatively limited, resulting in a lower number of virulence factors positioned in the
683 prophage regions of SDSE genomes. Similar behavior was detected in SESZ when compared with *S. equi*
684 *subsp. equi* (SESE). For instance, the SESE 4047 genome contains no CRISPR but contains prophage

associated genes encoding virulence factors. On the other hand, SESZ MGCS10565 and H70 contain which 26 and 18 spacers, respectively, but do not contain any prophages. Therefore, the CRISPR system in *streptococci* sharing prophages might play a significant role in the spread of virulence factors among species. Alternatively, these virulence factors might not benefit to SDSE during carriage or disease, thus, integration of these explicit phages is not under Darwinian selection (Shimomura et al., 2011).

To present a comprehensive contemporary SDSE pan-taxonomic/subspecies comparative genomics section, we included taxonomic/genomic information of the 55 publicly available SDSE strains/isolates. Our analysis included both SDSE GGS_124 and AC-2713 strains as reference/representative genomes, in addition to our isolate SDSE-SCDR1. As we mentioned earlier, strains NCTC11565 and strain NCTC5969 then excluded from the comparative investigation because of low quality and suggested genetic material contamination with *Pseudomonas aeruginosa* genetic material in case of strain NCTC11565 and *Staphylococcus aureus* strain NCTC5969. Besides, we believe that *Streptococcus dysgalactiae* Str. iSDSE230 and *Streptococcus dysgalactiae* Str. iSDSE140 available in the public database should be grouped with SDSE strains/isolates. Thus, we included 53 strains/genomes were included in our comparative investigation. The majority of isolates were of human origin; however, two isolates were collected from the genital tract of horses, one isolate was collected from a monkey's throat, and one is from a pleural exudate of a dog. Sites of isolation included blood cultures, sterile sites of invasive diseases, oral, throat (Pharyngitis) or upper respiratory samples, foot ulcers, and Joint fluid samples.

The protein-families-based Maximum likelihood phylogenetic analysis showed that SDSE isolates are distributed amongst six significant clades. The first clade, which contains our isolate SD-SCDR1, contains nine additional SDSE strains: SK1250, WCHSDSE-1, KNZ01, UT 4242 AB, NCTC11556, NCTC6407, NCTC10321, NCTC6179, and NCTC6181 strains with a bootstrap value of 79. Whereas strain SDSE-AC 2713 is located in the second clade that contains 11 SDSE strains, namely, UT-SS957, NCTC9414, NCTC5371, AKSDE4288, NCTC9603, UT_4255RC, NCTC11557, UT_4234_DH, ASDSE 99, and NCTC11555 strains with a bootstrap value of 91. The third clade contains seven SDSE strains:

UT_4231_KK, C161L1, UT-SS1069, 167, UT_4241_XS, UT_4966_RC, and strain NCTC5370 with a bootstrap value of 100. The fourth clade contains 14 SDSE strains, namely, the invasive strain GGS_124, SS1575, UT_4031CC, NCTC11564, KNZ04, KNZ15, KNZ07, RE378, KNZ01, KNZ03, KNZ12, NCTC7136, T642, and KNZ06 with a bootstrap value of 88. The fifth clade contains seven SDSE strains: ATCC 12394, KNZ16, NCTC11554, UT-5345, KNZ03-1, KNZ10 strain SK1249. The sixth clade contains two SDSE strains, namely, UT-5354 and NCTC8543. Strains UT_4277_BB, ASDSE_96, NCTC8546, and NCTC9413 each showed an independent phylogenetic lineage. Simultaneously, strains NCTC11565 and NCTC5969 confirmed the suggested low-quality genome by being phylogenetically outcast taxa. There is no doubt that advancements are required in the field of antimicrobial/antibiotic resistance (AMR) and antibiotic susceptibility testing. Both accessory genomic elements and CRISP-Cas-mediated immunity modulate resistance elements' acquisition (van Belkum and Dunne, 2013; van Belkum et al., 2015). Henceforth, further phenotype-genotype association investigations for critical bacterial pathogens are required. Studies have suggested that genomic antibiograms are as good as phenotypic ones. The initial study involved a mixture of bacterial species resulted in 99.7% accordance between genotypes and phenotypes (Zankari et al., 2012). Additionally, investigations involved *Proteus mirabilis*, *Staphylococcus aureus*, *Escherichia coli*, *Klebsiella pneumoniae*, and *Campylobacter spp.* have expanded these findings (Stoesser et al., 2013; Gordon et al., 2014; Zhao et al., 2016; Saeb et al., 2017b). Full accordance was detected for clonal bacterial taxa such as *Mycobacterium tuberculosis* (Walker et al., 2015). Genomic approaches embrace the potential for developing imminent antibiotic susceptibility testing schemes for routine usage in clinical microbiology laboratories, even though principal knowledge breaches still need to be overcome (Jaillard et al., 2017).

***Streptococcus dysgalactiae* subsp. *equisimilis* pan-taxonomic Antimicrobial Resistance Genes (Resistome):**

We observed the presence of a common core Resistome amongst SDSE genomes/isolates and strains. It is consisting of *folA*, *Dfr*, *Ddl*, *LiaS*, *LiaR*, *pgsA*, *Alr*, *kasA*, *gidB*, *EF-G*, *EF-Tu*, *folP*, *gyrA*, *gyrB*, Iso-

tRNA (*ileS*), *LiaF*, *S12p*, *S10p*, *rpoB*, *rpoC*, *MurA*, and *FabK* AMR genetic determinants (Supplementary Table 3). These AMR genetic determinants potentially confer resistance against Trimethoprim (*folA*, *Dfr*), Cycloserine (*Ddl*, *Alr*), Daptomycin (*LiaS*, *LiaR*, *pgsA*, *LiaF*), Isoniazid (*kasA*), Aminoglycoside (Streptomycin) (*gidB*), Fusidic acid (*EF-G*), Elfamycin (Pulvomycin) (*EF-Tu*), Sulfonamide antibiotics (*folP*), Fluoroquinolones (*gyrA*, *gyrB*), Mupirocin (Pseudomonic acid) (*ileS*), Streptomycin (*S12p*), Tigecycline (*S10p*), Rifampicin, Daptomycin (*rpoB*), Rifampicin, Vancomycin, Daptomycin (*rpoC*), Fosfomycin (*MurA*) and, (*FabK*) Triclosan and Supplementary table 3 and Table 3. Besides, all tested SDSE genomes contain *lmrP* except for two strains, namely, T642 and C161L1. All the strains hold *aadK* except for five strains, namely, NCTC6179, NCTC6181, NCTC9413, NCTC6407, NCTC10321. *rplF* gene was missing from 22 SDSE strains, namely, NCTC8543, NCTC11564, SD-SCDR1, NCTC6181, NCTC8546, NCTC9413, NCTC11554, NCTC 11556, NCTC11565, NCTC6407, NCTC10321, NCTC5969, NCTC9603, NCTC11555, NCTC11557, KNZ10 KNZ04, KNZ07, KNZ06, KNZ12, KNZ15, and KNZ16. Some SDSE strains showed some unique AMR phenotypes and genotypes. It was found that only SDSE strains AKSDE4288, strain ASDSE99, strain NCTC11557, and strain UT_4255RC have *ermA*. Moreover, only strains SDSE KNZ04 and WCHSDSE-1 have *ermB*. Besides, only two *Streptococcus dysgalactiae* subsp. *equisimilis* strain Sde558 has *erm(T)*. SDSE strain AC-2713 has ABC-F type ribosomal protection protein **Msr(D)** that confer resistance to the antibiotic telithromycin and MFS efflux pump, **Mef(A)**, which is responsible for resistance against Macrolides, Lincosamides, Streptogramins, Ketolides, Oxazolidinones (MLSKO). Furthermore, SDSE strains AKSDE4288, and SK1250 have the tetracycline resistance gene, *tetO*, that encodes for the ribosomal protection type (O) Tetracycline resistance. Whereas, SDSE strains NCTC11557, UT_4242_AB, UT_4234_DH, KNZ04, KNZ06, ASDSE_99, and ASDSE_96 have Tetracycline resistance gene *tetM*, that encodes for the ribosomal protection type (M) Tetracycline resistance. In addition, SDSE strain KNZ06 has *parC* gene that confers resistance against fluoroquinolone antibiotics. The SDSE strain T642 a point mutation in *ParE*, a subunit of topoisomerase IV that decatenates and relaxes DNA to allow access to genes for transcription or translation that prevent Anticoumarin antibiotics from inhibiting DNA synthesis, and thus conferring

resistance. Besides, it has the PhoB, phosphate regulon transcriptional regulatory protein (*SphR*), a direct regulator of the macAB efflux pump antibiotic out of a cell to confer resistance. Similarly, SDSE strain C161L1 has **PhoB** resistance against peptide antibiotics. Moreover, strain C161L1 has the *gyrB* (parE) gene that confers resistance against aminocoumarin (novobiocin). Indeed, elevated rates of macrolide resistance encoded by *erm* (erythromycin ribosome methylase) and *mef* (macrolide efflux) genes have been described in the last three decades amongst the main streptococcal pathogens, including SDSE (Brandt and Spellerberg, 2009; Broyles et al., 2009; Gherardi et al., 2014). It is worth mentioning that our isolate SDSE-SCDR1 is resistant to erythromycin, clindamycin, tetracycline, and tigecycline; however, it showed sensitivity against cefotaxime and vancomycin. Similarly, SDSE- WCHSDSE-1 was susceptible to penicillin, cefotaxime, levofloxacin, and vancomycin but was resistant to erythromycin, azithromycin, clindamycin, and tetracycline (Wang et al., 2016). As we mentioned above, the genomes of SDSE consists of a core virulence structure (virulome) consisting of 38 genetic determinants. These genetic determinants are involved in the infectivity process, *idnO*, *asnA*, *bcaT*, *pyrG*, *metQ*, *pflD*, *SP_0095*, *ykqC*, *fba*, *pstA*, *deoB*, *pstB*, and *leuS*, Streptococcal pathogenesis and full virulence, *oppA*, *vicK*, *gidA*, *lgt*, and *perR*, invasive skin and soft tissue infections and systemic dissemination of GAS from the skin, **BCHE**, adhesion, tissue invasion, and bacterial virulence *fbp54*, *dlt*, *hasC*, *lepA*, *guaA*, *cydA*, *SpyM3-0013*, *luxS*, and colonization ability, *purB*, *ccpA*, *glnA* (Holmes et al., 2001; Hava and Camilli, 2002; Wang et al., 2005; Zhu et al., 2009). Besides, we observed the presence of *covS* associated with hypervirulence (Hollands et al., 2010) amongst all the genomes of SDSE. Furthermore, the stress resistance genes *clpP*, *cpsY* and *SPy_1633* were also part of the SDSE core virulome (Shelver et al., 2003; Ibrahim et al., 2005; Kang et al., 2010). Moreover, *lsp*, the evasive immune mechanism, phagosomal escape, acquisition of zinc and virulence genetics determinant, and *purH*, involved in the synthesis of purines, required for pathogenesis and virulence, were also observed (Flashner et al., 2004; Moldovan and Fraunholz, 2019) (**Supplementary Table 4 and Table 2**). All tested SDSE genomes contain the eight potential drug targets that can provide a solution for severe untreatable infections (**Supplementary table 5 and table 4**). Amongst these two are

approved drugs and six with experimental drugs. The *gyrB* gene encoded for DNA topoisomerase IV subunit B (EC 5.99.1.3) can be targeted by Gatifloxacin, an approved drug, an antibiotic, and a member of the fourth-generation fluoroquinolone family. It functions through inhibiting the bacterial enzymes DNA gyrase and topoisomerase IV. It is indicated for the treatment of bronchitis, sinusitis, community-acquired pneumonia, and skin infections (abscesses, wounds) caused by *S. pneumoniae*, *H. influenzae*, *S. aureus*, *M. pneumoniae*, *C. pneumoniae*, *L. pneumophila*, and *S. pyogenes* (Park-Wyllie et al., 2006). Also, *rplC* encode for LSU ribosomal protein L3p (L3e) can be targeted by Retapamulin and Spiramycin, both approved drugs. Retapamulin is an antibiotic for skin infections like impetigo. The FDA approved it in April 2007. Retapamulin acts by selectively inhibiting protein synthesis initiation in bacteria at the bacterial 50S ribosome level against *S. aureus* (methicillin-susceptible isolates only) or *S. pyogenes* (Jones et al., 2006). Spiramycin is a chiefly bacteriostatic macrolide antimicrobial agent with activity against Gram-positive cocci and rods, Gram-negative cocci, and Legionellae mycoplasmas, chlamydiae, some types of spirochetes, *Toxoplasma gondii*, and *Cryptosporidium*. Spiramycin is a 16-membered macrolide that inhibits translocation by attaching to bacterial 50S ribosomal subunits with an evident 1:1 stoichiometry. This antibiotic is a powerful inhibitor of the binding to the ribosome of both donor and acceptor substrates (Brisson-Noël et al., 1988). Amongst the drug targets with experimental drugs is the *def* gene that encodes for peptide deformylase (EC 3.5.1.88) that is targeted by 2-[(Formyl-Hydroxy-Amino)-Methyl]-Heptanoic Acid [1-(2-Hydroxymethyl-Pyrrolidine-1-Carbonyl)-2-Methyl-Propyl]-Amide. This experimental drug removes the formyl group from the N-terminal Met of newly synthesized proteins (Meinzel et al., 1997). Moreover, the *guaB* gene encodes for Inosine-5'-monophosphate dehydrogenase/CBS domain is targeted by Merimepodib, an experimental drug, the active inhibitor of inosine monophosphate dehydrogenase (IMPDH). IMPDH inhibition leads to a decline in intracellular guanosine triphosphate (GTP) required for DNA and RNA synthesis. It is also affected by Inosinic acid, an experimental drug that catalyzes the transformation of inosine 5'-phosphate (IMP) to xanthosine 5'-phosphate (XMP), essential for the de-novo synthesis of guanine nucleotides. As a result, it plays a vital role in regulating cell growth (Ashbaugh and

Wessels, 1995; Markland et al., 2000). **Supplementary Table 6 and supplementary figure 1** presents detailed information about the complete phage genomes annotated in SDSE genomes. To date, 16 complete phages are associated with SDSE strains. Streptococcal phages are considered critical in horizontal gene transfer, especially in the transport of virulence factors in Streptococci (Ikebe et al., 2002; Holden et al., 2009). We observed the presence of **Streptodornase D** encoding gene *sda* in both strains SDSE UT_4242_AB and SDSE UT_4966_RC. This gene is homologous to phage-encoded extracellular streptodornase D (*sdaI*) of GAS (PHA01790), responsible of the degradation of neutrophil extracellular traps (NETs), which are composed of granule proteins and chromatin released by neutrophils and can catch and kill surrounding bacteria (Buchanan et al., 2006). Besides, we observed the presence of the pleiotropic regulator of exopolysaccharide synthesis, competence, and biofilm formation Ftr, XRE family homologous with ftr protein of GAS strain SPY5003 in Streptococcus phage Javan161 and Streptococcus phage Javan170. Also, we observed the presence of the Streptococcal extracellular nuclease 3 (Mitogenic factor 3, MF3) encoded by *SPyI436*, of GAS serotype M1, in both Streptococcus phage Javan161 and Streptococcus phage Javan170. Mitogenic factor 3 plays a crucial role in the evasion of the innate immune response by degrading neutrophil extracellular traps. Besides, MF3 acts as an extracellular DNase in most clinically isolated GAS strains. Moreover, all GAS strains have extracellular nuclease activity and produce significantly more activity compared with other groups of streptococci, signifying that the activity might contribute to virulence (Wen et al., 2011). Mitogenic factor 3 is also found in Prophage GGS_124.2 of the invasive strain GGS_124 (Shimomura et al., 2011). Also, we observed the presence of pathogenicity island SaPI_n2 in Streptococcus phage phi3396. The pathogenicity island SaPI_n2 appears to be important to the pathogen, i.e. *S. aureus*, as it is maintained in the genome by two restriction-modification genes in the center of the island. Remnants of gene duplication and transfer may also be found in the transposase gene at the 5' end of SaPI_n2. This finding further supports the importance and prevalence of these proteins as agents of host/pathogen interactions (Arcus et al., 2002).

CRISPR relies on small RNAs for sequence-specific detection and cleaving of foreign nucleic acids, including bacteriophages and plasmids. Watanabe et al. reported that All SDSE strains sequenced previously harbor at least one probable CRISPR with more than 20 spacers. In contrast, their newly sequenced 167 strain does not harbor any probable CRISPR or CAS. They suggested that presumably, there is no interference by the CRISPR system in 167, and SDSE 167 may be prone to infection by phages. Moreover, analysis of a larger number of group C SDSE isolates is necessary to determine whether the absence of CRISPR interference is common to these isolates (Watanabe et al., 2013). Thus, in the 55 genomes of SDSE, we discovered that most SDSE strains/genomes (89%) have CRISPR-CAS systems associated with genes/products (**Supplementary Table 7**).

Moreover, GGS_124, the STSS causing strain, contains a CRISPR/Cas system consisting of an array of genes, *can1*, *cas1*, *cas2*, *csn2*, and CRISPR (19 direct repeats of 36 bp each and 18 spacer sequences 30 or 32 bp in length). Besides, it was found that the mean number of CRISPR spacers was higher in seven SDSE (18.0 ± 3.3 spacers) than in GAS strains (4.0 ± 1.0 spacers; range, 0 to 9), signifying that prophage infection of SDSE is to some extent constrained, resulting in a lesser number of virulence factors positioned in the prophage regions of SDSE. Then Again, SDSE may contact phages more frequently, with integrated phages having a fitness cost for SDSE (Shimomura et al., 2011). Therefore, the CRISPR system in streptococci sharing prophages may play a significant role in spreading virulence factors among species. On the other hand, these virulence factors might not be beneficial for SDSE throughout the carriage or disease, such that the integration of these individual phages is not under positive selection. We noticed two CAS systems among the investigated SDSE genomes, namely, CAS-TypeIIA and CAS-TypeIC. The signature gene for type II CRISPR-Cas systems is *cas9*, which encodes a multi-domain protein that combines all the functions of effector complexes and the target DNA cleavage and is essential for the maturation of the CRISPR RNA (crRNA) (Jinek et al., 2012). Type II-A systems encompass an additional gene, known as *csn2*, which is considered a signature gene for this subtype. The Csn2 protein is not required for interference but has an unclear role in spacer integration (Barrangou et al., 2007). All type I systems

contain the signature gene *cas3*, which encodes a large protein with a helicase possessing a single-stranded DNA (ssDNA)-stimulated ATPase activity coupled to the unwinding of DNA-DNA and RNA-DNA duplexes [40]. The signature gene for this subtype C is Cas8c and usually lacks the *cas6* gene. The Cas5 is catalytically active and replaces the Cas6 function (Makarova and Koonin, 2015). In conclusion, in the present work, we present a comprehensive genomic analysis for the SDSE strain SCDR1 that belong to Lancefield group G, emm type (stG6) and (MLST) sequence type (ST44) isolated from neuroischaemic diabetic foot ulcer, the first time for this emerging pathogen to be documented in Saudi Arabia and middle east. Strain SCDR1 has a superior spectrum of virulence factors with SDSE strains causing invasive diseases, suggesting that strains causing invasive and non-invasive diseases may not be distinguished by the presence or absence of certain virulence factors.

Moreover, we presented the most comprehensive comparative genomics analysis for the emerging human pathogen SDSE. Our study confirmed that the SDSE strains possess different characteristics in producing virulence factors for pathogenicity to humans and their genome sequence and close relationship with GAS. Our study shed light on the proposed pathogenic mechanisms of SDSE and may form the basis of molecular epidemiological research on these highly virulent bacteria.

Reference:

- Abdelsalam, M., Chen, S.-C., and Yoshida, T. (2010). Dissemination of streptococcal pyrogenic exotoxin G (speG) with an IS-like element in fish isolates of *Streptococcus dysgalactiae*. *FEMS Microbiol Lett* 309, 105–113. doi:10.1111/j.1574-6968.2010.02024.x.
- Ahmad, Y., Gertz, R. E., Li, Z., Sakota, V., Broyles, L. N., Van Beneden, C., et al. (2009). Genetic relationships deduced from emm and multilocus sequence typing of invasive *Streptococcus dysgalactiae* subsp. *equisimilis* and *S. canis* recovered from isolates collected in the United States. *J Clin Microbiol* 47, 2046–2054. doi:10.1128/JCM.00246-09.
- Akarsu, H., Bordes, P., Mansour, M., Bigot, D.-J., Genevaux, P., and Falquet, L. (2019). TASmania: A bacterial Toxin-Antitoxin Systems database. *PLOS Computational Biology* 15, e1006946. doi:10.1371/journal.pcbi.1006946.
- Amgarten, D., Braga, L. P. P., da Silva, A. M., and Setubal, J. C. (2018). MARVEL, a Tool for Prediction of Bacteriophage Sequences in Metagenomic Bins. *Front. Genet.* 9. doi:10.3389/fgene.2018.00304.

889 Arcus, V. L., Langley, R., Proft, T., Fraser, J. D., and Baker, E. N. (2002). The Three-dimensional
890 structure of a superantigen-like protein, SET3, from a pathogenicity island of the *Staphylococcus*
891 *aureus* genome. *J Biol Chem* 277, 32274–32281. doi:10.1074/jbc.M203914200.

892 Armstrong, D., Blevins, A., Louria, D. B., Henkel, J. S., Moody, M. D., and Sukany, M. (1970). Groups
893 B, C, and G streptococcal infections in a cancer hospital. *Ann. N. Y. Acad. Sci.* 174, 511–522.
894 doi:10.1111/j.1749-6632.1970.tb45577.x.

895 Ashbaugh, C. D., and Wessels, M. R. (1995). Cloning, sequence analysis and expression of the group A
896 streptococcal *guaB* gene encoding inosine monophosphate dehydrogenase. *Gene* 165, 57–60.
897 doi:10.1016/0378-1119(95)00422-3.

898 Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene
899 ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25–
900 29. doi:10.1038/75556.

901 Auckenthaler, R., Hermans, P. E., and Washington, J. A. (1983). Group G streptococcal bacteremia:
902 clinical study and review of the literature. *Rev. Infect. Dis.* 5, 196–204.
903 doi:10.1093/clinids/5.2.196.

904 Aziz, R. K., and Kotb, M. (2008). Rise and persistence of global M1T1 clone of *Streptococcus pyogenes*.
905 *Emerging Infect. Dis.* 14, 1511–1517. doi:10.3201/eid1410.071660.

906 Barnham, M. R. D., and Weightman, N. C. (2004). Changing incidence of detected streptococcal
907 bacteraemia in North Yorkshire, England. *Indian J. Med. Res.* 119 Suppl, 160–163.

908 Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., et al. (2007). CRISPR
909 provides acquired resistance against viruses in prokaryotes. *Science* 315, 1709–1712.
910 doi:10.1126/science.1138140.

911 Barroso, V., Rohde, M., Davies, M. R., Gillen, C. M., Nitsche-Schmitz, D. P., Dinkla, K., et al. (2009).
912 Identification of active variants of PARF in human pathogenic group C and group G streptococci
913 leads to an amended description of its consensus motif. *Int J Med Microbiol* 299, 547–553.
914 doi:10.1016/j.ijmm.2009.04.004.

915 Brandt, C. M., Haase, G., Schnitzler, N., Zbinden, R., and Lütticken, R. (1999). Characterization of blood
916 culture isolates of *Streptococcus dysgalactiae* subsp. *equisimilis* possessing Lancefield's group A
917 antigen. *J Clin Microbiol* 37, 4194–4197. doi:10.1128/JCM.37.12.4194-4197.1999.

918 Brandt, C. M., and Spellerberg, B. (2009). Human infections due to *Streptococcus dysgalactiae*
919 subspecies *equisimilis*. *Clin. Infect. Dis.* 49, 766–772. doi:10.1086/605085.

920 Brisson-Noël, A., Trieu-Cuot, P., and Courvalin, P. (1988). Mechanism of action of spiramycin and other
921 macrolides. *Journal of Antimicrobial Chemotherapy* 22, 13–23.
922 doi:10.1093/jac/22.Supplement_B.13.

923 Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., et al.
924 (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* 321, 960–964.
925 doi:10.1126/science.1159689.

926 Broyles, L. N., Van Beneden, C., Beall, B., Facklam, R., Shewmaker, P. L., Malpiedi, P., et al. (2009).
 927 Population-based study of invasive disease due to beta-hemolytic streptococci of groups other
 928 than A and B. *Clin. Infect. Dis.* 48, 706–712. doi:10.1086/597035.

929 Buchanan, J. T., Simpson, A. J., Aziz, R. K., Liu, G. Y., Kristian, S. A., Kotb, M., et al. (2006). DNase
 930 expression allows the pathogen group A *Streptococcus* to escape killing in neutrophil
 931 extracellular traps. *Curr Biol* 16, 396–400. doi:10.1016/j.cub.2005.12.039.

932 Cohen-Poradosu, R., Jaffe, J., Lavi, D., Grisariu-Greenzaid, S., Nir-Paz, R., Valinsky, L., et al. (2004).
 933 Group G streptococcal bacteremia in Jerusalem. *Emerging Infect. Dis.* 10, 1455–1460.
 934 doi:10.3201/eid1008.030840.

935 Davis, J. J., Gerdes, S., Olsen, G. J., Olson, R., Pusch, G. D., Shukla, M., et al. (2016). PATtyFams:
 936 Protein Families for the Microbial Genomes in the PATRIC Database. *Front Microbiol* 7, 118.
 937 doi:10.3389/fmicb.2016.00118.

938 Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput.
 939 *Nucleic Acids Res.* 32, 1792–1797. doi:10.1093/nar/gkh340.

940 Flashner, Y., Mamroud, E., Tidhar, A., Ber, R., Aftalion, M., Gur, D., et al. (2004). Generation of
 941 *Yersinia pestis* attenuated strains by signature-tagged mutagenesis in search of novel vaccine
 942 candidates. *Infect Immun* 72, 908–915. doi:10.1128/iai.72.2.908-915.2004.

943 Genteluci, G. L., Silva, L. G., Souza, M. C., Glatthardt, T., de Mattos, M. C., Ejzemberg, R., et al. (2015).
 944 Assessment and characterization of biofilm formation among human isolates of *Streptococcus*
 945 *dysgalactiae* subsp. *equisimilis*. *Int. J. Med. Microbiol.* 305, 937–947.
 946 doi:10.1016/j.ijmm.2015.10.004.

947 Gherardi, G., Imperi, M., Palmieri, C., Magi, G., Facinelli, B., Baldassarri, L., et al. (2014). Genetic
 948 diversity and virulence properties of *Streptococcus dysgalactiae* subsp. *equisimilis* from different
 949 sources. *J Med Microbiol* 63, 90–98. doi:10.1099/jmm.0.062109-0.

950 Gherardi, G., Vitali, L. A., and Creti, R. (2018). Prevalent emm Types among Invasive GAS in Europe
 951 and North America since Year 2000. *Front Public Health* 6. doi:10.3389/fpubh.2018.00059.

952 Gordon, N. C., Price, J. R., Cole, K., Everitt, R., Morgan, M., Finney, J., et al. (2014). Prediction of
 953 *Staphylococcus aureus* antimicrobial resistance by whole-genome sequencing. *J Clin Microbiol*
 954 52, 1182–1191. doi:10.1128/JCM.03117-13.

955 Haidan, A., Talay, S. R., Rohde, M., Sriprakash, K. S., Currie, B. J., and Chhatwal, G. S. (2000).
 956 Pharyngeal carriage of group C and group G streptococci and acute rheumatic fever in an
 957 Aboriginal population. *Lancet* 356, 1167–1169. doi:10.1016/S0140-6736(00)02765-3.

958 Hava, D. L., and Camilli, A. (2002). Large-scale identification of serotype 4 *Streptococcus pneumoniae*
 959 virulence factors. *Mol Microbiol* 45, 1389–1406.

960 Hawkey, J., Hamidian, M., Wick, R. R., Edwards, D. J., Billman-Jacobe, H., Hall, R. M., et al. (2015).
 961 ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence
 962 data. *BMC Genomics* 16, 667. doi:10.1186/s12864-015-1860-2.

963 Holden, M. T. G., Heather, Z., Paillot, R., Steward, K. F., Webb, K., Ainslie, F., et al. (2009). Genomic
964 evidence for the evolution of *Streptococcus equi*: host restriction, increased virulence, and genetic
965 exchange with human pathogens. *PLoS Pathog* 5, e1000346. doi:10.1371/journal.ppat.1000346.

966 Hollands, A., Pence, M. A., Timmer, A. M., Osvath, S. R., Turnbull, L., Whitchurch, C. B., et al. (2010).
967 Genetic switch to hypervirulence reduces colonization phenotypes of the globally disseminated
968 group A streptococcus MIT1 clone. *J Infect Dis* 202, 11–19. doi:10.1086/653124.

969 Holmes, A. R., McNab, R., Millsap, K. W., Rohde, M., Hammerschmidt, S., Mawdsley, J. L., et al.
970 (2001). The *pavA* gene of *Streptococcus pneumoniae* encodes a fibronectin-binding protein that
971 is essential for virulence. *Mol Microbiol* 41, 1395–1408. doi:10.1046/j.1365-2958.2001.02610.x.

972 Horvath, P., and Barrangou, R. (2010). CRISPR/Cas, the immune system of bacteria and archaea. *Science*
973 327, 167–170. doi:10.1126/science.1179555.

974 Ibrahim, Y. M., Kerr, A. R., Silva, N. A., and Mitchell, T. J. (2005). Contribution of the ATP-dependent
975 protease ClpCP to the autolysis and virulence of *Streptococcus pneumoniae*. *Infect Immun* 73,
976 730–740. doi:10.1128/IAI.73.2.730-740.2005.

977 Ikebe, T., Okuno, R., Sasaki, M., Kanda, Y., Otsuka, H., Kawahara, R., et al. (2018). Molecular
978 characterization and antibiotic resistance of *Streptococcus dysgalactiae* subspecies *equisimilis*
979 isolated from patients with streptococcal toxic shock syndrome. *J. Infect. Chemother.* 24, 117–
980 122. doi:10.1016/j.jiac.2017.09.013.

981 Ikebe, T., Wada, A., Inagaki, Y., Sugama, K., Suzuki, R., Tanaka, D., et al. (2002). Dissemination of the
982 phage-associated novel superantigen gene *speL* in recent invasive and noninvasive *Streptococcus*
983 *pyogenes* M3/T3 isolates in Japan. *Infect Immun* 70, 3227–3233. doi:10.1128/iai.70.6.3227-
984 3233.2002.

985 Jaillard, M., van Belkum, A., Cady, K. C., Creely, D., Shortridge, D., Blanc, B., et al. (2017). Correlation
986 between phenotypic antibiotic susceptibility and the resistome in *Pseudomonas aeruginosa*.
987 *International Journal of Antimicrobial Agents* 50, 210–218.
988 doi:10.1016/j.ijantimicag.2017.02.026.

989 Jensen, A., and Kilian, M. (2012). Delineation of *Streptococcus dysgalactiae*, its subspecies, and its
990 clinical and phylogenetic relationship to *Streptococcus pyogenes*. *J Clin Microbiol* 50, 113–126.
991 doi:10.1128/JCM.05900-11.

992 Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., and Charpentier, E. (2012). A
993 programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*
994 337, 816–821. doi:10.1126/science.1225829.

995 Jones, R. N., Fritsche, T. R., Sader, H. S., and Ross, J. E. (2006). Activity of retapamulin (SB-275833), a
996 novel pleuromutilin, against selected resistant gram-positive cocci. *Antimicrob Agents Chemother*
997 50, 2583–2586. doi:10.1128/AAC.01432-05.

998 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2016). KEGG as a reference
999 resource for gene and protein annotation. *Nucleic Acids Res.* 44, D457–462.
1000 doi:10.1093/nar/gkv1070.

1001 Kang, S. O., Caparon, M. G., and Cho, K. H. (2010). Virulence gene regulation by CvfA, a putative
1002 RNase: the CvfA-enolase complex in *Streptococcus pyogenes* links nutritional stress, growth-
1003 phase control, and virulence gene expression. *Infect. Immun.* 78, 2754–2767.
1004 doi:10.1128/IAI.01370-09.

1005 Kittang, B. R., Skrede, S., Langeland, N., Haanshuus, C. G., and Mylvaganam, H. (2011). emm gene
1006 diversity, superantigen gene profiles and presence of SlaA among clinical isolates of group A, C
1007 and G streptococci from western Norway. *Eur J Clin Microbiol Infect Dis* 30, 423–433.
1008 doi:10.1007/s10096-010-1105-x.

1009 Leitner, E., Zollner-Schwetz, I., Zarfel, G., Masoud-Landgraf, L., Gehrler, M., Wagner-Eibel, U., et al.
1010 (2015). Prevalence of emm types and antimicrobial susceptibility of *Streptococcus dysgalactiae*
1011 subsp. *equisimilis* in Austria. *Int. J. Med. Microbiol.* 305, 918–924.
1012 doi:10.1016/j.ijmm.2015.10.001.

1013 Lu, B., Fang, Y., Huang, L., Diao, B., Du, X., Kan, B., et al. (2016). Molecular characterization and
1014 antibiotic resistance of clinical *Streptococcus dysgalactiae* subsp. *equisimilis* in Beijing, China.
1015 *Infect Genet Evol* 40, 119–125. doi:10.1016/j.meegid.2016.01.030.

1016 Ma, J.-S., Chen, S.-Y., and Lo, H.-H. (2017). Biofilm formation of beta-hemolytic group G *Streptococcus*
1017 *dysgalactiae* subspecies *equisimilis* isolates and its association with emm polymorphism. *APMIS*
1018 125, 1027–1032. doi:10.1111/apm.12746.

1019 Makarova, K. S., and Koonin, E. V. (2015). Annotation and Classification of CRISPR-Cas Systems.
1020 *Methods Mol Biol* 1311, 47–75. doi:10.1007/978-1-4939-2687-9_4.

1021 Markland, W., McQuaid, T. J., Jain, J., and Kwong, A. D. (2000). Broad-spectrum antiviral activity of the
1022 IMP dehydrogenase inhibitor VX-497: a comparison with ribavirin and demonstration of antiviral
1023 additivity with alpha interferon. *Antimicrob Agents Chemother* 44, 859–866.
1024 doi:10.1128/aac.44.4.859-866.2000.

1025 Marraffini, L. A., and Sontheimer, E. J. (2008). CRISPR interference limits horizontal gene transfer in
1026 staphylococci by targeting DNA. *Science* 322, 1843–1845. doi:10.1126/science.1165771.

1027 Matsue, M., Ogura, K., Sugiyama, H., Miyoshi-Akiyama, T., Takemori-Sakai, Y., Iwata, Y., et al. (2020).
1028 Pathogenicity Characterization of Prevalent-Type *Streptococcus dysgalactiae* subsp. *equisimilis*
1029 Strains. *Front Microbiol* 11. doi:10.3389/fmicb.2020.00097.

1030 Matuschek, E., Brown, D. F. J., and Kahlmeter, G. (2014). Development of the EUCAST disk diffusion
1031 antimicrobial susceptibility testing method and its implementation in routine microbiology
1032 laboratories. *Clin. Microbiol. Infect.* 20, O255-266. doi:10.1111/1469-0691.12373.

1033 McMillan, D. J., Bessen, D. E., Pinho, M., Ford, C., Hall, G. S., Melo-Cristino, J., et al. (2010).
1034 Population genetics of *Streptococcus dysgalactiae* subspecies *equisimilis* reveals widely dispersed
1035 clones and extensive recombination. *PLoS ONE* 5, e11741. doi:10.1371/journal.pone.0011741.

1036 McMillan, D. J., Kaul, S. Y., Bramhachari, P. V., Smeesters, P. R., Vu, T., Karmarkar, M. G., et al.
1037 (2011). Recombination drives genetic diversification of *Streptococcus dysgalactiae* subspecies
1038 *equisimilis* in a region of streptococcal endemicity. *PLoS One* 6, e21346.
1039 doi:10.1371/journal.pone.0021346.

1040 McShan, W. M., Ferretti, J. J., Karasawa, T., Suvorov, A. N., Lin, S., Qin, B., et al. (2008). Genome
1041 sequence of a nephritogenic and highly transformable M49 strain of *Streptococcus pyogenes*. *J*
1042 *Bacteriol* 190, 7773–7785. doi:10.1128/JB.00672-08.

1043 Meinnel, T., Lazennec, C., Villoing, S., and Blanquet, S. (1997). Structure-function relationships within
1044 the peptide deformylase family. Evidence for a conserved architecture of the active site involving
1045 three conserved motifs and a metal ion. *J Mol Biol* 267, 749–761. doi:10.1006/jmbi.1997.0904.

1046 Miyoshi-Akiyama, T., Zhao, J., Kato, H., Kikuchi, K., Totsuka, K., Kataoka, Y., et al. (2003).
1047 *Streptococcus dysgalactiae*-derived mitogen (SDM), a novel bacterial superantigen:
1048 characterization of its biological activity and predicted tertiary structure. *Mol Microbiol* 47,
1049 1589–1599. doi:10.1046/j.1365-2958.2003.03411.x.

1050 Moldovan, A., and Fraunholz, M. J. (2019). In or out: Phagosomal escape of *Staphylococcus aureus*. *Cell*
1051 *Microbiol* 21, e12997. doi:10.1111/cmi.12997.

1052 Nakagawa, I., Amano, A., Mizushima, N., Yamamoto, A., Yamaguchi, H., Kamimoto, T., et al. (2004).
1053 Autophagy defends cells against invading group A *Streptococcus*. *Science* 306, 1037–1040.
1054 doi:10.1126/science.1103966.

1055 Nitsche, D. P., Johansson, H. M., Frick, I.-M., and Mörgelin, M. (2006). Streptococcal protein FOG, a
1056 novel matrix adhesin interacting with collagen I in vivo. *J Biol Chem* 281, 1670–1679.
1057 doi:10.1074/jbc.M506776200.

1058 Ogura, K., Okumura, K., Shimizu, Y., Kirikae, T., and Miyoshi-Akiyama, T. (2018). Pathogenicity
1059 Induced by Invasive Infection of *Streptococcus dysgalactiae* subsp. *equisimilis* in a Mouse Model
1060 of Diabetes. *Front Microbiol* 9, 2128. doi:10.3389/fmicb.2018.02128.

1061 Ondov, B. D., Treangen, T. J., Melsted, P., Mallonee, A. B., Bergman, N. H., Koren, S., et al. (2016).
1062 Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol.* 17, 132.
1063 doi:10.1186/s13059-016-0997-x.

1064 Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H.-Y., Cohoon, M., et al. (2005). The
1065 subsystems approach to genome annotation and its use in the project to annotate 1000 genomes.
1066 *Nucleic Acids Res.* 33, 5691–5702. doi:10.1093/nar/gki866.

1067 Park-Wyllie, L. Y., Juurlink, D. N., Kopp, A., Shah, B. R., Stukel, T. A., Stumpo, C., et al. (2006).
1068 Outpatient gatifloxacin therapy and dysglycemia in older adults. *N Engl J Med* 354, 1352–1361.
1069 doi:10.1056/NEJMoa055191.

1070 Pinho, M. D., Melo-Cristino, J., and Ramirez, M. (2006). Clonal relationships between invasive and
1071 noninvasive Lancefield group C and G streptococci and emm-specific differences in invasiveness.
1072 *J Clin Microbiol* 44, 841–846. doi:10.1128/JCM.44.3.841-846.2006.

1073 Plainvert, C., Doloy, A., Loubinoux, J., Lepoutre, A., Collobert, G., Touak, G., et al. (2012). Invasive
1074 group A streptococcal infections in adults, France (2006–2010). *Clin. Microbiol. Infect.* 18, 702–
1075 710. doi:10.1111/j.1469-0691.2011.03624.x.

1076 Rantala, S. (2014). *Streptococcus dysgalactiae* subsp. *equisimilis* bacteremia: an emerging infection. *Eur.*
1077 *J. Clin. Microbiol. Infect. Dis.* 33, 1303–1310. doi:10.1007/s10096-014-2092-0.

1078 Rantala, S., Vähäkuopus, S., Vuopio-Varkila, J., Vuento, R., and Syrjänen, J. (2010). *Streptococcus*
1079 *dysgalactiae* subsp. *equisimilis* Bacteremia, Finland, 1995–2004. *Emerg Infect Dis* 16, 843–846.
1080 doi:10.3201/eid1605.080803.

1081 Reid, H. F., Bassett, D. C., Poon-King, T., Zabriskie, J. B., and Read, S. E. (1985). Group G streptococci
1082 in healthy school-children and in patients with glomerulonephritis in Trinidad. *J Hyg (Lond)* 94,
1083 61–68. doi:10.1017/s0022172400061131.

1084 Rider, M. A., and McGregor, J. C. (1994). Group G streptococcus--an emerging cause of graft loss? *Br J*
1085 *Plast Surg* 47, 346–348. doi:10.1016/0007-1226(94)90094-9.

1086 Saeb, A. T., Abouelhoda, M., Selvaraju, M., Althawadi, S. I., Mutabagani, M., Adil, M., et al. (2017a).
1087 The Use of Next-Generation Sequencing in the Identification of a Fastidious Pathogen: A Lesson
1088 From a Clinical Setup. *Evol Bioinform Online* 13. doi:10.1177/1176934316686072.

1089 Saeb, A. T. M. (2018). Current Bioinformatics resources in combating infectious diseases. *Bioinformation*
1090 14, 31–35. doi:10.6026/97320630014031.

1091 Saeb, A. T. M., Al-Rubeaan, K. A., Abouelhoda, M., Selvaraju, M., and Tayeb, H. T. (2017b). Genome
1092 sequencing and analysis of the first spontaneous Nanosilver resistant bacterium *Proteus mirabilis*
1093 strain SCDR1. *Antimicrob Resist Infect Control* 6, 119. doi:10.1186/s13756-017-0277-x.

1094 Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., et al. (2004). BRENDA, the
1095 enzyme database: updates and major new developments. *Nucleic Acids Res.* 32, D431-433.
1096 doi:10.1093/nar/gkh081.

1097 Sevin, E. W., and Barloy-Hubler, F. (2007). RASTA-Bacteria: a web-based tool for identifying toxin-
1098 antitoxin loci in prokaryotes. *Genome Biol* 8, R155. doi:10.1186/gb-2007-8-8-r155.

1099 Shelver, D., Rajagopal, L., Harris, T. O., and Rubens, C. E. (2003). MtaR, a regulator of methionine
1100 transport, is critical for survival of group B streptococcus in vivo. *J. Bacteriol.* 185, 6592–6599.
1101 doi:10.1128/jb.185.22.6592-6599.2003.

1102 Shimomura, Y., Okumura, K., Murayama, S. Y., Yagi, J., Ubukata, K., Kirikae, T., et al. (2011).
1103 Complete genome sequencing and analysis of a Lancefield group G *Streptococcus dysgalactiae*
1104 subsp. *equisimilis* strain causing streptococcal toxic shock syndrome (STSS). *BMC Genomics* 12,
1105 17. doi:10.1186/1471-2164-12-17.

1106 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large
1107 phylogenies. *Bioinformatics* 30, 1312–1313. doi:10.1093/bioinformatics/btu033.

1108 Stamatakis, A., Hoover, P., and Rougemont, J. (2008). A rapid bootstrap algorithm for the RAxML Web
1109 servers. *Syst. Biol.* 57, 758–771. doi:10.1080/10635150802429642.

1110 Stoesser, N., Batty, E. M., Eyre, D. W., Morgan, M., Wyllie, D. H., Del Ojo Elias, C., et al. (2013).
1111 Predicting antimicrobial susceptibilities for *Escherichia coli* and *Klebsiella pneumoniae* isolates
1112 using whole genomic sequence data. *J Antimicrob Chemother* 68, 2234–2244.
1113 doi:10.1093/jac/dkt180.

1114 *Streptococcus dysgalactiae* subsp. *equisimilis* GGS_124::Genome Overview Available at:
1115 <https://www.patricbrc.org/view/Genome/486410.3> [Accessed December 10, 2020].

1116 Sunaoshi, K., Murayama, S. Y., Adachi, K., Yagoshi, M., Okuzumi, K., Chiba, N., et al. (2010).
 1117 Molecular emm genotyping and antibiotic susceptibility of *Streptococcus dysgalactiae* subsp.
 1118 *equisimilis* isolated from invasive and non-invasive infections. *J Med Microbiol* 59, 82–88.
 1119 doi:10.1099/jmm.0.013201-0.

1120 Takahashi, T., Sunaoshi, K., Sunakawa, K., Fujishima, S., Watanabe, H., Ubukata, K., et al. (2010).
 1121 Clinical aspects of invasive infections with *Streptococcus dysgalactiae* ssp. *equisimilis* in Japan:
 1122 differences with respect to *Streptococcus pyogenes* and *Streptococcus agalactiae* infections. *Clin.*
 1123 *Microbiol. Infect.* 16, 1097–1103. doi:10.1111/j.1469-0691.2009.03047.x.

1124 Takahashi, T., Ubukata, K., and Watanabe, H. (2011). Invasive infection caused by *Streptococcus*
 1125 *dysgalactiae* subsp. *equisimilis*: characteristics of strains and clinical features. *Journal Of*
 1126 *Infection And Chemotherapy: Official Journal Of The Japan Society Of Chemotherapy* 17, 1–10.
 1127 doi:10.1007/s10156-010-0084-2.

1128 van Belkum, A., and Dunne, W. M. (2013). Next-generation antimicrobial susceptibility testing. *J Clin*
 1129 *Microbiol* 51, 2018–2024. doi:10.1128/JCM.00313-13.

1130 van Belkum, A., Soriaga, L. B., LaFave, M. C., Akella, S., Veyrieras, J.-B., Barbu, E. M., et al. (2015).
 1131 Phylogenetic Distribution of CRISPR-Cas Systems in Antibiotic-Resistant
 1132 *Pseudomonas aeruginosa*. *mBio* 6, e01796-01715. doi:10.1128/mBio.01796-15.

1133 van der Ploeg, J. R. (2009). Analysis of CRISPR in *Streptococcus mutans* suggests frequent occurrence of
 1134 acquired immunity against infection by M102-like bacteriophages. *Microbiology (Reading)* 155,
 1135 1966–1976. doi:10.1099/mic.0.027508-0.

1136 Wajima, T., Morozumi, M., Hanada, S., Sunaoshi, K., Chiba, N., Iwata, S., et al. (2016). Molecular
 1137 Characterization of Invasive *Streptococcus dysgalactiae* subsp. *equisimilis*, Japan. *Emerging*
 1138 *Infect. Dis.* 22, 247–254. doi:10.3201/eid2202.141732.

1139 Wajima, T., Murayama, S. Y., Sunaoshi, K., Nakayama, E., Sunakawa, K., and Ubukata, K. (2008).
 1140 Distribution of emm type and antibiotic susceptibility of group A streptococci causing invasive
 1141 and noninvasive disease. *J. Med. Microbiol.* 57, 1383–1388. doi:10.1099/jmm.0.2008/002642-0.

1142 Walker, T. M., Kohl, T. A., Omar, S. V., Hedge, J., Del Ojo Elias, C., Bradley, P., et al. (2015). Whole-
 1143 genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and
 1144 resistance: a retrospective cohort study. *Lancet Infect Dis* 15, 1193–1202. doi:10.1016/S1473-
 1145 3099(15)00062-6.

1146 Wang, C.-H., Lin, C.-Y., Luo, Y.-H., Tsai, P.-J., Lin, Y.-S., Lin, M. T., et al. (2005). Effects of
 1147 oligopeptide permease in group A streptococcal infection. *Infect. Immun.* 73, 2881–2890.
 1148 doi:10.1128/IAI.73.5.2881-2890.2005.

1149 Wang, X., Zhang, X., and Zong, Z. (2016). Genome sequence and virulence factors of a group G
 1150 *Streptococcus dysgalactiae* subsp. *equisimilis* strain with a new element carrying *erm(B)*. *Sci Rep*
 1151 6, 20389. doi:10.1038/srep20389.

1152 Watanabe, S., Kirikae, T., and Miyoshi-Akiyama, T. (2013). Complete Genome Sequence of
 1153 *Streptococcus dysgalactiae* subsp. *equisimilis* 167 Carrying Lancefield Group C Antigen and
 1154 Comparative Genomics of *S. dysgalactiae* subsp. *equisimilis* Strains. *Genome Biol Evol* 5, 1644–
 1155 1651. doi:10.1093/gbe/evt117.

1156 Watanabe, S., Takemoto, N., Ogura, K., and Miyoshi-Akiyama, T. (2016). Severe invasive streptococcal
1157 infection by *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subsp. *equisimilis*.
1158 *Microbiol. Immunol.* 60, 1–9. doi:10.1111/1348-0421.12334.

1159 Wen, Y.-T., Tsou, C.-C., Kuo, H.-T., Wang, J.-S., Wu, J.-J., and Liao, P.-C. (2011). Differential
1160 secretomics of *Streptococcus pyogenes* reveals a novel peroxide regulator (PerR)-regulated
1161 extracellular virulence factor mitogen factor 3 (MF3). *Mol Cell Proteomics* 10, M110.007013.
1162 doi:10.1074/mcp.M110.007013.

1163 Yung, L., Rashid, M., and Bräu, N. (2019). Group C *Streptococcus dysgalactiae* subsp. *equisimilis*
1164 (SDSE) Endocarditis with Endogenous Endophthalmitis and Aortic Root Abscess. *IDCases* 15,
1165 e00513. doi:10.1016/j.idcr.2019.e00513.

1166 Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., et al. (2012).
1167 Identification of acquired antimicrobial resistance genes. *J. Antimicrob. Chemother.* 67, 2640–
1168 2644. doi:10.1093/jac/dks261.

1169 Zhao, S., Tyson, G. H., Chen, Y., Li, C., Mukherjee, S., Young, S., et al. (2016). Whole-Genome
1170 Sequencing Analysis Accurately Predicts Antimicrobial Resistance Phenotypes in *Campylobacter*
1171 spp. *Appl Environ Microbiol* 82, 459–466. doi:10.1128/AEM.02873-15.

1172 Zhu, H., Liu, M., Sumby, P., and Lei, B. (2009). The secreted esterase of group A streptococcus is
1173 important for invasive skin infection and dissemination in mice. *Infect Immun* 77, 5225–5232.
1174 doi:10.1128/IAI.00636-09.

1175

1176

1177

1178

1179

1180

1181

1182

1183

1184

1185

1186

1187

1188

Figures titles:

Figure 1: Genome representation shows from outer to inner rings, the contigs, CDS on the forward strand, CDS on the reverse strand, RNA genes, CDS with homology to known antimicrobial resistance genes, CDS with homology to known virulence factors, GC content, and GC skew.

Figure 2: Mauve Multiple Whole Genome Alignment of SDSE strain SD-SCDR1 against reference genomes SDSE strain NCTC7136, SDSE strain WCHSDSE-1 and *Streptococcus pyogenes* strain MGAS15252.

Figure 3: An overview of the biological subsystems and genes for *Streptococcus dysgalactiae* subsp. *equisimilis* strain SD-SCDR1 genome.

Figure 4: (a) Phylogenetic tree of SDSE based on whole-genome sequences (reference and outgroup genomes) and (b) selected 33 SDSE genome sequences, *Streptococcus dysgalactiae* subsp. *equisimilis* strain SD-SCDR1, *Streptococcus pyogenes* MGAS15252, and *Staphylococcus aureus* MR1 genomes.

Figure 5: incomplete prophage region identified in *Streptococcus dysgalactiae* subsp. *equisimilis* strain SD-SCDR1 genome

Figure 6: protein families based on Maximum likelihood phylogenetic analysis for the publicly available SDSE strains/isolates genomes.

Figure 7: protein families phylogeny based proteomic comparison among SDSE-SD-SCDR1 clade containing strains (a) All gene/proteins (b) Essential gene/proteins

Tables titles:

Table1: The genome-genome comparison between SDSE strain SCDR-SD1 and other SDSE strains.

Table 2: Virulence factors genetic determinants observed in SD-SCDR-1 genome.

Table 3: antimicrobial resistance genes (AMR) genes annotated in SD-SCDR1 genome and corresponding AMR mechanism

Table 4: SDSE SD-SCDR1 genome potential novel drug targets.

Table 5: SDSE SD-SCDR1 genome CRISPR repeats, spaces, and arrays.

- 1221 **Table 6:** CRISPR repeats information identified in SD-SCDR-1 genome.
- 1222 **Table 7:** CAS systems information identified in SD-SCDR-1 genome.
- 1223 **Table 8:** incomplete prophage region/CDS identified in *Streptococcus dysgalactiae* subsp. *equisimilis*
- 1224 strain SD-SCDR1 genome
- 1225 **Table 9:** 40 Insertion sequences in SDSE-SCDR1 genome
- 1226
- 1227 **Supplemental Figures titles:**
- 1228 **Supplementary Figure 1:** complete phage identified in *Streptococcus dysgalactiae* subsp. *equisimilis*
- 1229 strain/genomes.
- 1230 **Supplementary Figure 2:** *Streptococcus dysgalactiae* subsp. *equisimilis* pan-taxonomic CRISPER-
- 1231 CAS-TypeIC systems.
- 1232 **Supplementary Figure 3:** *Streptococcus dysgalactiae* subsp. *equisimilis* pan-taxonomic CRISPER-CAS-
- 1233 TypeIIA systems.
- 1234 **Supplementary Tables titles:**
- 1235 **Supplementary Table 1:** represents the taxonomic/genomic information of the 55 publicly available
- 1236 SDSE strains/isolates.
- 1237 **Supplementary Table 2:** Reference strain SDSE-AC 2713 essential 189 genes predicted using LB
- 1238 media's metabolic flux-balance analysis.
- 1239
- 1240 **Supplementary Table 3:** Antimicrobial resistance genes (AMR) genes annotated in SDSE genomes and
- 1241 corresponding AMR mechanism
- 1242
- 1243 **Supplementary Table 4:** Major virulence factors (VF) genes annotated in SDSE genomes and
- 1244 corresponding products.
- 1245 **Supplementary Table 5:** Drug target genes/products annotated in SDSE genomes and proposed drug.
- 1246 **Supplementary Table 6:** Detailed information for the complete phage genomes annotated in SDSE
- 1247 genomes.
- 1248 **Supplementary Table 7:** Detailed information on CRISPER-CAS systems detected in SDSE genomes.
- 1249

1250 *** List of abbreviations**

1251 **SDSE:** *Streptococcus dysgalactiae* subsp. *equisimilis*

1252 **GAS:** *Streptococcus pyogenes*

1253 **MLST:** Multilocus sequence typing

1254 **PATRIC:** Bacterial bioinformatics database and analysis resource

1255 **ARDB:** Antibiotic Resistance Genes Database

1256 **CARD:** The Comprehensive Antibiotic Resistance Database

1257 **DFU:** Diabetic foot ulcer

1258 **ABPI:** The ankle-brachial pressure index

1259 **CRISPR:** Clustered regularly interspaced short palindromic repeats

1260 **AMR:** The antimicrobial resistance genes

1261 **MLSKO:** Macrolides, Lincosamides, Streptogramins, Ketolides, and Oxazolidinones

1262 **VF:** Virulence factors

1263 **GRAB:** G-related-alpha-2 macroglobulin-binding protein

1264 **WI/Ni:** Weak invasive/ Non-invasive

1265 **HI:** High invasive

1266 **STSS:** Streptococcal toxic shock syndrome

1267 ***erm*:** erythromycin ribosome methylase

1268 ***mef*:** macrolide efflux

1269

1270 **Declarations:**

1271 *** Ethics approval and consent to participate**

1272 The institutional review board approved this study at King Saud University, College of Medicine Riyadh,

1273 Kingdom of Saudi Arabia (IRB-E-16-1839). The subjects provided written informed consent for

1274 participating in this study.

1275 *** Consent to publish**

1276 All authors have consented to the publication of this manuscript.

1277 *** Availability of data and materials**

1278 The genome sequence was submitted to NCBI with an accession number CP033391 under BioProject
1279 accession number PRJNA498183, BioSample accession number SAMN10285107 and assembly
1280 accession number GCA_009497975.1.

1281 *** Competing interests**

1282 The authors declare that they have no competing interests

1283 *** Funding**

1284 The authors received an internal research fund from King Faisal specialist hospital and research center to
1285 support the publication.

1286 *** Authors' contributions**

1287 **ATMS:** Involved in study conception and design, data analysis, and interpretation and involved in
1288 drafting the manuscript or revising it critically for important intellectual content. Preparing the final
1289 approval of the version to be published.

1290 **HT:** Involved in study conception and design and drafted the manuscript or revised it critically for
1291 important intellectual content. She was involved in preparing the final approval of the version to be
1292 published.

1293 **Acknowledgment:**

1294 The authors want to thank the University Diabetes Center members at King Saud University for their help
1295 throughout the study. Besides, we want to acknowledge that NGS experiments and analysis were supported
1296 by the Saudi Human Genome Program (SHGP) at KACST and KFSHRC.