



HAL
open science

Integrative genomic, virulence, and transcriptomic analysis of emergent *Streptococcus dysgalactiae* subspecies *equisimilis* (SDSE) emm type stG62647 isolates causing human infections

Jesus M Eraso, Randall J Olsen, S Wesley Long, Ryan Gadd, Sarrah Boukthir, Ahmad Faili, Samer Kayal, James M. Musser

► To cite this version:

Jesus M Eraso, Randall J Olsen, S Wesley Long, Ryan Gadd, Sarrah Boukthir, et al.. Integrative genomic, virulence, and transcriptomic analysis of emergent *Streptococcus dysgalactiae* subspecies *equisimilis* (SDSE) emm type stG62647 isolates causing human infections. *mBio*, 2024, pp.e0257824. 10.1128/mbio.02578-24 . hal-04765930

HAL Id: hal-04765930

<https://hal.science/hal-04765930v1>

Submitted on 4 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Integrative genomic, virulence, and transcriptomic analysis of emergent *Streptococcus dysgalactiae* subspecies *equisimilis* (SDSE) *emm* type *stG62647* isolates causing human infections

Jesus M. Eraso,^{1,2,3} Randall J. Olsen,^{1,2,3,4} S. Wesley Long,^{1,2,3,4} Ryan Gadd,^{1,2} Sarrah Boukthir,^{5,6,7} Ahmad Faili,^{6,8,9} Samer Kayal,^{5,6,7,9} James M. Musser^{1,2,3,4}

AUTHOR AFFILIATIONS See affiliation list on p. 17.

ABSTRACT *Streptococcus dysgalactiae* subspecies *equisimilis* (SDSE) is a Gram-positive bacterial pathogen that infects humans and is closely related to group A streptococcus (GAS). Compared with GAS, far less is known about SDSE pathobiology. Increased rates of invasive SDSE infections have recently been reported in many countries. One SDSE *emm* type (*stG62647*) is known to cause severe diseases, including necrotizing soft-tissue infections, endocarditis, and osteoarticular infections. To increase our understanding of the molecular pathogenesis of *stG62647* SDSE isolates causing human infections, we sequenced to closure the genomes of 120 *stG62647* SDSE isolates. The genomes varied in size from 2.1 to 2.24 Mb pairs. The great majority of *stG62647* isolates had *IS1548* integrated into the *silB* gene, thereby inactivating it. Regions of difference, such as mobile genetic elements, were the largest source of genomic diversity. All 120 *stG62647* isolates were assayed for virulence using a well-established mouse model of necrotizing myositis. An unexpectedly wide range of virulence was identified (20% to 95%), as assessed by near-mortality data. To explore the molecular mechanisms underlying virulence differences, we analyzed RNAseq transcriptome profiles for 38 *stG62647* isolates (comprising the 19 least and most virulent) grown *in vitro*. Genetic polymorphisms were identified from whole-genome sequence data. Collectively, the results suggest that these SDSE isolates use multiple genetic pathways to alter virulence phenotype. The data also suggest that human genetics and underlying medical conditions contribute to disease severity. Our study integrates genomic, mouse virulence, and RNAseq data to advance our understanding of SDSE pathobiology and its molecular pathogenesis.

IMPORTANCE This study integrated genomic sequencing, mouse virulence assays, and bacterial transcriptomic analysis to advance our understanding of the molecular mechanisms contributing to *Streptococcus dysgalactiae* subsp. *equisimilis emm* type *stG62647* pathogenesis. We tested a large cohort of genetically closely related *stG62647* isolates for virulence using an established mouse model of necrotizing myositis and discovered a broad spectrum of virulence phenotypes, with near-mortality rates ranging from 20% to 95%. This variation was unexpected, given their close genetic proximity. Transcriptome analysis of *stG62647* isolates responsible for the lowest and highest near-mortality rates suggested that these isolates used multiple molecular pathways to alter their virulence. In addition, some genes encoding transcriptional regulators and putative virulence factors likely contribute to SDSE *emm* type *stG62647* pathogenesis. These data underscore the complexity of pathogen–host interactions in an emerging SDSE clonal group.

KEYWORDS *Streptococcus dysgalactiae*, genomics, pathogenesis, emerging clone

Editor Nancy E. Freitag, University of Illinois Chicago, Chicago, Illinois, USA

Address correspondence to James M. Musser, jmmusser@houstonmethodist.org.

The authors declare no conflict of interest.

Received 23 August 2024

Accepted 16 September 2024

Published 17 October 2024

Copyright © 2024 Eraso et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Streptococcus dysgalactiae subspecies *equisimilis* (SDSE) is a Gram-positive bacterial pathogen that infects humans and a variety of animals, including pigs, cows, sheep, horses, dogs, fish, and skunks (1–5). Whole genome sequencing has shown that SDSE isolates are composed of several host-specific sublineages, indicating that cross-species transmission is rare (6). The various host-specific sublineages have been reported to have distinct arrays of virulence factors, for example including adhesions and toxins that likely contribute to host specificity (6). SDSE isolates generally express Lancefield group C or G antigen, although some isolates can be group A or L. Genetic analysis has shown that SDSE is phylogenetic closely related to *Streptococcus pyogenes* (group A streptococcus; GAS) (7), likely sharing a common ancestor, although the timing is not clear (8). Compared with GAS, relatively little is known about the genomic, transcriptomic, and molecular pathogenesis mechanisms of SDSE (9–11). Unlike GAS, few experimentally confirmed virulence factors have been identified in SDSE (10–15), although several have been proposed based on homology to GAS genes (16–22).

Similar to GAS, SDSE can cause a spectrum of human infections, ranging from mild pharyngitis, tonsillitis, and skin infections, to severe invasive infections (23, 24), such as necrotizing myositis (24, 25), bacteremia (26), osteoarticular infections (OAI) (21, 27), and toxic shock syndrome (STSS) (25, 28, 29). Rarer infections include pneumonia, meningitis, and endocarditis (22, 30–32). Although the exact disease burden of SDSE remains unquantified, increased rates of invasive SDSE infections have recently been reported in several countries (22, 28, 33, 34). Concurrently, the emergence of SDSE isolates resistant to key antibiotics, including macrolides, fluoroquinolones, erythromycin, clindamycin, tetracycline, chloramphenicol, and beta-lactams, is an increasing public health concern (22, 25, 35–41).

The recently emerged SDSE *emm* type (*stG62647*) has been repeatedly associated with severe clinical manifestations, including STSS and necrotizing soft-tissue infections, in several countries (9, 18, 28, 42–47). We recently examined the virulence attributes of three SDSE isolates, including two *emm* type *stG62647* and one *emm* type *stC74a* isolate, in a mouse model of necrotizing myositis (9). The two *stG62647* isolates were more virulent than the *stC74a* isolate in this infection model. The *stG62647* isolates were multilocus sequence type clonal complex 20 (CC20), whereas the *stC74a* isolate was a genetically distinct CC17. This observation is consistent with recent epidemiological reports indicating an increase in both the frequency and severity of human infections caused by *stG62647* isolates. However, our initial analysis was limited to only a few isolates.

The goal of the current study was to expand our understanding of the pathobiology of an emerging SDSE lineage using an integrated approach involving: (i) complete genomic sequencing of all 120 *emm* type *stG62647* isolates identified in a comprehensive sample of 499 isolates cultured from humans and recovered between 2010 and 2018 in French Brittany (9); (ii) virulence assessment of the 120 *stG62647* isolates using a well-established mouse model of necrotizing myositis; and (iii) comparative transcriptome analysis of a subset of 38 *stG62647* isolates (35 CC20 and 3 CC17) grown *in vitro*. This subset included the 19 isolates causing the lowest and 19 isolates causing the highest near-mortality rates in the mouse necrotizing myositis virulence assay. The data from this integrative analysis address several knowledge deficits about an important emerging human bacterial pathogen and provide new data on the emergent *stG62647* clonal group causing severe human infections.

MATERIALS AND METHODS

Bacterial isolates and growth media

SDSE isolates were grown *in vitro* in Todd–Hewitt broth (Becton, Dickinson; BD) supplemented with 0.2% yeast extract (THY medium). To ensure consistency and minimize variations in the transcriptome due to differences in growth media composition, all media were prepared from a single master lot. Todd–Hewitt Broth (Research

Products International, Mount Prospect, IL) (3×500 g; lot 7274–7951) and Yeast Extract (Life Technologies Corporation, Miami, FL) (100 g; lot 2397659) were combined and mixed. Aliquots were dispensed to make liquid media (32 g/L of water) as needed. In total, 6 L of broth was prepared by autoclaving 2 L in three 4 L flasks (KIMAX Kimble), as described (48). For solid media, SDSE isolates were grown on trypticase soy agar plates supplemented with 5% sheep red blood cells (Becton, Dickinson and Company). RNAseq experiments were performed in triplicate on SDSE isolates harvested at mid-exponential growth phase ($OD_{600} = 1$).

Whole genome sequencing

The SDSE genomes were sequenced with methods described previously (9, 48–50). Briefly, isolates were grown at 37°C with 5% CO_2 either on tryptic soy agar with 5% sheep blood (Becton, Dickinson, Franklin Lakes, NJ) or in Todd–Hewitt broth with 2% (wt/vol) yeast extract (THY; Difco Laboratories, Franklin Lakes, NJ). For Illumina short-read sequencing, chromosomal DNA was isolated with the RNAdvance viral kit (Beckman Coulter, Brea, CA) and a BioMek i7 instrument (Beckman Coulter). Libraries were prepared with the NexteraXT kit (Illumina, San Diego, CA) and sequenced with a NovaSeq instrument (Illumina) using a 2×250 bp protocol. The average depth of coverage for the *stG62647* isolates was 382 (range, 157–1639-fold coverage). For Oxford Nanopore Technologies long-read sequencing, chromosomal DNA was isolated with a DNeasy blood and tissue kit (Qiagen, Germantown, MD). Libraries were prepared with either a native barcoding kit or rapid barcoding kit (Oxford Nanopore Technologies, UK) and sequenced with a GridION instrument using version R10.4 or R9.4.1 flow cells (Oxford Nanopore Technologies), respectively.

Phylogeny among the isolates was inferred by neighbor joining based on core chromosomal SNPs. Dendroscope (51) was used for the graphical representation of phylogenetic trees.

Genome assembly, closure, and annotation

The bioinformatics pipeline used to close genome assemblies and process RNAseq data is shown in Fig. S1. Genomes were assembled from the short- and long-read sequencing data. Initially, FASTQ files were preprocessed using Trimmomatic 0.39 to adapter and quality trim and filter the short-read Illumina reads, and Filtlong 0.2.1 to lengthen and quality filter the long-read data (1 kb minimum length and 90% retention threshold). Subsequently, Unicycler 0.5.0 was used for hybrid assembly of genomes with Spades 3.15.4 (52). Racon 1.5.0 (53) was used for assembly and polishing using the “normal” Unicycler setting (54). Genomes that did not close and circularize with this approach were assembled as described above using Unicycler with the “bold” setting. For genomes that did not assemble using either hybrid approach, a long-read-first approach was taken using Tricycler, with Minimap2 v2.17-r941, Miniasm v0.3-r179, Raven 1.8.1, and Minipolish, as described previously (9). Tricycler assemblies were polished with Polypolish and POLCA (55) and annotated with PROKKA (56) using SDSE isolate *emm* type *stG62647* MGCS36044 as the reference genome (9) (Fig. S1). Artemis (57) was used to browse and manually curate the annotated genomes. Multilocus sequence types were determined using MLST (<https://github.com/tseemann/mlst>) relative to the PubMLST database (<https://pubmlst.org/>), *emm* types were determined using Emmtyper (<https://github.com/MDU-PHL/emmtyper>) relative to the CDC *emm* type database (<https://www2.cdc.gov/vaccines/biotech/strepblast.asp>).

Virulence assessment of *emm*-type *stG62647* isolates in a mouse model of necrotizing myositis

Mouse necrotizing myositis virulence studies were performed as described previously (48, 58–60). All 120 *stG62647* SDSE isolates identified in our comprehensive collection of 499 isolates from French Brittany were included (Table S1). MCGS36089, which is

genetically representative of the *stG62647* isolates recovered in French Brittany and was previously characterized for mouse virulence (9), was used as the virulence reference isolate. Frozen stocks of each isolate were prepared by growth in Todd–Hewitt broth supplemented with 2% w/v yeast extract (Difco Laboratories) and quantified by counting colony-forming units (CFUs) recovered from cultures. For the virulence assays, immunocompetent 3-week-old female CD1 mice weighing 10 g (Envigo Laboratories) were randomly assigned to strain treatment groups. Each mouse was inoculated in the right lower hind limb with 1×10^8 CFUs in 100 μ L ($n = 20$ mice/strain; 2,380 mice total). This inoculation dose was selected based on previous experiments establishing that 1×10^8 CFUs of the reference strain MCGS36089 resulted in 50% near-mortality (9). To validate the initial mouse near-mortality findings, 19 isolates representing the highest and lowest mortality groups were re-tested using the same parameters ($n = 20$ mice/strain; 380 mice total). This strategy identified isolates with significantly increased or decreased virulence. The mouse sample size was determined using a power calculation with the following variables: $\alpha = 0.05$, power $(1-\beta) = 0.8$, difference in survival rates between groups = 0.4, and ratio of group size = 1. After inoculation, mice were monitored at least daily, and near-mortality was determined with internationally recognized criteria (61). Survival data were plotted as Kaplan–Meier curves or percent near-mortality graphs, and statistically significant differences were determined using either the log-rank test or Mann–Whitney test, respectively (Prism10, GraphPad Software).

RNAseq library preparation and sequencing

For RNAseq analysis, 38 *stG62647* isolates exhibiting a low- or high-virulence phenotype in the mouse model of necrotizing myositis were selected (9). Specifically, 19 isolates with near-mortality rates ranging between 20% and 30% were classified as low-virulence isolates, and 19 isolates with near-mortality rates ranging between 70% and 95% were classified as high-virulence isolates. For RNAseq library generation, these 38 isolates were grown *in vitro* in THY broth and analyzed in triplicate at mid-exponential growth phase ($OD_{600} = 1$). Briefly, 2.2 mL from each culture was added to 4.4 mL RNAProtect (Qiagen, Germantown, MD), incubated for 20 min at RT, centrifuged for 15 min at 4,000 rpm, and the pellets were frozen at -80°C . RNA was extracted with an RNeasy kit (Qiagen) according to the manufacturer's instructions. Ribosomal RNA (rRNA) was depleted using the NEBNext rRNA depletion kit (bacteria) (NEB, Beverly, MA), and the ribodepleted RNA was purified with Agencourt RNAClean XP paramagnetic beads (Beckman Coulter, Brea, CA). The quality of the total RNA and rRNA-depleted RNA was evaluated with RNA Nano and Pico chips, respectively, on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). RNA integrity numbers for the 114 RNA samples (38 isolates with three biological replicates) ranged from 9.5 to 10, and rRNA contamination was negligible.

A total of 114 cDNA sequencing libraries were generated from the rRNA-depleted RNA using the NEBNext Ultra II Directional RNA Library Preparation Kit for Illumina (NEB), according to manufacturer's instructions. SPRIselect paramagnetic beads (Beckman Coulter) were used for product purification in all subsequent steps. The cDNA libraries were bar-coded with 48 indexed reverse primers from the NEBNext multiplex oligonucleotides for Illumina (NEB). The quality of cDNA libraries was evaluated with DNA ScreenTape on a TapeStation 2200 (Agilent Technologies). The concentration of each sample was measured fluorometrically with Qubit dsDNA Broad Range and High Sensitivity kits (Thermo Fisher Scientific, Waltham, MA). All sequencing libraries had a cDNA concentration ≥ 100 nM.

Resultant RNAseq cDNA libraries were sequenced with an Illumina NextSeq550 instrument. Fastq files were demultiplexed, sequencing data quality was checked with FastQC (62), and the mean per-base sequence quality was $\geq Q30$ for all libraries. Adapter contamination and read quality filtering were performed with Trimmomatic (63). The RNAseq reads were mapped to the reference CC20 strain MGCS36044 genome (9), which contains a complement of regions of difference (RODs) representative of most isolates in this *stG62647* sample using EDGE-Pro. In some cases, RNAseq reads required individual

mapping to their corresponding isolate genome due to variation in ROD content (Fig. S1). Library replicates produced on average 22 million reads (Fig. S2), corresponding to 100% genomic gene content coverage. To reduce noise and enhance statistical power genes expressed at low levels (that is, with normalized counts <20, and in $\geq 60\%$ of the samples) were excluded from the expression analysis.

Analysis of RNAseq data

Individual genome annotations generated with PROKKA (56) provided the necessary fasta, ptt, and rnt files for read mapping with EDGEpro (64), which in turn generated the raw reads for differential expression (DE) analysis. DE analysis was performed with DESeq2 (65) (Fig. S1) using RStudio. The analysis in RStudio was streamlined with pcaExplorer (66) and ideal (67). Genes were considered to be differentially expressed based on a greater than 1.5-fold change in transcript levels and an adjusted *P* value (*p*_{adj}; Benjamini–Hochberg corrected) of ≤ 0.05 . rRNA and tRNA genes were excluded from the analysis. To compare CC17 and CC20 transcript levels, genes in RODs were also excluded due to the lack of shared RODs between CC17 isolates and the CC20 reference isolate MGCS36044.

Identification of the core genome, accessory genome, and pangenome

The core, accessory, and pangenome gene content for the 38 *stG62647* isolates investigated was determined using Panaroo (68) (Fig. S1). SDSE putative virulence genes were inferred by BLAST search based on homology to a panel of known GAS virulence genes (Table S2).

Polymorphism analysis

Illumina paired-end reads were mapped to the reference isolate MGCS36044 genome with SMALT (<https://www.sanger.ac.uk/tool/smalt/>), and polymorphisms between the aligned reads and the reference genome were identified with FreeBayes (<https://doi.org/10.48550/arXiv.1207.3907>). The statistical significance of the distribution of the polymorphisms between the low- and high-virulence isolate genomes was assessed with Pyseer and by χ analysis.

The lengths of homopolymeric T tracts in the genome of reference strain MGCS36044 were analyzed with the “find base pattern” function in Artemis (57), with an initial range of 20 consecutive T nucleotides, and decreasing to a final range of nine nucleotides (Table S3).

Regions of difference and putative mobile genetic elements

Regions of variably present accessory gene content as defined by Panaroo were used to identify regions of difference (RODs) in consecutive gene content in all isolates. Subsequently, RODs unique to specific isolates or shared by multiple isolates were identified (Table 1). Sites of integration for putative MGEs, genes flanking the RODs identified with Artemis (57); genes flanking the RODs common to most of the isolates are shown in Table S4.

Statistical analysis

The log-rank test was used to determine significant *P*-values for mouse survival data expressed as Kaplan–Meier curves and the Mann–Whitney test for percent near-mortality graphs (Table 1). Single-factor analysis of variance (ANOVA) was used for comparisons of mouse near-mortality (Fig. S3). DESeq2 was used to determine the adjusted *P*-value for differential expression analysis. Pyseer and χ were used to determine the significance of the difference in the distribution of polymorphisms (SNPs and InDels). χ analysis was used to assess the significance of differences in the distribution of the accessory gene content.

TABLE 1 Characteristics of the 38 *stG62647* isolates

Number	Isolate	% near-mortality	CC ^a	Virulence	Genome size in nt.	Number of locus tags	ROD count ^b	Number of genes in RODs	% ROD genes
1	MGCS35681	20	CC20	Low	2,171,406	2,221	15	249	11.2
2	MGCS35679	25	CC20	Low	2,105,313	2,150	13	191	8.9
3	MGCS35983	25	CC20	Low	2,156,048	2,211	14	244	11.0
4	MGCS35868	25	CC20	Low	2,156,576	2,206	14	246	11.2
5	MGCS35970	25	CC20	Low	2,157,490	2,220	14	243	10.9
6	MGCS35893	25	CC20	Low	2,215,594	2,292	15	315	13.7
7	MGCS35977	25	CC17 ^c	Low	2,244,088	2,300	16	318	13.8
8	MGCS36023 ^d	30	CC20	Low	2,156,090	2,206	14	246	11.2
9	MGCS35678	30	CC20	Low	2,156,317	2,203	14	244	11.1
10	MGCS35869	30	CC20	Low	2,158,434	2,208	14	243	11.0
11	MGCS35706	30	CC20	Low	2,160,476	2,217	14	244	11.0
12	MGCS35987	30	CC20	Low	2,167,341	2,227	15	252	11.3
13	MGCS35777	30	CC20	Low	2,172,722	2,223	15	249	11.2
14	MGCS36051	30	CC20	Low	2,202,749	2,246	16	276	12.3
15	MGCS36030	30	CC17	Low	2,203,793	2,230	15	242	10.9
16	MGCS35705	30	CC20	Low	2,203,864	2,297	17	319	13.9
17	MGCS35873	30	CC20	Low	2,205,608	2,258	15	294	13.0
18	MGCS36046	30	CC20	Low	2,219,796	2,311	16	331	14.3
19	MGCS36021 ^e	30	CC17	Low	2,241,044	2,272	15	264	11.6
20	MGCS35941	70	CC20	High	2,155,982	2,202	14	246	11.2
21	MGCS35921	70	CC20	High	2,156,631	2,212	14	244	11.0
22	MGCS36067	70	CC20	High	2,156,854	2,207	14	245	11.1
23	MGCS35844	70	CC20	High	2,157,320	2,205	14	244	11.1
24	MGCS35838	70	CC20	High	2,157,501	2,214	14	244	11.0
25	MGCS35904	70	CC20	High	2,175,679	2,221	15	241	10.9
26	MGCS36035	70	CC20	High	2,199,191	2,276	15	314	13.8
27	MGCS36086	70	CC20	High	2,210,996	2,264	15	281	12.4
28	MGCS35730	75	CC20	High	2,109,976	2,169	13	194	8.9
29	MGCS36020	75	CC20	High	2,177,607	2,236	15	262	11.7
30	MGCS36061	75	CC20	High	2,195,165	2,239	15	282	12.6
31	MGCS36071 ^f	80	CC20	High	2,156,154	2,206	14	247	11.2
32	MGCS35939	80	CC20	High	2,158,828	2,209	14	247	11.2
33	MGCS35957	80	CC20	High	2,233,554	2,326	17	327	14.1
34	MGCS35798	85	CC20	High	2,157,408	2,203	14	244	11.1
35	MGCS35978	85	CC20	High	2,159,390	2,214	14	245	11.1
36	MGCS35725	85	CC20	High	2,160,452	2,215	14	244	11.0
37	MGCS35785	90	CC20	High	2,157,613	2,207	14	244	11.1
38	MGCS35823	95	CC20	High	2,189,545	2,269	15	297	13.1
	Min.	20			2,105,313	2,150	13	191	8.9
	Max.	95			2,244,088	2,326	17	331	14.3

^aCC, clonal complex.^bROD, region of difference.^cThe CC17 isolates are separated from the CC20s by approximately 10,700 core SNPs.^dMGCS36023 is one locus variant from CC20.^eMGCS36021 is one locus variant from CC17.^fMGCS36071 is one locus variant from CC20.

RESULTS

Genome sequence data

The study is based on a comprehensive analysis of all 499 SDSE isolates recovered from humans in French Brittany between 2010 and 2018. Illumina short-read sequencing found that *stG62647* was the most prevalent *emm* type ($n = 120$; 24%). These 120 isolates

included organisms from invasive infections ($n = 67$), probably invasive infections ($n = 11$), non-invasive infections ($n = 24$), and carriers ($n = 18$). *Emm* type *stG62647* isolates have been associated with severe invasive infections and have recently increased in frequency across multiple countries (28, 42, 69, 70). Given their abundance in this French Brittany sample, their emergent nature, medical importance, and the relative lack of fundamental pathobiology data, we focused our study on the *stG62647* isolates.

To gain a fuller understanding of the population genomics of these *stG62647* isolates, we sequenced the genomes to closure by combining Illumina paired-end short-read and Oxford Nanopore (ONT) long-read data (Fig. S1). The vast majority of the *stG62647* isolates were multilocus sequence type clonal complex 20 (CC20; $n = 116$) and were closely related genetically, differing from each other by an average of only 240 core-genome SNPs (Fig. 1; Table S1). These data indicated that the great majority of French Brittany CC20 *stG62647* isolates are clonally related, descending from a relatively recent common ancestor. In contrast, three *stG62647* isolates were identified as clonal complex CC17, and one was *st128*, differing from the 116 CC20 isolates on average by 10,700 and 13,500 core-genome SNPs, respectively (Fig. 1).

***stG62647* isolates differ widely in virulence in a mouse model of necrotizing myositis**

Despite most of the isolates differing modestly genetically, the 120 *emm*-type *stG62647* isolates had a wide range of virulence in a well-established mouse model of necrotizing myositis (48, 58–60) (Fig. 2; Table S1). Twenty mice each were inoculated in the lower hind limb with 1×10^8 CFUs of each isolate. Mouse virulence varied considerably strain-to-strain, with near-mortality ranging from 20% to 95% (Fig. 2). As a group, the 19 isolates with highest near-mortality were significantly more virulent than the 19 isolates with lowest near-mortality ($P < 0.0001$, log-rank test). Among these 38 *stG62647* isolates, 35 were CC20, and three were CC17. Because of the unexpectedly wide range of near-mortality, we repeated the virulence assessment with 12 high-virulence isolates and seven low-virulence isolates, chosen from the 19 isolates in each corresponding group. The repeat near-mortality data confirmed the initial wide-range of mouse virulence.

The three CC17 *stG62647* isolates reproducibly had significantly lower mouse near-mortality compared with the 19 high-mortality CC20 isolates (Fig. S3).

Analysis of *stG62647* closed-genome data

To identify potential molecular mechanisms underlying the substantial differences in mouse virulence, we analyzed the closed-genome data for the 38 *stG62647* isolates in the low- and high-virulence groups. Genomic and virulence characteristics for these 38 isolates are shown in Table 1. Their genomes varied in size from 2.16 to 2.24 Mbp and contained between 2,150 and 2,326 inferred genes. The variation in genome size and overall gene content was primarily due to discrete regions of difference (RODs), predominantly attributable to the integration of putative mobile genetic elements (MGEs). Among the 2,916 genes identified in the pangenome, 1,751 were present in all isolates and constitute the core genome, and 1,165 were variably present and constitute the accessory genome (Fig. S1). Importantly, in comparing the gene content of the low- and high-virulence isolates, no gene was found to be significantly non-randomly associated with either phenotypic group.

In total, 43 different RODs were identified across the 38 isolates, with each ROD containing between 4 and 51 inferred genes. Some RODs had gene content characteristic of integrative conjugative elements (ICEs), such as integrase/recombinase genes, or bacteriophage genes as described previously (9, 71). The core genes flanking the integrase genes in the RODs, including tRNA genes, *traG*, *rpmG*, and *rpsL*, were consistent with known streptococcal target genes for ICE insertion (71). Notably, several genes encoding small non-coding RNAs were also identified as putative ICE integration targets, an observation that may warrant future investigation (Table S4). In addition to putative ICE and bacteriophage genes, other ROD genes had the capacity to encode inferred

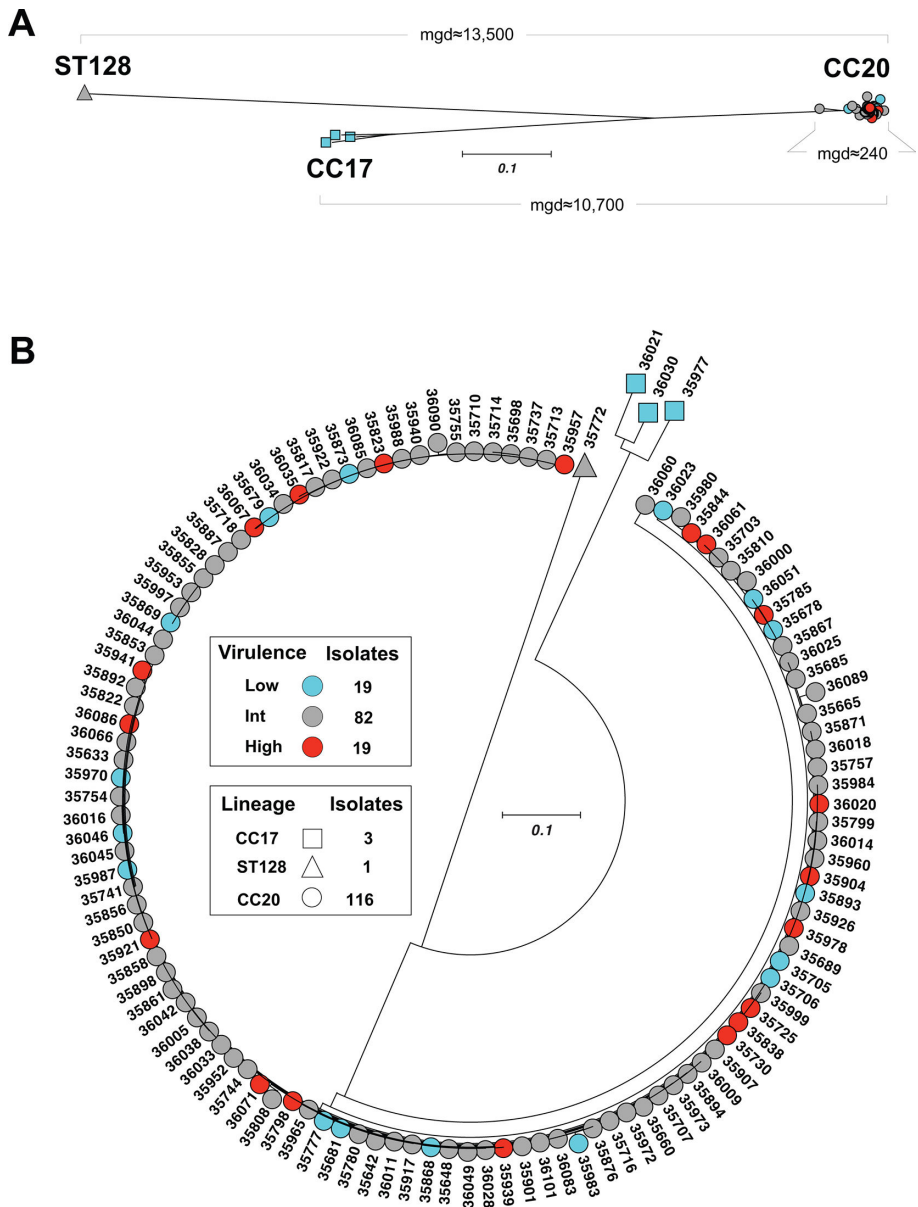


FIG 1 Genetic relationships among 120 SDSE *stG62647* isolates. The genetic relationships were inferred by neighbor joining based on 25,790 core SNPs defined relative to the reference isolate MGCS36044. (A) Three different genetic groups were identified. The 116 CC20 isolates clustered together and were separated, on average, by approximately 240 core-genome SNPs. The three CC17 isolates were separated, on average, by approximately 3,100 core-genome SNPs, and as a group were separated from the CC20s by approximately 10,700 core-genome SNPs. The ST128 isolate was separated from the CC20s by approximately 13,500 core-genome SNPs. mgd, mean genetic distance (in SNPs). (B) Radial phylogram depicting the 120 SDSE *stG62647* isolates. CC20 isolates ($n = 116$) are represented by circles, CC17s ($n = 3$) by squares, and the ST128 ($n = 1$) by a triangle. The virulence phenotype is shown in colors, as high virulence (red; $n = 19$), low virulence (blue; $n = 19$), and intermediate virulence (int, gray; $n = 82$).

transporters, transcriptional regulators, methyltransferases, toxin-antitoxin systems, and bacteriocins.

The number of RODs per isolate ranged from 13 to 17, comprising between 191 to 331 genes, accounting for 8.9% to 14.3% of each genome (Table 1). Isolates with the largest genomes had more RODs and corresponding accessory genes (Table 1). Consistent with the genomic differentiation between CC17 and CC20 isolates, only one ROD

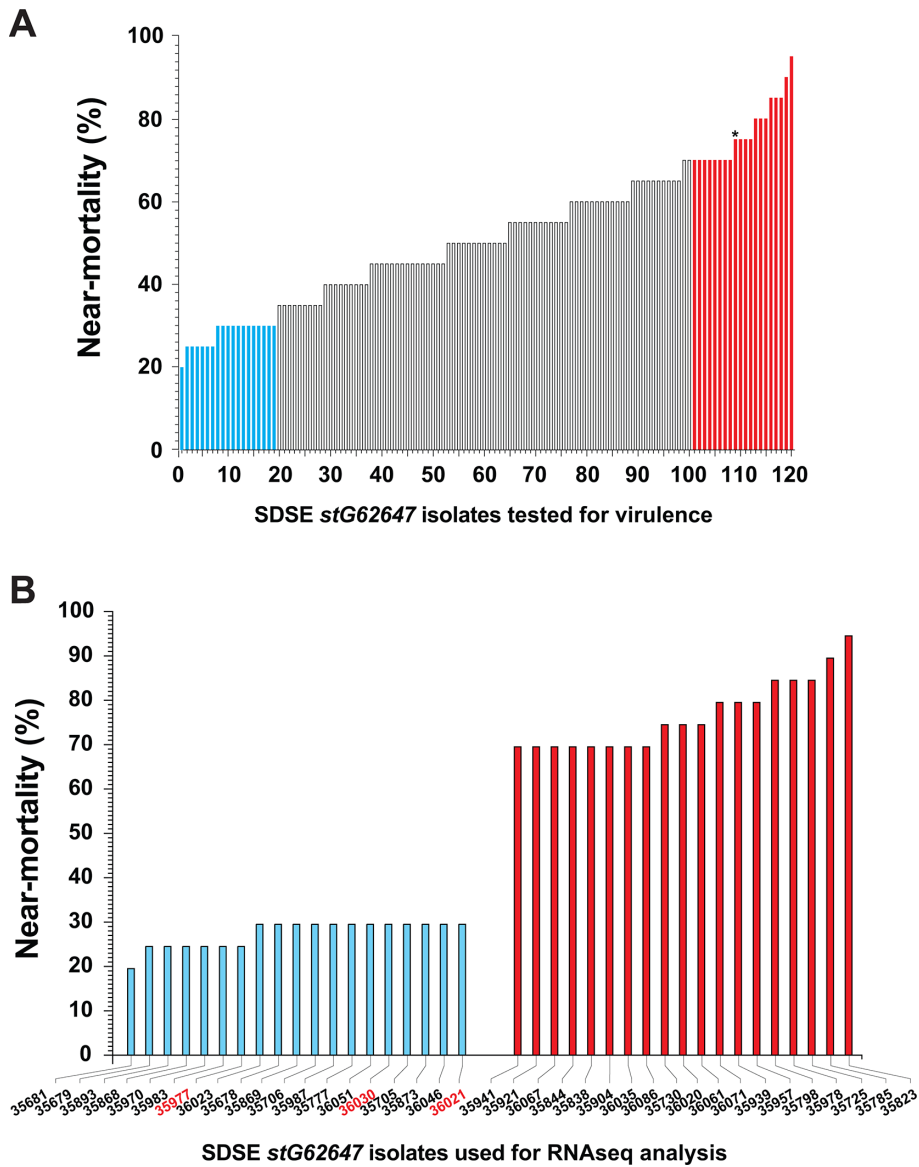


FIG 2 Mouse virulence of SDSE *stG62647* isolates. The 120 *emm*-type *stG62647* isolates from the comprehensive French Brittany collection of SDSE clinical isolates ($n = 499$) were inoculated into mice in the right lower hind limb (dose = 1×10^8 CFUs; $n = 20$ mice/strain; 2,400 total mice). (A) Percent near-mortality for the 120 isolates. The isolates were separated into three virulence groups: (i) low virulence (blue bars; percent near-mortality [PM] = 20%–30%), (ii) isolates with intermediate virulence (white bars; PM = 35%–70%), and (iii) high-virulence strains (red bars; PM = 70%–95%). *, MGCS36044 was not assayed for RNAseq analysis. (B) Percent near-mortality values for isolates selected for RNAseq analysis. Thirty-five isolates were CC20 (black font), and three were CC17 (red font).

(36044 ROD.7 containing 50 genes on average) was shared between these two multilocus sequence-type groups. ROD.7 was present in one CC17 and two CC20 isolates. Twenty-five RODs were identified among the 35 CC20 isolates that were not present in the CC17 isolates. Fourteen had genes present in all 35 isolates, while the remaining were shared by only one to four isolates. Of the 18 RODs in the three CC17 isolates, 14 were present in all three isolates (Table S4). No individual gene was found significantly non-randomly distributed between the low- and high-virulence groups of isolates.

RNAseq transcriptome analysis comparing low- and high-virulence SDSE isolates

To gain insight into how altered gene expression may contribute to differences in mouse virulence among *stG62647* isolates, we conducted RNAseq analysis on the 38 isolates with the lowest ($n = 19$) and highest ($n = 19$) mouse near-mortality groups. SDSE transcriptome data has only been done for a few isolates, largely focused on the two-component regulator CsrRS (aka CovRS). A pangenome-spotted microarray analysis was done for one strain and its *csrS* mutant derivative (11), and a genome-wide RNAseq analysis was done on one non-invasive human isolate and its *csrS* gene deletion mutant derivative (10). Importantly, no transcriptome studies have yet been conducted with SDSE *stG62647* isolates. Thus, this analysis provides much needed information on the transcriptome characteristics of these medically important emergent isolates.

The 38 SDSE isolates were grown in triplicate in THY broth and harvested at the mid-exponential growth phase. We used THY broth because (i) it has been previously used successfully in extensive transcriptomic studies of the genetically related pathogen GAS (48), and (ii) currently, there are no media that mimic *in vivo* growth in mammalian hosts.

To explore commonality in variation patterns in the RNAseq expression data set, principal component analysis (PCA) was conducted for these 38 isolates (38 isolates \times 3 = 114 biological replicates) (Fig. 3A). First, for nearly all of the isolates, the three biological replicates were located in close proximity, indicating good within-isolate reproducibility. Second, the three CC17 isolates were well separated from the 35 CC20 isolates along principal component 1 (PC1). Among the 35 CC20 isolates two less distantly separated groups were found along PC2 (Fig. 3A). Importantly, these PCA parsings did not significantly correlate with virulence phenotype, as the CC20 low- and high-virulence isolates were interspersed with each other in the separate clusters.

Comparison of the transcriptome data from the CC17 and CC20 isolates identified 380 differentially expressed (DE) genes (at a fold-change [FC] ≥ 1.5 and an adjusted $P \leq 0.05$), with 236 genes upregulated in the CC20 isolates, and 144 genes upregulated in the CC17 isolates. These DE genes included several putative virulence genes (Table S5). Despite the high reproducibility of the isolates in the RNAseq and mouse virulence data, the PCA clustering shows that there is no simple pattern of variance in gene expression shared by the *stG62647* isolates that correlate with differences in virulence phenotype.

CC20 isolates separate into two distinct clusters by PCA

To further investigate why the CC20 isolates separate into two clusters differing in expression variance patterns (Fig. 3A), the analysis was repeated, omitting the three CC17 isolates. PCA parsed the 35 CC20 isolates (35 isolates \times 3 = 105 biological replicates) into two well-resolved clusters along PC1 (Fig. 3B, C1 = 22 and C2 = 13). Again, when comparing the *stG62647* CC20 low- ($n = 16$) and high-virulence ($n = 19$) isolates as groups, the PCA clustering did not correlate with the virulence phenotype. Consistent with this lack of correlation, only 53 of 1,751 (3.0%) core genes were significantly DE between the virulence groups. Moreover, these 53 DE genes lacked any putative virulence factors. We interpret this lack of correlation to indicate that either (i) multiple different molecular pathways enable the *stG62647* CC20 isolates to acquire an altered virulence phenotype, or (ii) the *in vitro* growth conditions in THY broth do not adequately replicate the *in vivo* mouse necrotizing myositis growth environment. We note it is possible that *in vitro* growth to a different logarithmic phase may yield additional relevant data, but more extensive studies are required to address this idea.

In contrast to the low number of DE genes identified when comparing the low- and high-virulence groups, when comparing the two *stG62647* CC20 PCA clusters, 409 genes were found significantly DE. Notably between C1 and C2, the most highly DE genes included five contiguous genes in the second of two fibronectin-collagen-T antigen (FCT-2) regions present in the SDSE genome (Fig. 4). These pilus-encoding genes had

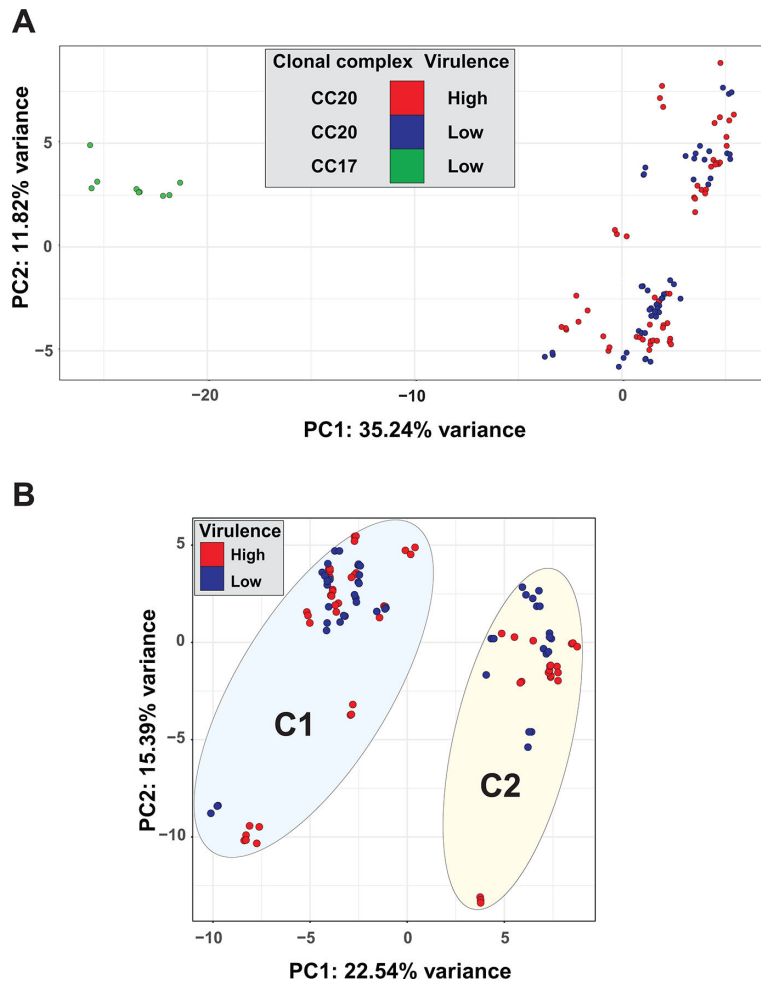


FIG 3 Transcriptome analysis of SDSE *stG62647* isolates grown *in vitro*: principal component analyses. SDSE *stG62647* isolates were grown *in vitro* in triplicate (114 replicates) to the mid-exponential phase. Thirty-five isolates were CC20, and three were CC17. (A) The CC17 replicates (green circles) clustered together and grouped distinctly apart from the CC20 replicates along principal component 1 (PC1). The CC20 high- (red circles) and low-virulence (blue circles) replicates were similarly interspersed along two somewhat distinct clusters along principal component 2 (PC2). (B) Upon exclusion of the three CC17 isolates, the CC20 replicates corresponding to 35 isolates were assigned to two distinct clusters designated C1 ($n = 22$ strains) and C2 ($n = 13$ strains), demarcated with ovals. The low- and high-virulence replicates were interspersed with each other.

higher transcript levels in cluster 1 (C1 = high pilus expression) compared with cluster 2 (C2 = low pilus expression). This difference in expression was uniform, in that all five FCT-2 region genes were higher in expression in all 22 C1 isolates and lower in expression in all 13 C2 isolates. This observation, together with the molecular architecture of this region, suggests that these five contiguous genes comprise an operon (Fig. 4). An inferred stand-alone transcriptional regulator gene encoding an RofA-like protein (RALP) is divergently transcribed from these five genes analogous to pilus regulons in GAS (72, 73).

Analysis of the intergenic region between the divergently transcribed pilus genes and the RALP regulator gene identified a homopolymeric nucleotide tract (HPNT) of 14 Ts in the reference strain MGCS36044 genome. This HPNT is one of only two such HPNTs in the MGCS36044 genome that is longer than 10 Ts in length (Table S3), suggesting that this is rare and unlikely represents a random occurrence. The length of this HPNT varied from 9 to 19 Ts among the 35 CC20 *stG62647* isolates (Table 2). Importantly, all

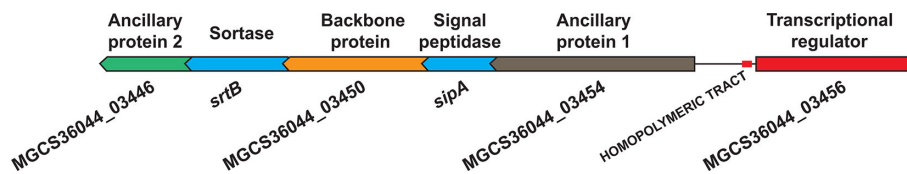


FIG 4 Organization of FCT-2 region genes. Genes and corresponding gene products in the FCT-2 region of the reference SDSE strain MGCS36044. From left to right: (i) MGCS36044_03446, pilus ancillary protein 2 (AP2); (ii) *srtB*, pilus sortase; (iii) MGCS36044_03450, pilus backbone protein (BP); (iv) *sipA*, signal peptidase; (v), MGCS36044_03454, pilus ancillary protein 1 (AP1); and (vi) MGCS36044_03456, putative transcriptional regulator. The location of the homopolymeric T-tract in the intergenic region is shown.

13 CC20 C2 isolates with high transcript levels of the five DE pilus genes had an HPNT of either 13 or 14 Ts, whereas 21 of 22 CC20 C1 isolates with low transcript levels of the pilus genes had an HPNT either shorter or longer than 13–14 Ts in length. The exception was observed in isolate MGCS35970, where the putative RALP transcriptional regulator gene was disrupted due to a frame-shift mutation caused by the insertion of one T nucleotide in an intragenic HPNT (Table 2). This observation is consistent with this putative transcriptional regulator activating the expression of the five contiguous pilus genes.

RNAseq transcriptome analysis comparing individual strains within low- or high-virulence groups

To explore the hypothesis that the differences in virulence observed among the *stG62647* isolates might be multifactorial in nature, involving multiple different regulatory pathways, such that individual isolates achieve altered virulence following different routes, we reanalyzed the expression data using a one-against-all comparison approach. In this approach, we maintain an association with the virulence phenotype by assessing for gene content that an individual isolate of one virulence group differentially expresses relative to all isolates of the opposing virulence group. Therefore, we compared all 16 of the low-virulence CC20 isolates individually against all 19 high-virulence CC20 isolates as a group, and vice versa (*i.e.*, 35 one-against-all comparisons).

Using a stringent cutoff for differential gene expression ($p_{adj} \leq 0.05$, and $FC \geq 1.5$ -fold), the number of DE genes varied significantly among isolates (Fig. 5; Tables S6 and S7). For example, comparisons of individual low-virulence isolates to the grouped high-virulence isolates revealed DE genes ranging from 1 to 167, while the converse comparison yielded 1 to 216 DE genes. Importantly, regardless of the comparison, there were no common genes or biochemical pathways (as assessed by use of FUNAGE-Pro) (74) whose genes were differentially expressed across all comparisons. These data provide additional evidence for the existence of multiple molecular pathways for SDSE *stG62647* isolates to achieve a distinct mouse virulence phenotype. Possible mechanisms might include (i) polymorphisms in key genes, either in major transcriptional regulators (MTRs), such as those seen with *covS* in GAS (48, 75), or in genes encoding a virulence-related protein (60, 76), and/or (ii) through gene content differences in the accessory genome, which might not be apparent when sequencing reads from all isolates are mapped to a single reference isolate MGCS36044. Thus, we next investigated these possibilities.

Polymorphism analysis

To test the hypothesis that nonsynonymous polymorphisms (nsSNPs) in MTR genes and putative virulence genes (Table S2) contribute to differences in mouse virulence among the SDSE *stG62647* isolates, we identified nsSNPs and Indels in the CC20 isolates ($n = 35$) relative to the genome of the CC20 reference isolate MGCS36044 (Table S8). Although nsSNPs were identified in genes known to regulate virulence in GAS (*e.g.*, *covRS* TCS,

TABLE 2 The length of the homopolymeric tract determines FCT-2 pilus gene expression

	Isolate	Length	Homopolymeric tract ^a	Virulence	Cluster ^b	Pilus gene expression ^c
1	MGCS35730	9	TTG TTTTTTTT CTTTACGA	High	C1	Low
2	MGCS35844	9	TTG TTTTTTTT CTTTACGA	High	C1	Low
3	MGCS35904	9	TTG TTTTTTTT CTTTACGA	High	C1	Low
4	MGCS35798	11	TTG TTTTTTTTT CTTTACGA	High	C1	Low
5	MGCS35838	11	TTG TTTTTTTTT CTTTACGA	High	C1	Low
6	MGCS35777	12	TTG TTTTTTTTTT CTTTACGA	Low	C1	Low
7	MGCS35939	12	TTG TTTTTTTTTT CTTTACGA	High	C1	Low
8	MGCS35987	12	TTG TTTTTTTTTT CTTTACGA	Low	C1	Low
9	MGCS36020	12	TTG TTTTTTTTTT CTTTACGA	High	C1	Low
10	MGCS35970	13 ^d	TTG TTTTTTTTTTT CTTTACGA	Low	C1	Low
11	MGCS35681	15	TTG TTTTTTTTTTTT CTTTACGA	Low	C1	Low
12	MGCS35725	15	TTG TTTTTTTTTTTT CTTTACGA	High	C1	Low
13	MGCS35893	15	TTG TTTTTTTTTTTT CTTTACGA	Low	C1	Low
14	MGCS35957	15	TTG TTTTTTTTTTTT CTTTACGA	High	C1	Low
15	MGCS35978	15	TTG TTTTTTTTTTTT CTTTACGA	High	C1	Low
16	MGCS36023	15	TTG TTTTTTTTTTTT CTTTACGA	Low	C1	Low
17	MGCS36051	15	TTG TTTTTTTTTTTT CTTTACGA	Low	C1	Low
18	MGCS36071	15	TTG TTTTTTTTTTTT CTTTACGA	High	C1	Low
19	MGCS35706	16	TTG TTTTTTTTTTTTT CTTTACGA	Low	C1	Low
20	MGCS35983	16	TTG TTTTTTTTTTTTT CTTTACGA	Low	C1	Low
21	MGCS36086	16	TTG TTTTTTTTTTTTT CTTTACGA	High	C1	Low
22	MGCS35678	19	TTG TTTTTTTTTTTTTTT ACGA	Low	C1	Low
1	MGCS35705	13	TTG TTTTTTTTTTTT CTTTACGA	Low	C2	High
2	MGCS35873	13	TTG TTTTTTTTTTTT CTTTACGA	Low	C2	High
3	MGCS35921	13	TTG TTTTTTTTTTTT CTTTACGA	High	C2	High
4	MGCS35679	14	TTG TTTTTTTTTTTTT CTTTACGA	Low	C2	High
5	MGCS35785	14	TTG TTTTTTTTTTTTT CTTTACGA	High	C2	High
6	MGCS35823	14	TTG TTTTTTTTTTTTT CTTTACGA	High	C2	High
7	MGCS35868	14	TTG TTTTTTTTTTTTT CTTTACGA	Low	C2	High
8	MGCS35869	14	TTG TTTTTTTTTTTTT CTTTACGA	Low	C2	High
9	MGCS35941	14	TTG TTTTTTTTTTTTT CTTTACGA	High	C2	High
10	MGCS36035	14	TTG TTTTTTTTTTTTT CTTTACGA	High	C2	High
11	MGCS36046	14	TTG TTTTTTTTTTTTT CTTTACGA	Low	C2	High
12	MGCS36061	14	TTG TTTTTTTTTTTTT CTTTACGA	High	C2	High
13	MGCS36067	14	TTG TTTTTTTTTTTTT CTTTACGA	High	C2	High

^aHomopolymeric tract in bold and flanking common nucleotides in all isolates.

^bCluster refers to the two clusters of CC20 strains in the PCA plot in Fig. 3B.

^cExpression of the five consecutive pilus genes in the FCT-2 region.

^dDisrupted transcriptional regulator gene.

mga, and *fas* operon), nsSNPs in these genes were relatively uncommon, and there was no consistent association between nsSNPs or Indels and virulence phenotype in these isolates. Similarly, genome-wide analyses with Pyseer (77) also did not identify polymorphisms in genes significantly associated with differences in mouse virulence categories in this isolate sample.

RNAseq transcriptome analysis of the RODs

We next tested the hypothesis that gene content differences in RODs contribute to mouse virulence phenotype. No ROD was uniquely present in either the low- or high-virulence groups and absent among isolates in the alternative virulence phenotypic category. The same was true for individual ROD genes. These results indicate that differences in ROD gene content between isolates are not responsible for the observed differences in mouse virulence.

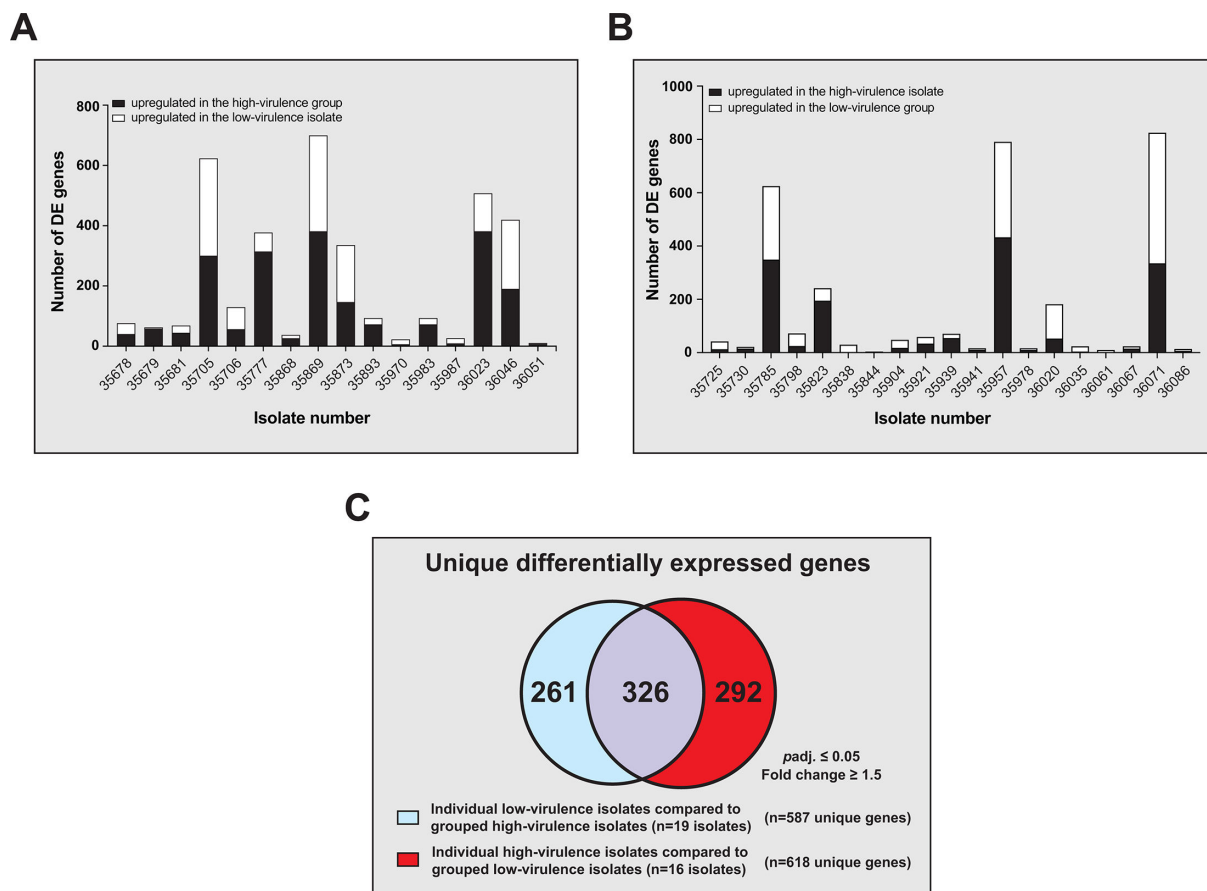


FIG 5 Summary of transcriptome analysis of SDSE *stG62647* CC20 isolates. SDSE *stG62647* isolates were grown in triplicate *in vitro* and collected at mid-exponential growth phase. The graphs represent the total number of differentially expressed genes in each comparison, either upregulated in the high-virulence isolates (black bars), or in the low-virulence isolates (white bars). The fold-change value cutoff and adjusted *P*-value cutoff was 1.5 and ≤0.05, respectively. (A) Sixteen comparisons for individual low-virulence isolates against grouped high-virulence isolates. (B) Nineteen comparisons for individual high-virulence isolates against grouped low-virulence isolates. (C) Venn diagram showing the number of unique differentially expressed genes, either for the individual low-virulence isolates compared with grouped high-virulence isolates (blue), or for the reciprocal set of comparisons (red). Genes present in both comparisons are shaded in purple.

We next analyzed differential gene expression in the RODs by comparing the 14 RODs shared by most CC20 isolates (encompassing a total of 244 genes), as two groups comprising low- or high-virulence isolates. The mean reads per kilobase of transcript per million (RPKM) values were calculated for all shared genes in a ROD, and fold-differences were calculated from the ratios of mean RPKM high virulence/mean RPKM low virulence. Nearly all of the 244 genes had mean RPKM ratios close to one when comparing the high- and low-virulence isolates. This lack of differential expression between the high- and low-virulence groups further reinforces that there are likely multiple molecular pathways responsible for altered virulence in this model.

DISCUSSION

SDSE isolates of *emm*-type *stG62647* have recently emerged as a prominent cause of invasive human infections in multiple countries (9, 18, 28, 42–47). We discovered that *emm*-type *stG62647* isolates are the most common cause of SDSE infections in French Brittany, with a notable increase in occurrence over the 8-year study period (Fig. S4). These observations, coupled with the limited understanding of the molecular mechanisms contributing to SDSE pathogenesis, motivated this integrative study of the genomics, virulence, and transcriptomics of *stG62647* isolates.

We sequenced to closure the genomes of 120 *stG62647* isolates, including 116 CC20 and 3 CC17 organisms, thereby substantively expanding existing population genomics data for SDSE. These CC20 organisms are closely related, differing on average by only 240 core-genome SNPs (Fig. 1). The data indicate a relatively recent divergence from a common ancestor, although the exact timing is not known. Complete comparative genomic analysis of the 120 *stG62647* isolates was beyond the scope of this study but will be presented in a subsequent manuscript.

The genome sizes of the 38 SDSE isolates we analyzed were approximately 300–400 kbp larger on average than those of the closely related GAS genome. This increase is mainly due to the presence of discrete RODs in SDSE, some of which have gene content consistent with ICEs. The ROD content in these 38 *stG62647* isolates varied, but many RODs were shared among the majority of either CC20 or CC17 isolates.

In principle, these accessory genes could confer a survival advantage by expanding the coding capacity of the SDSE genomes. In this regard, we note that the 43 RODs included genes encoding inferred transporters, transcriptional regulators, methyltransferases, toxin–antitoxin systems, bacteriocins, and many hypothetical proteins. Additionally, as inferred from gene annotations, some of the inferred gene products have the potential to result in additional metabolic pathways or confer resistance to heavy metals, as previously suggested (78, 79).

Despite the close genomic relationships among the *stG62647* CC20 isolates in this SDSE sample, there was a strikingly wide range of virulence observed in a well-established mouse model of necrotizing myositis, a model that replicates many aspects of human disease caused by pathogenic beta-hemolytic streptococci (9, 48, 58, 59). We hypothesized the differences in virulence were caused, in part, by differences in gene expression, a phenomenon extensively documented in GAS (48, 60, 80). The *stG62647* transcriptome data revealed no simple relationship between mouse virulence and the overall pattern of *in vitro* gene expression. However, we did identify several potential molecular contributors to altered mouse virulence.

The 326 genes differentially expressed in both comparisons included several inferred putative virulence genes. Notable among these are genes important for SDSE virulence, such as streptolysin S region-encoding genes (10, 14), and *fasA*, which encodes the TCS response regulator FasA (14), albeit in different SDSE *emm* types. Additional differentially expressed putative virulence genes included the adhesin genes *fbp* and *fbpB*, encoding fibronectin-binding proteins; the transcriptional regulators *rofA* and *mga*; *emm*, which encodes the M protein; *slr*, which encodes an InIA-like protein; and *shr*, which encodes a heme-binding protein (Table S7). In addition, 38% of these common 326 differentially expressed genes are involved in carbohydrate and amino acid metabolism, consistent with a report that carbohydrate metabolism genes are differentially expressed during SDSE infection in an intraperitoneal mouse model (11).

Three additional SDSE genes or gene groups have been suggested or shown to affect SDSE virulence, including (i) disruption of *silB* in SDSE *stG62647* isolates (28); (ii) deletion of one of the three TTAAAGA heptanucleotide tandem direct repeats within *srrG* in SDSE *stG6792* isolates (10); and (iii) inactivation of *covRS* (*csrRS*), also in SDSE *stG6792* isolates (10, 11). These changes are either known or postulated to enhance SDSE virulence. The results show that a virulence phenotype can be influenced *via* mutational polymorphisms in SDSE, a phenomenon well-documented in GAS (48, 60, 75, 76).

Consistent with the hypothesis that multiple molecular pathways contribute to altered mouse virulence phenotype, our polymorphism analysis confirmed that there was no single polymorphism present in all isolates either in the low- or high-virulence group and absent in the alternative group. Given the numerous polymorphisms identified among the 35 CC20 isolates (CC17 isolates were more genetically distant), we concentrated on two groups of genes most likely to affect virulence: (i) genes encoding putative virulence factors, and (ii) genes encoding major transcriptional regulators (MTRs). Many of the gene products identified have not been directly studied in SDSE; however, their involvement in virulence has been inferred from similar studies conducted

in GAS (10, 11, 14, 16, 18–22, 81). We chose this approach based on DNA sequence similarity between GAS and SDSE and the overlap in human disease manifestations observed between GAS and SDSE.

Importantly, compared with other MTR genes involved in virulence, *covS* ($n = 6$), *covR* ($n = 2$), *fas* ($n = 5$), and *mga* ($n = 3$), exhibited the largest number of polymorphisms in this isolate cohort, which is consistent with previous reports implicating CovRS, Fas, and Mga in GAS virulence (48, 75, 82, 83). Interestingly, in cases where nonsynonymous polymorphisms occurred in the same gene (e.g., *covR* or *fas*) in both low- and high-virulence isolates, the SNPs were located at different gene positions in each virulence group. Previous studies on different SDSE *emm* types have reported *covRS* polymorphisms and the effect of this TCS on virulence (10, 11, 14), but the SNPs we identified in *stG62647* differ from those reports.

When individual isolates were compared with an isolate (MGCS35679) lacking MTR or virulence gene polymorphisms, isolates harboring specific mutations, including CovS V52G, Mga E309K, and RofA L202F, showed differences in gene expression. A substantial number of the differentially expressed genes identified in these comparisons have been previously identified as members of the CovRS, Mga, and RofA regulons in GAS, respectively (83–85). We hypothesize that additional polymorphisms found in this SDSE cohort could similarly influence virulence phenotype, although confirmation through creation and analysis of isogenic strains is required to definitively determine these effects.

Genetic variation in the *sil* operon and *srrG* gene has also been implicated in virulence (10, 28). Oppegaard et al. (28) reported that 18 of 19 *stG62647* isolates causing invasive infections in Norway had a *silB* gene disrupted by insertion of *IS1548*. This was an important observation as *silB* encodes a sensory kinase that has been reported to increase virulence when inactivated (86, 87). The investigators speculated that inactivation of *silB* by insertion of *IS1548* contributed to enhanced virulence of *stG62647* CC20 isolates (28). The presence of this *IS1548* insertion at the same location in *silB* in the great majority of the French Brittany *stG62647* CC20 isolates (Table S1) is consistent with genome sequence data, indicating that the *stG62647* CC20 isolates had a relatively recent clonal origin. Conversely, the three *stG62647* CC17 isolates lacked the *sil* locus, analogous to data reported previously (28). The *sil* locus transcript abundance values were essentially identical in both virulence groups, indicating that there was not an association between expression level and differences in virulence phenotype in this animal model.

For the *srrG* gene, all 35 low- and high-virulence CC20 *stG62647* isolates had deletion of one of the three TTAAAGA tandem repeats, which Ikebe et al. reported enhances *sagA* gene expression and increases mouse virulence (10). This mutation might contribute to the increased virulence observed in CC20 *stG62647* compared with other SDSE *emm* types, although it does not fully account for the low- or high-virulence phenotypes observed here. Of the three CC17 *stG62647* isolates, two contained three TTAAAGA tandem repeats, and one had only two tandem repeats. Taken together, these data are consistent with there being multiple pathways for SDSE *stG62647* isolates to modulate virulence phenotype.

Excluding the three CC17 *stG62647* isolates from the transcriptome analysis allowed the CC20 *stG62647* isolates to cluster into two discrete groups (Fig. 3B). However, there was not a simple association between variance in gene expression and mouse virulence phenotype. The variance was driven mainly by differences in the transcript levels of five genes in the pilus FCT-2 region (32- to 35-fold difference between the two groups). We found a strong association between the length of a homopolymeric tract of T residues in the presumed regulatory region, located between the divergently transcribed five-gene pilus operon (Table 2), and a gene encoding a RALP-like protein typically found in streptococcal FCT regions (88). Homopolymeric tracts are highly variable in length (48, 76), and in this cohort of 35 CC20 35 isolates, they ranged from 9 to 19T residues. With a single exception, lengths of 13 and 14T residues were associated with high

transcript levels for the five pilus genes; any other length resulted in lower expression. In isolate MGCS35970, which had low transcript levels of the FCT-2 pilus genes despite having a 13T length, the gene encoding the divergently transcribed putative RALP-like transcriptional regulator had a single nucleotide (T) deletion at position 346, a polymorphism resulting in an out of frame mutation in the coding region of the gene. These data are consistent with the role of this putative RALP as an activator of FCT-2 pilus gene expression. The lack of a uniform association between pilus gene expression and enhanced virulence mirrors observations in GAS studies, where increased pilus gene expression results in decreased virulence in mouse models of invasive infection (89, 90). This suggests that in SDSE, as in GAS, pilus expression may enhance colonization and carriage but paradoxically decrease virulence in invasive infections. However, further investigation using isogenic mutant strains will be required to test this idea.

To summarize, we integrated full genome sequencing, mouse virulence, and *in vitro* transcriptome data from a cohort of 38 human SDSE isolates of the *stG62647 emm* type from French Brittany. Our comprehensive analysis leads us to conclude that multiple molecular pathways contribute to the diverse virulence phenotypes observed in this emerging group of organisms, which is now responsible for a significant number of infections across several geographic areas. In addition, the data suggest that human genetics and underlying medical conditions contribute to the range of disease severity. More work is required to further define specific molecular factors that contribute to host–pathogen interactions in SDSE organisms. Given that the isolates are from a relatively limited geographic region (Brittany, France), analysis of *stG62647* SDSE from more widely dispersed areas is warranted and might contribute information to the generalizability of our findings.

ACKNOWLEDGMENTS

These studies were funded in part by the Fondren Foundation.

We thank Alma Amaya, Akanksha Batajoo, Nicole Kanellopoulos, Shelby Kvinta, Regan Mangham, Eleanor Nichols, Matthew Ojeda Saavedra, Jordan Pachuca, Sindy Pena, and Kristina Reppond for expert technical assistance and Drs. Sasha Pejerrey and Heather McConnell for editorial assistance.

AUTHOR AFFILIATIONS

¹Laboratory for Molecular and Translational Human Infectious Diseases Research, Center for Infectious Diseases, Houston Methodist Research Institute, Houston, Texas, USA

²Department of Pathology and Genomic Medicine, Houston Methodist Hospital, Houston, Texas, USA

³Department of Pathology and Laboratory Medicine, Weill Medical College of Cornell University, New York, New York, USA

⁴Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, New York, USA

⁵CHU de Rennes, Service de Bactériologie-Hygiène Hospitalière, Rennes, France

⁶INSERM, CIC 1414, Rennes, France

⁷Université Rennes 1, Faculté de Médecine, Rennes, France

⁸Université Rennes 1, Faculté de Pharmacie, Rennes, France

⁹OSS-Oncogenesis, Stress, and Signaling, INSERM 1242, Rennes, France

AUTHOR ORCIDs

Jesus M. Eraso  <http://orcid.org/0000-0003-1383-8702>

Randall J. Olsen  <http://orcid.org/0000-0002-5295-4178>

S. Wesley Long  <http://orcid.org/0000-0003-3043-5307>

Ryan Gadd  <http://orcid.org/0009-0007-8236-3030>

Samer Kayal  <http://orcid.org/0000-0002-4734-1713>

James M. Musser  <http://orcid.org/0000-0002-7765-4956>

DIRECT CONTRIBUTION

This article is a direct contribution from James M. Musser, a Fellow of the American Academy of Microbiology, who arranged for and secured reviews by David Morens, National Institute of Allergy and Infectious Diseases; Allison Carey, University of Utah; and Barbara Spellerberg, University Hospital Ulm.

DATA AVAILABILITY

Genome sequences have been submitted to the National Center for Biotechnology Information (NCBI) under Bioproject accession number [PRJNA925803](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA925803) ([CP158994](https://www.ncbi.nlm.nih.gov/submitter/CP158994)-[CP159107](https://www.ncbi.nlm.nih.gov/submitter/CP159107)).

ETHICS APPROVAL

Mouse infection studies were performed in accordance with protocol IS00007779 approved by the Houston Methodist Research Institute Institutional Animal Care and Use Committee.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Fig. S1 (mBio02578-24-s0001.eps). Flowchart depicting processes for analysis of SDSE stG62647 genomes, mouse virulence, and transcriptomes.

Fig. S2 (mBio02578-24-s0002.eps). Genome sequencing reads obtained for 38 SDSE stG62647 isolates.

Fig. S3 (mBio02578-24-s0003.eps). Comparison of mouse near-mortality for CC17 and CC20 isolates.

Fig. S4 (mBio02578-24-s0004.eps). Temporal distribution of stG62647 isolates within the 499 French Brittany clinical isolates.

Supplemental Tables (mBio02578-24-s0005.docx). Tables S1–S8.

REFERENCES

- Alves-Barroco C, Brito PH, Santos-Sanches I, Fernandes AR. 2022. Phylogenetic analysis and accessory genome diversity reveal insight into the evolutionary history of *Streptococcus dysgalactiae*. *Front Microbiol* 13:952110. <https://doi.org/10.3389/fmicb.2022.952110>
- Didkowska A, Żmuda P, Kwiecień E, Rzewuska M, Klich D, Krajewska-Wędzina M, Witkowski L, Żychska M, Kaczmarkowska A, Orłowska B, Anusz K. 2020. Microbiological assessment of sheep lymph nodes with lymphadenitis found during post-mortem examination of slaughtered sheep: implications for veterinary-sanitary meat control. *Acta Vet Scand* 62:48. <https://doi.org/10.1186/s13028-020-00547-x>
- Hwang YT, Wobeser G, Larivière S, Messier F. 2002. *Streptococcus equisimilis* infection in striped skunks (*Mephitis mephitis*) in Saskatchewan. *J Wildl Dis* 38:641–643. <https://doi.org/10.7589/0090-3558-38.3.641>
- Oh S-I, Kim JW, Jung J-Y, Chae M, Lee Y-R, Kim JH, So B, Kim H-Y. 2018. Pathologic and molecular characterization of *Streptococcus dysgalactiae* subsp. *equisimilis* infection in neonatal piglets. *J Vet Sci* 19:313–317. <https://doi.org/10.4142/jvs.2018.19.2.313>
- Preziuso S, Laus F, Tejada AR, Valente C, Cuteri V. 2010. Detection of *Streptococcus dysgalactiae* subsp. *equisimilis* in equine nasopharyngeal swabs by PCR. *J Vet Sci* 11:67–72. <https://doi.org/10.4142/jvs.2010.11.1.67>
- Porcellato D, Smistad M, Skeie SB, Jørgensen HJ, Austbø L, Oppegaard O. 2021. Whole genome sequencing reveals possible host species adaptation of *Streptococcus dysgalactiae*. *Sci Rep* 11:17350. <https://doi.org/10.1038/s41598-021-96710-z>
- Xie O, Morris JM, Hayes AJ, Towers RJ, Jespersen MG, Lees JA, Ben Zakour NL, Berking O, Baines SL, Carter GP, et al. 2024. Inter-species gene flow drives ongoing evolution of *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subsp. *equisimilis*. *Nat Commun* 15:2286. <https://doi.org/10.1038/s41467-024-46530-2>
- Jensen A, Kilian M. 2012. Delineation of *Streptococcus dysgalactiae*, its subspecies, and its clinical and phylogenetic relationship to *Streptococcus pyogenes*. *J Clin Microbiol* 50:113–126. <https://doi.org/10.1128/JCM.05900-11>
- Beres SB, Olsen RJ, Long SW, Eraso JM, Boukthir S, Faili A, Kayal S, Musser JM. 2023. Analysis of the genomics and mouse virulence of an emergent clone of *Streptococcus dysgalactiae* subspecies *equisimilis*. *Microbiol Spectr* 11:e0455022. <https://doi.org/10.1128/spectrum.04550-22>
- Ikebe T, Otsuka H, Chiba K, Kazawa Y, Yamaguchi T, Okuno R, Date Y, Sasaki M, Isobe J, Ohnishi M, Akeda Y. 2022. Natural mutation in the regulatory gene (*srrG*) influences virulence-associated genes and enhances invasiveness in *Streptococcus dysgalactiae* subsp. *equisimilis* strains isolated from cases of streptococcal toxic shock syndrome. *EBioMedicine* 81:104133. <https://doi.org/10.1016/j.ebiom.2022.104133>
- Watanabe S, Shimomura Y, Ubukata K, Kirikae T, Miyoshi-Akiyama T. 2013. Concomitant regulation of host tissue-destroying virulence factors and carbohydrate metabolism during invasive diseases induced by

- group G streptococci. *J Infect Dis* 208:1482–1493. <https://doi.org/10.1093/infdis/jit353>
12. Humar D, Datta V, Bast DJ, Beall B, De Azavedo JCS, Nizet V. 2002. Streptolysin S and necrotising infections produced by group G streptococcus. *Lancet* 359:124–129. [https://doi.org/10.1016/S0140-6736\(02\)07371-3](https://doi.org/10.1016/S0140-6736(02)07371-3)
 13. Ogura K, Okumura K, Shimizu Y, Kirikae T, Miyoshi-Akiyama T. 2018. Pathogenicity induced by invasive infection of *Streptococcus dysgalactiae* subsp. *equisimilis* in a mouse model of diabetes. *Front Microbiol* 9:2128. <https://doi.org/10.3389/fmicb.2018.02128>
 14. Steiner K, Malke H. 2002. Dual control of streptokinase and streptolysin S production by the *covRS* and *fasCAX* two-component regulators in *Streptococcus dysgalactiae* subsp. *equisimilis*. *Infect Immun* 70:3627–3636. <https://doi.org/10.1128/IAI.70.7.3627-3636.2002>
 15. Yoshida H, Takahashi T, Nakamura M, Ørverby A, Takahashi T, Ubukata K, Matsui H. 2016. A highly susceptible CD46 transgenic mouse model of subcutaneous infection with *Streptococcus dysgalactiae* subspecies *equisimilis*. *J Infect Chemother* 22:229–234. <https://doi.org/10.1016/j.jiac.2016.01.001>
 16. Davies MR, McMillan DJ, Beiko RG, Barroso V, Geffers R, Sriprakash KS, Chhatwal GS. 2007. Virulence profiling of *Streptococcus dysgalactiae* subspecies *equisimilis* isolated from infected humans reveals 2 distinct genetic lineages that do not segregate with their phenotypes or propensity to cause diseases. *Clin Infect Dis* 44:1442–1454. <https://doi.org/10.1086/516780>
 17. Geyer A, Schmidt KH. 2000. Genetic organisation of the M protein region in human isolates of group C and G streptococci: two types of multigene regulator-like (*mgrC*) regions. *Mol Gen Genet* 262:965–976. <https://doi.org/10.1007/pl00008665>
 18. Kaci A, Jonassen CM, Skrede S, Sivertsen A, Norwegian Study Group on *Streptococcus dysgalactiae*, Steinbakk M, Oppegaard O. 2023. Genomic epidemiology of *Streptococcus dysgalactiae* subsp. *equisimilis* strains causing invasive disease in Norway during 2018. *Front Microbiol* 14:1171913. <https://doi.org/10.3389/fmicb.2023.1171913>
 19. Lo HH, Cheng WS. 2015. Distribution of virulence factors and association with *emm* polymorphism or isolation site among beta-hemolytic group G *Streptococcus dysgalactiae* subspecies *equisimilis*. *APMIS* 123:45–52. <https://doi.org/10.1111/apm.12305>
 20. Matsue M, Ogura K, Sugiyama H, Miyoshi-Akiyama T, Takemori-Sakai Y, Iwata Y, Wada T, Okamoto S. 2020. Pathogenicity characterization of prevalent-type *Streptococcus dysgalactiae* subsp. *equisimilis* strains. *Front Microbiol* 11:97. <https://doi.org/10.3389/fmicb.2020.00097>
 21. Oppegaard O, Mylvaganam H, Skrede S, Kittang BR. 2018. Exploring the arthritogenicity of *Streptococcus dysgalactiae* subspecies *equisimilis*. *BMC Microbiol* 18:17. <https://doi.org/10.1186/s12866-018-1160-5>
 22. Takahashi T, Ubukata K, Watanabe E. 2011. Invasive infection caused by *Streptococcus dysgalactiae* subsp. *equisimilis*: characteristics of strains and clinical features. *J Infect Chemother* 17:1–10. <https://doi.org/10.1007/s10156-010-0084-2>
 23. Ekelund K, Skinhøj P, Madsen J, Konradsen HB. 2005. Invasive group A, B, C and G streptococcal infections in Denmark 1999–2002: epidemiological and clinical aspects. *Clin Microbiol Infect* 11:569–576. <https://doi.org/10.1111/j.1469-0691.2005.01169.x>
 24. Rantala S, Vuopio-Varkila J, Vuento R, Huhtala H, Syrjänen J. 2009. Clinical presentations and epidemiology of beta-haemolytic streptococcal bacteraemia: a population-based study. *Clin Microbiol Infect* 15:286–288. <https://doi.org/10.1111/j.1469-0691.2008.02672.x>
 25. Broyles LN, Van Beneden C, Beall B, Facklam R, Shewmaker PL, Malpiedi P, Daily P, Reingold A, Farley MM. 2009. Population-based study of invasive disease due to beta-hemolytic streptococci of groups other than A and B. *Clin Infect Dis* 48:706–712. <https://doi.org/10.1086/597035>
 26. Rantala S. 2014. *Streptococcus dysgalactiae* subsp. *equisimilis* bacteremia: an emerging infection. *Eur J Clin Microbiol Infect Dis* 33:1303–1310. <https://doi.org/10.1007/s10096-014-2092-0>
 27. Oppegaard O, Skrede S, Mylvaganam H, Kittang BR. 2016. Temporal trends of β-haemolytic streptococcal osteoarticular infections in western Norway. *BMC Infect Dis* 16:535. <https://doi.org/10.1186/s12879-016-1874-7>
 28. Oppegaard O, Mylvaganam H, Skrede S, Lindemann PC, Kittang BR. 2017. Emergence of a *Streptococcus dysgalactiae* subspecies *equisimilis* stG62647-lineage associated with severe clinical manifestations. *Sci Rep* 7:7589. <https://doi.org/10.1038/s41598-017-08162-z>
 29. Shimomura Y, Okumura K, Murayama SY, Yagi J, Ubukata K, Kirikae T, Miyoshi-Akiyama T. 2011. Complete genome sequencing and analysis of a Lancefield group G *Streptococcus dysgalactiae* subsp. *equisimilis* strain causing streptococcal toxic shock syndrome (STSS). *BMC Genomics* 12:17. <https://doi.org/10.1186/1471-2164-12-17>
 30. Hikone M, Nagai K, Inoue K, Noji S, Sugiyama K, Hamabe Y. 2021. Invasive infection due to *Streptococcus dysgalactiae* subspecies *equisimilis* causing endocarditis and ventriculitis: a case report. *Clin Case Rep* 9:e04638. <https://doi.org/10.1002/ccr3.4638>
 31. Jourani M, Duprez T, Roelants V, Rodriguez-Villalobos H, Hantson P. 2017. Acute bacterial meningitis and systemic abscesses due to *Streptococcus dysgalactiae* subsp. *equisimilis* infection. *Case Rep Infect Dis* 2017:8645859. <https://doi.org/10.1155/2017/8645859>
 32. Watanabe S, Takemoto N, Ogura K, Miyoshi-Akiyama T. 2016. Severe invasive streptococcal infection by *Streptococcus pyogenes* and *Streptococcus dysgalactiae* subsp. *equisimilis*. *Microbiol Immunol* 60:1–9. <https://doi.org/10.1111/1348-0421.12334>
 33. Chochua S, Rivers J, Mathis S, Li Z, Velusamy S, McGee L, Van Beneden C, Li Y, Metcalf BJ, Beall B. 2019. Emergent invasive group A *Streptococcus dysgalactiae* subsp. *equisimilis*, United States, 2015–2018. *Emerg Infect Dis* 25:1543–1547. <https://doi.org/10.3201/eid2508.181758>
 34. Wright CM, Moorin R, Pearson G, Dyer J, Carapetis J, Manning L. 2022. Invasive infections caused by Lancefield groups C/G and A *Streptococcus*, Western Australia, Australia, 2000–2018. *Emerg Infect Dis* 28:2190–2197. <https://doi.org/10.3201/eid2811.220029>
 35. Barros RR. 2021. Antimicrobial resistance among Beta-hemolytic *Streptococcus* in Brazil: an overview. *Antibiotics (Basel)* 10:973. <https://doi.org/10.3390/antibiotics10080973>
 36. Behera B, Mathur P, Bhardwaj N, Jain N, Misra MC, Kapil A, Singh S. 2014. Antibiotic susceptibilities, streptococcal pyrogenic exotoxin gene profiles among clinical isolates of group C or G *Streptococcus dysgalactiae* subsp. *equisimilis* & of group G *S. anginosus* group at a tertiary care centre. *Indian J Med Res* 139:438–445.
 37. Biedenbach DJ, Toleman MA, Walsh TR, Jones RN. 2006. Characterization of fluoroquinolone-resistant beta-hemolytic *Streptococcus* spp. isolated in North America and Europe including the first report of fluoroquinolone-resistant *Streptococcus dysgalactiae* subspecies *equisimilis*: report from the SENTRY antimicrobial surveillance program (1997–2004). *Diagn Microbiol Infect Dis* 55:119–127. <https://doi.org/10.1016/j.diagmicrobio.2005.12.006>
 38. Liu L-C, Tsai J-C, Hsueh P-R, Tseng S-P, Hung W-C, Chen H-J, Teng L-J. 2008. Identification of *tet(S)* gene area in tetracycline-resistant *Streptococcus dysgalactiae* subsp. *equisimilis* clinical isolates. *J Antimicrob Chemother* 61:453–455. <https://doi.org/10.1093/jac/dkm500>
 39. Pinho MD, Melo-Cristino J, Ramirez M, Portuguese Group for the Study of Streptococcal Infections. 2010. Fluoroquinolone resistance in *Streptococcus dysgalactiae* subsp. *equisimilis* and evidence for a shared global gene pool with *Streptococcus pyogenes*. *Antimicrob Agents Chemother* 54:1769–1777. <https://doi.org/10.1128/AAC.01377-09>
 40. Sunaoshi K, Murayama SY, Adachi K, Yagoshi M, Okuzumi K, Chiba N, Morozumi M, Ubukata K. 2010. Molecular *emm* genotyping and antibiotic susceptibility of *Streptococcus dysgalactiae* subsp. *equisimilis* isolated from invasive and non-invasive infections. *J Med Microbiol* 59:82–88. <https://doi.org/10.1099/jmm.0.013201-0>
 41. Uh Y, Hwang GY, Jang IH, Cho HM, Noh SM, Kim HY, Kwon O, Yoon KJ. 2007. Macrolide resistance trends in beta-hemolytic streptococci in a tertiary Korean hospital. *Yonsei Med J* 48:773–778. <https://doi.org/10.3349/ymj.2007.48.5.773>
 42. Itzek A, Weißbach V, Meinrup D, Rieß B, van der Linden M, Borgmann S. 2023. Epidemiological and clinical features of *Streptococcus dysgalactiae* ssp. *equisimilis* stG62647 and other *emm* types in Germany. *Pathogens* 12:589. <https://doi.org/10.3390/pathogens12040589>
 43. Leitner E, Zollner-Schwetz I, Zarfel G, Masoud-Landgraf L, Gehrler M, Wagner-Eibel U, Grisold AJ, Feierl G. 2015. Prevalence of *emm* types and antimicrobial susceptibility of *Streptococcus dysgalactiae* subsp. *equisimilis* in Austria. *Int J Med Microbiol* 305:918–924. <https://doi.org/10.1016/j.ijmm.2015.10.001>
 44. Lother SA, Demczuk W, Martin I, Mulvey M, Dufault B, Lagacé-Wiens P, Keynan Y. 2017. Clonal clusters and virulence factors of group C and G

- Streptococcus* causing severe infections, Manitoba, Canada, 2012–2014. *Emerg Infect Dis* 23:1079–1088. <https://doi.org/10.3201/eid2307.161259>
45. Rojo-Bezares B, Toca L, Azcona-Gutiérrez JM, Ortega-Unanue M, Toledano P, Sáenz Y. 2021. *Streptococcus dysgalactiae* subsp. *equisimilis* from invasive and non-invasive infections in Spain: combining epidemiology, molecular characterization, and genetic diversity. *Eur J Clin Microbiol Infect Dis* 40:1013–1021. <https://doi.org/10.1007/s10096-020-04119-9>
 46. Rößler S, Berner R, Jacobs E, Toepfner N. 2018. Prevalence and molecular diversity of invasive *Streptococcus dysgalactiae* and *Streptococcus pyogenes* in a German tertiary care medical centre. *Eur J Clin Microbiol Infect Dis* 37:1325–1332. <https://doi.org/10.1007/s10096-018-3254-2>
 47. Traverso F, Blanco A, Villalón P, Beratz N, Sáez Nieto JA, Lopardo H, National Collaborative Group for the Study of Streptococci and Related Bacteria. 2016. Molecular characterization of invasive *Streptococcus dysgalactiae* subsp. *equisimilis*. Multicenter study: Argentina 2011–2012. *Rev Argent Microbiol* 48:279–289. <https://doi.org/10.1016/j.ram.2016.07.001>
 48. Kachroo P, Eraso JM, Beres SB, Olsen RJ, Zhu L, Nasser W, Bernard PE, Cantu CC, Saavedra MO, Arredondo MJ, et al. 2019. Integrated analysis of population genomics, transcriptomics and virulence provides novel insights into *Streptococcus pyogenes* pathogenesis. *Nat Genet* 51:548–559. <https://doi.org/10.1038/s41588-018-0343-1>
 49. Beres SB, Kachroo P, Nasser W, Olsen RJ, Zhu L, Flores AR, de la Riva I, Paez-Mayorga J, Jimenez FE, Cantu C, Vuopio J, Jalava J, Kristinsson KG, Gottfredsson M, Corander J, Fittipaldi N, Di Luca MC, Petrelli D, Vitali LA, Raiford A, Jenkins L, Musser JM. 2016. Transcriptome remodeling contributes to epidemic disease caused by the human pathogen *Streptococcus pyogenes*. *MBio* 7:e00403-16. <https://doi.org/10.1128/mBio.00403-16>
 50. Beres SB, Zhu L, Pruitt L, Olsen RJ, Faili A, Kayal S, Musser JM. 2022. Integrative reverse genetic analysis identifies polymorphisms contributing to decreased antimicrobial agent susceptibility in *Streptococcus pyogenes*. *MBio* 13:e0361821. <https://doi.org/10.1128/mbio.03618-21>
 51. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. 2007. Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8:460. <https://doi.org/10.1186/1471-2105-8-460>
 52. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>
 53. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 27:737–746. <https://doi.org/10.1101/gr.214270.116>
 54. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
 55. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, Wyres KL, Holt KE. 2021. Tricycler: consensus long-read assemblies for bacterial genomes. *Genome Biol* 22:266. <https://doi.org/10.1186/s13059-021-02483-z>
 56. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
 57. Carver T, Harris SR, Berriman M, Parkhill J, McQuillan JA. 2012. Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* 28:464–469. <https://doi.org/10.1093/bioinformatics/btr703>
 58. Olsen RJ, Sitkiewicz I, Ayeras AA, Gonulal VE, Cantu C, Beres SB, Green NM, Lei B, Humbird T, Greaver J, Chang E, Ragasa WP, Montgomery CA, Cartwright J Jr, McGeer A, Low DE, Whitney AR, Cagle PT, Blasdel TL, DeLeo FR, Musser JM. 2010. Decreased necrotizing fasciitis capacity caused by a single nucleotide mutation that alters a multiple gene virulence axis. *Proc Natl Acad Sci U S A* 107:888–893. <https://doi.org/10.1073/pnas.0911811107>
 59. Zhu L, Olsen RJ, Lee JD, Porter AR, DeLeo FR, Musser JM. 2017. Contribution of secreted NADase and streptolysin O to the pathogenesis of epidemic serotype M1 *Streptococcus pyogenes* infections. *Am J Pathol* 187:605–613. <https://doi.org/10.1016/j.ajpath.2016.11.003>
 60. Zhu L, Olsen RJ, Nasser W, Beres SB, Vuopio J, Kristinsson KG, Gottfredsson M, Porter AR, DeLeo FR, Musser JM. 2015. A molecular trigger for intercontinental epidemics of group A *Streptococcus*. *J Clin Invest* 125:3545–3559. <https://doi.org/10.1172/JCI82478>
 61. National Research Council, Committee for the Update of the Guide for the Care and Use of Laboratory Animals. 2011. *Guide for the care and use of laboratory animals*. National Academies press, Washington, DC.
 62. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data.
 63. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
 64. Magoc T, Wood D, Salzberg SL. 2013. EDGE-pro: estimated degree of gene expression in prokaryotic genomes. *Evol Bioinform Online* 9:127–136. <https://doi.org/10.4137/EBO.S11250>
 65. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15:550. <https://doi.org/10.1186/s13059-014-0550-8>
 66. Marini F, Binder H. 2019. pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components. *BMC Bioinformatics* 20:331. <https://doi.org/10.1186/s12859-019-2879-1>
 67. Marini F, Linke J, Binder H. 2020. Ideal: an R/Bioconductor package for interactive differential expression analysis. *BMC Bioinformatics* 21:565. <https://doi.org/10.1186/s12859-020-03819-5>
 68. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo S, Beaudoin C, Floto RA, Frost SDW, Corander J, Bentley SD, Parkhill J. 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 21:180. <https://doi.org/10.1186/s13059-020-02090-4>
 69. Lambertsen LM, Ingels H, Schønheyder HC, Hoffmann S, Danish Streptococcal Surveillance Collaboration Group. 2011. 2014. Nationwide laboratory-based surveillance of invasive beta-haemolytic streptococci in Denmark from 2005 to 2011. *Clin Microbiol Infect* 20:216–223. <https://doi.org/10.1111/1469-0691.12378>
 70. Ruppen C, Rasmussen M, Casanova C, Sendi P. 2017. A 10-year observational study of *Streptococcus dysgalactiae* bacteraemia in adults: frequent occurrence among female intravenous drug users. *Swiss Med Wkly* 147:w14469. <https://doi.org/10.4414/smww.2017.14469>
 71. Ambroset C, Coluzzi C, Guédon G, Devignes M-D, Loux V, Lacroix T, Payot S, Leblond-Bourget N. 2015. New insights into the classification and integration specificity of *Streptococcus* integrative conjugative elements through extensive genome exploration. *Front Microbiol* 6:1483. <https://doi.org/10.3389/fmicb.2015.01483>
 72. Kreikemeyer B, Mclver KS, Podbielski A. 2003. Virulence factor regulation and regulatory networks in *Streptococcus pyogenes* and their impact on pathogen-host interactions. *Trends Microbiol* 11:224–232. [https://doi.org/10.1016/s0966-842x\(03\)00098-2](https://doi.org/10.1016/s0966-842x(03)00098-2)
 73. Vega LA, Malke H, Mclver KS. 2016. Edited by Ferretti JJ, Stevens DL, and Fischetti VA. *Virulence-related transcriptional regulators of Streptococcus pyogenes*, in *Streptococcus pyogenes: basic biology to clinical manifestations*. Oklahoma City, OK.
 74. de Jong A, Kuipers OP, Kok J. 2022. FUNAGE-Pro: comprehensive web server for gene set enrichment analysis of prokaryotes. *Nucleic Acids Res* 50:W330–W336. <https://doi.org/10.1093/nar/gkac441>
 75. Sumbly P, Whitney AR, Graviss EA, DeLeo FR, Musser JM. 2006. Genome-wide analysis of group A streptococci reveals a mutation that modulates global phenotype and disease specificity. *PLoS Pathog* 2:e5. <https://doi.org/10.1371/journal.ppat.0020005>
 76. Eraso JM, Kachroo P, Olsen RJ, Beres SB, Zhu L, Badu T, Shannon S, Cantu CC, Saavedra MO, Kubiak SL, Porter AR, DeLeo FR, Musser JM. 2020. Genetic heterogeneity of the Spy1336/R28-Spy1337 virulence axis in *Streptococcus pyogenes* and effect on gene transcript levels and pathogenesis. *PLoS One* 15:e0229064. <https://doi.org/10.1371/journal.pone.0229064>
 77. Lees JA, Galardini M, Bentley SD, Weiser JN, Corander J. 2018. Pyseer: a comprehensive tool for microbial pangenome-wide association studies. *Bioinformatics* 34:4310–4312. <https://doi.org/10.1093/bioinformatics/bty539>
 78. Delavat F, Miyazaki R, Carraro N, Pradervand N, van der Meer JR. 2017. The hidden life of integrative and conjugative elements. *FEMS Microbiol Rev* 41:512–537. <https://doi.org/10.1093/femsre/fux008>

79. Johnson CM, Grossman AD. 2015. Integrative and conjugative elements (ICES): what they do and how they work. *Annu Rev Genet* 49:577–601. <https://doi.org/10.1146/annurev-genet-112414-055018>
80. Treviño J, Perez N, Ramirez-Peña E, Liu Z, Shelburne SA 3rd, Musser JM, Sumbly P. 2009. CovS simultaneously activates and inhibits the CovR-mediated repression of distinct subsets of group A *Streptococcus virulence* factor-encoding genes. *Infect Immun* 77:3141–3149. <https://doi.org/10.1128/IAI.01560-08>
81. Turner CE, Bubba L, Efstratiou A. 2019. Pathogenicity factors in group C and G streptococci. *Microbiol Spectr* 7. <https://doi.org/10.1128/microbiolspec.gpp3-0020-2018>
82. Kreikemeyer B, Boyle MD, Buttaro BA, Heinemann M, Podbielski A. 2001. Group A streptococcal growth phase-associated virulence factor regulation by a novel operon (Fas) with homologies to two-component-type regulators requires a small RNA molecule. *Mol Microbiol* 39:392–406. <https://doi.org/10.1046/j.1365-2958.2001.02226.x>
83. Ribardo DA, McIver KS. 2006. Defining the Mga regulon: Comparative transcriptome analysis reveals both direct and indirect regulation by Mga in the group A streptococcus. *Mol Microbiol* 62:491–508. <https://doi.org/10.1111/j.1365-2958.2006.05381.x>
84. Levin JC, Wessels MR. 1998. Identification of *csrR/csrS*, a genetic locus that regulates hyaluronic acid capsule synthesis in group A *Streptococcus*. *Mol Microbiol* 30:209–219. <https://doi.org/10.1046/j.1365-2958.1998.01057.x>
85. Roshika R, Jain I, Glenaldo T, Sickler T, Musser JM, Sumbly P. 2022. Characterization of M-type-specific pilus expression in group A *Streptococcus*. *J Bacteriol* 204:e0027022. <https://doi.org/10.1128/jb.00270-22>
86. Belotserkovsky I, Baruch M, Peer A, Dov E, Ravins M, Mishalian I, Persky M, Smith Y, Hanski E. 2009. Functional analysis of the quorum-sensing streptococcal invasion locus (*sil*). *PLoS Pathog* 5:e1000651. <https://doi.org/10.1371/journal.ppat.1000651>
87. Eran Y, Getter Y, Baruch M, Belotserkovsky I, Padalon G, Mishalian I, Podbielski A, Kreikemeyer B, Hanski E. 2007. Transcriptional regulation of the *sil* locus by the SilCR signalling peptide and its implications on group A streptococcus virulence. *Mol Microbiol* 63:1209–1222. <https://doi.org/10.1111/j.1365-2958.2007.05581.x>
88. Nakata M, Kreikemeyer B. 2021. Genetics, structure, and function of group a streptococcal pili. *Front Microbiol* 12:616508. <https://doi.org/10.3389/fmicb.2021.616508>
89. Crotty Alexander LE, Maisey HC, Timmer AM, Rooijackers SHM, Gallo RL, von Köckritz-Blickwede M, Nizet V. 2010. M1T1 group A streptococcal pili promote epithelial colonization but diminish systemic virulence through neutrophil extracellular entrapment. *J Mol Med (Berl)* 88:371–381. <https://doi.org/10.1007/s00109-009-0566-9>
90. Flores AR, Olsen RJ, Cantu C, Pallister KB, Guerra FE, Voyich JM, Musser JM. 2017. Increased pilus production conferred by a naturally occurring mutation alters host-pathogen interaction in favor of carriage in *Streptococcus pyogenes*. *Infect Immun* 85:e00949-16. <https://doi.org/10.1128/IAI.00949-16>