Original Article
Clinical Microbiology

Ann Lab Med 2022;42:438-446
https://doi.org/10.3343/alm.2022.42.4.438
ISSN 2234-3806 • eISSN 2234-3814

ANNALS OF LABORATORY MEDICINE

Comparing Genomic Characteristics of *Streptococcus pyogenes* Associated with Invasiveness over a 20-year Period in Korea

Hyoshim Shin, M.D.[1], Takashi Takahashi, M.D., Ph.D.[2], Seungjun Lee, M.D.[3], Eun Hwa Choi, M.D., Ph.D.[4], Takahiro Maeda, B.P.[2], Yasuto Fukushima, B.P.[2], and Sunjoo Kim, M.D., Ph.D.[3,5]

[1]Department of Laboratory Medicine, Gyeongsang National University Hospital, Jinju, Korea; [2]Laboratory of Infectious Diseases, Graduate School of Infection Control Sciences & Ōmura Satoshi Memorial Institute, Kitasato University, Tokyo, Japan; [3]Department of Laboratory Medicine, Gyeongsang National University Changwon Hospital, Changwon, Korea; [4]Department of Pediatrics, Seoul National University College of Medicine, Seoul, Korea; [5]Department of Laboratory Medicine, Gyeongsang National University College of Medicine, Institute of Health Sciences, Jinju, Korea

Background: Few studies have investigated the invasiveness of *Streptococcus pyogenes* based on whole-genome sequencing (WGS). Using WGS, we determined the genomic features associated with invasiveness of *S. pyogenes* strains in Korea.

Methods: Forty-five *S. pyogenes* strains from 1997, 2006, and 2017, including common emm types, were selected from the repository at Gyeongsang National University Hospital in Korea. In addition, 48 *S. pyogenes* strains were randomly selected depending on their invasiveness between 1997 and 2017 to evaluate the genetic evolution and the associations between invasiveness and genetic profiles. Using WGS datasets, we conducted virulence-associated DNA sequence determination, *emm* genotyping, multi-locus sequence typing (MLST), and superantigen gene profiling.

Results: In total, 87 strains were included in this study. There were no significant differences in the genomic features throughout the study periods. Four genes, *csn1*, *ispE*, *nisK*, and *citC*, were detected only in invasive strains. There was a significant association between invasiveness and *emm* cluster type A-C3, including, *emm*1.0, *emm*1.18, *emm*1.3, and *emm*1.76 ($P<0.05$). The predominant *emm*1 lineage belonged to ST28. There were no associations between invasiveness and superantigen gene profiles.

Conclusions: This is the first study using WGS datasets of *S. pyogenes* strains collected between 1997 and 2017 in Korea. Streptococcal invasiveness is associated with the presence of *csn1*, *ispE*, *nisK*, and *citC*. The *emm*1 lineage and ST28 clone are explicitly associated with invasiveness, whereas genomic features remained stable over the 20-year period.

Key Words: *Streptococcus pyogenes*, Invasiveness, Whole genome sequencing

Received: June 11, 2021
Revision received: July 23, 2021
Accepted: December 6, 2021

Corresponding author:
Sunjoo Kim, M.D., Ph.D.
Department of Laboratory Medicine, Gyeongsang National University Changwon Hospital, 11 Samjungja-ro, Seongsan-gu, Changwon 51472, Korea
Tel: +82-55-214-3072
Fax: +82-55-214-3087
E-mail: sjkim8239@hanmail.net

INTRODUCTION

Streptococcus pyogenes causes a wide spectrum of diseases in humans, from mild tonsillopharyngitis, impetigo, and scarlet fever to severe and invasive sepsis, arthritis, necrotizing fasciitis, and streptococcal toxic shock syndrome (STSS) [1]. More than 500,000 deaths due to streptococcal infections are reported worldwide each year, making *S. pyogenes* a major pathogen associated with high morbidity and mortality [2]. *S. pyogenes* strains are genetically diverse and have various virulence factors, including adhesion molecules, superantigens, DNases, proteases, and M protein, that are involved in its complex pathoge-

nicity.

S. pyogenes strains can be classified by *emm* typing that is based on PCR amplification and amplicon sequencing of the *emm* gene, encoding M protein. Multilocus sequence typing (MLST) based on the amplification and sequencing of seven housekeeping genes and pulsed-field gel electrophoresis (PFGE) of the large genomic fragment have also been used in epidemiological studies [3, 4]. Whole-genome sequencing (WGS) datasets help in discriminating closely related strains and allow epidemiological analysis of small infection clusters. WGS analysis is an ideal molecular typing method for bacteria, as it provides complete genetic information of a strain. Until recently, because of the high cost and technical complexity, WGS was beyond the reach of average diagnostic laboratories [5]. Little has been published on the invasiveness of *S. pyogenes* strains collected in Korea. Characterization of *S. pyogenes* in a longitudinal surveillance study of WGS datasets would provide important information about the genomic characteristics, virulence-associated gene profiles, and genomic dynamics during the long period of time.

We genomically characterized *S. pyogenes* strains collected in Korea between 1997 and 2017 by determining their emm types, MLST-based sequence types (STs), and superantigen gene profiles. We then investigated whether the genomic characteristics of *S. pyogenes* strains differed based on invasiveness and isolation year.

MATERIALS AND METHODS

Bacterial strain selection

All strains used in this study were collected between 1997 and 2017 and stored in the repository at Gyeongsang National University Hospital (GNUH) in Gyeongnam Province, Korea. Forty-five *S. pyogenes* strains were selected according to the common *emm* types in three years: 1997, 2006, and 2017. Forty-eight strains were randomly selected based on their invasiveness between 1997 and 2017. An "invasive strain" was defined as one isolated from a normally sterile body fluid, such as blood, cerebrospinal fluid, pleural fluid, pericardial fluid, joint fluid, bone aspirate, or a deep-tissue abscess [6]. Fig. 1 shows a flow chart of the strain selection procedure. In total, 87 strains were included in this study. For the first analysis to evaluate the genetic evolution over a 20-year time span, non-invasive strains (N=45) were selected every 10 years: 1997, 2006, and 2017. In the second analysis to evaluate assocations between invasiveness and genetic profiles, non-invasive and invasive strains were compared. Sixty-three non-invasive strains were isolated from the throats of carriers who did not have any symptoms or signs of tonsillitis. The other 24 invasive strains were isolated from blood (N=21) or joint fluid (N=3) of patients (Fig. 1).

Bacteria were identified using a Vitek-2 automated identification system (BioMérieux Inc., Marcy l'Étoile, France). All strains were inoculated in 30% glycerol in Todd-Hewitt broth and stored at –70°C. They were recovered on blood agar plates for

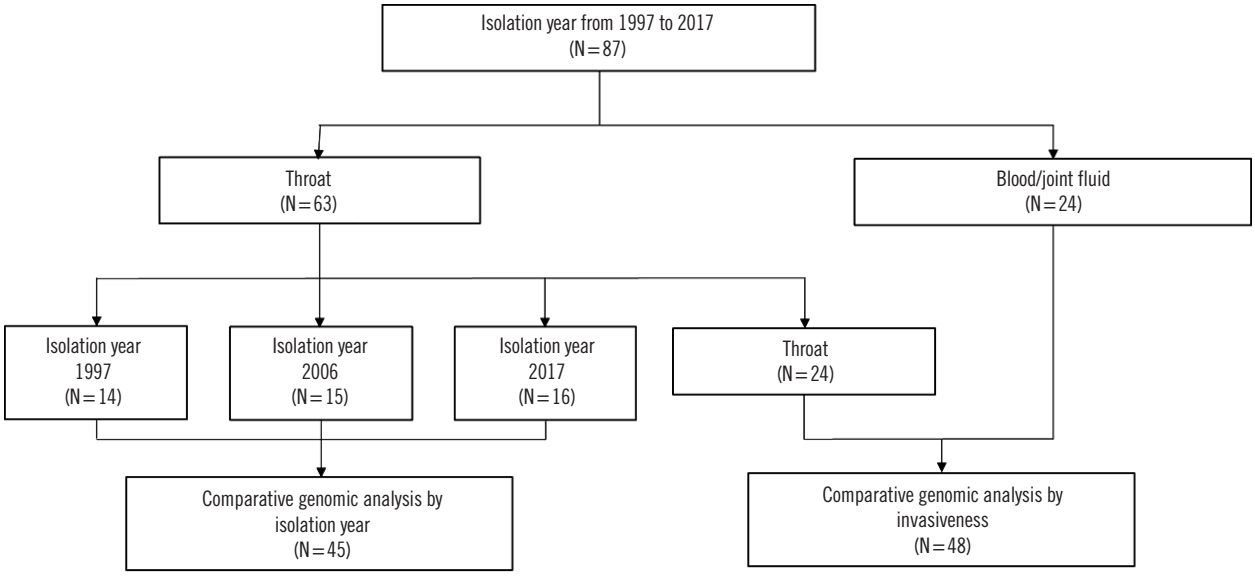

Fig. 1. Flow chart of strain selection according to isolation year and source of *Streptococcus pyogenes* isolates.

genetic analysis.

The study protocol was approved by the Institutional Review Board of GNUH (approval number: GNUCH 2018-01-008). Informed consent was waived because of the retrospective nature of the study.

Genomic DNA extraction and WGS

Genomic DNA was extracted using a Wizard Genomic DNA Isolation Kit (Promega, Madison, WI, USA). The DNA and potential culture contamination were checked by 16S rRNA gene sequencing using an ABI 3730 DNA sequencer (Applied Biosystems, Foster City, CA, USA). A draft genome sequence of each strain was generated by MiSeq sequencing (300-bp, paired-end) using a MiSeq Reagent Kit v3 (Illumina, San Diego, CA, USA). Sequencing libraries were prepared using the TruSeq DNA LT sample Prep Kit (Illumina). The Illumina sequencing data were assembled with SPAdes v3.13.0 (Algorithmic Biology Lab, St. Petersburg Academic University of the Russian Academy of Sciences, St. Petersburg, Russia). The EzBioCloud genome database was used for gene finding and functional annotation of the whole-genome assemblies (https://www.ezbiocloud.net). Protein-coding DNA sequences (CDSs) were predicted using Prodigal 2.6.2 [7]. The CDSs were classified based on their roles, with reference to orthologous groups (EggNOG v4.5; http://eggnogdb.embl.de). For more detailed functional annotation, the predicted CDSs were compared with those from the Swiss-Prot (https://www.uniprot.org), KEGG (http://www.genome.jp/kegg/), and SEED (http://pubseed.theseed.org) databases using the UBLAST (https://www.drive5.com/) program.

Comparative genome (CG) analyses

CG analyses comprised two steps (Fig. 1). The first analysis was conducted according to the isolation year (strains of 1997 vs. those of 2006 vs. those of 2017). The second analysis was conducted according to invasiveness (strains from the throat vs. those from blood/joint fluid). CG analysis was conducted by comparing functional genes based on the clustering of orthologous genes. The genome sequences of all strains were obtained from the EzBioCloud database (http://www.ezbiocloud.net/), and average nucleotide identity (ANI) values were calculated. For ANI calculation, the query genomes were cut into small fragments (1,020 bp), and high-scoring pairs between two genome sequences were selected using the USEARCH program (http://www.drive5.com/usearch). Using the calculated ANI values, a dendrogram was constructed using the unweighted pair group method. Homologous regions in a target genome to query open reading frames were determined using the USEARCH program and were aligned using pair-wise global alignment. The matched regions in the subject contig were extracted and saved as homologs [8].

emm genotyping

We used DDBJ Fast Annotation and Submission Tool v1.2.4 (DFAST; https://dfast.nig.ac.jp) for annotation and searched the sequences around the *mga* annotation encoding multiple virulence gene regulators based on the annotation data [9]. If the sequences around *mga* were not found, the sequences around the *emm*1 primer (forward: 5′-TATT(C/G)GCTTAGAAAATTAA-3′) were searched throughout the contig sequences using the FASTA format. We extracted the *emm* sequences between emm1 and *emm2* primers (reverse: 5′-GCAAGTTCTTCAGCTTGTTT-3′) using the corresponding sequences recovered from the contig data. By inserting the extracted sequences into the Centers for Disease Control and Prevention (CDC) database (https://www2.cdc.gov/vaccines/biotech/strepblast.asp), *emm* genotypes (including subtypes) were assigned to the extracted sequences.

When sequences were incompletely matched with an *emm* genotype in the CDC database, we directly PCR-amplified *emm* using bacterial DNA templates and the *emm*1/*emm2* primer set and sequenced the amplicons after purification using an AccuPrep Purification Kit (Bioneer Corp., Daejeon, Korea). The *emm* genotypes were assigned directly to the amplified sequences based on the CDC database [10, 11].

Phylogenetic tree and superantigen gene profiling

Phylogenetic analysis was accomplished using ~1.4 million bps of orthologous protein-coding regions for 45 strains according to the isolation year and 48 strains according to invasiveness, respectively (data not shown).

To determine five target genes (*speA*, *speB*, *speC*, *ssa*, and *smeZ*) encoding the superantigens (also known as exotoxins), we conducted PCR simulation analysis using the Serial Cloner (http://serialbasics.free.fr/Serial_Cloner.html) application with the contig sequences, as previously reported [12-14]. The primer sets used to amplify the *speA*, *speB*, *speC*, *ssa*, and *smeZ* are listed in Table 1. The *speB* product was included as an internal control in the PCR simulation analysis because all *S. pyogenes* strains possess the *speB* sequence (955 bp). Superantigen gene profiles were determined for each strain.

https://doi.org/10.3343/alm.2022.42.4.438

Table 1. Primer sets used to amplify *speA, speB, speC, ssa*, and *smeZ*

Superantigen gene	Forward	Reverse
speA	5′-TAAGAACCAAGAGATGG-3′	5′-ATTCTTGAGCAGTTACC-3′
Alternative *speA*	5′-CAAGAACCGAGAGATGT-3′	
speB	5′-AAGAAGCAAAAGATAGC-3′	5′-TGGTAGAAGTTACGTCC-3′
speC	5′-GATTTCTACTATTTCACC-3′	5′-AAATATCTGATCTAGTCCC-3′
ssa	5′-GTGTAGAATTGAGGTAATTG-3′	5′-TAATATAGCCTGTCTCGTAC-3′
smeZ	5′-TAACTCCTGAAAAGAGGCT-3′	5′-TTGTAGCTAGAACCAGAAG-3′
Alternative *smeZ*	5′-TAGCTCCTGAAAAGAGGCT-3′	5′-TTGTAGTTAGAACCAGAAG-3′

MLST

We determined the STs using allelic profiles consisting of seven housekeeping genes (*gki*, *gtr*, *murI*, *mutS*, *recP*, *xpt*, and *yqiL*) by inserting the contig sequences obtained into the online application MLST v2.0 (https://cge.cbs.dtu.dk/services/MLST/), which is managed by the Center for Genomic Epidemiology at the Technical University of Denmark [15]. The STs were grouped into clonal complexes (CC), whereby related STs were classified as single locus variants, differing in only one housekeeping gene. An expansion of the goeBURST program implemented in PHYLOViZ was used to produce a minimum-spanning tree representing possible relationships among the STs [16].

For novel allelic numbers/STs, we submitted the data (i.e., bacterial genotypic/phenotypic data and patient backgrounds) to the *S. pyogenes* PubMLST (http://pubmlst.org/organisms/streptococcus-pyogenes) database. The PubMLST curator assigned novel allelic numbers/STs to our strains.

Statistical analysis

We used Fisher's exact test (two-sided) to determine significant differences in categorical variables, and the chi-square test to compare the proportions in each *emm* genotype/cluster between invasive and non-invasive strains. SPSS Statistics v22.0 (IBM Corp., Armonk, NY, USA) was used for the analysis. $P<0.05$ was considered significant.

RESULTS

emm genotypes/clusters

The *emm* genotypes and cluster types are presented in Supplemental Data Table 1. In total, 21 *emm* genotypes were identified, with *emm*1 (*emm*1.00, *emm*1.18, *emm*1.30, and *emm* 1.76), *emm*4 (*emm*4.00), and *emm*12 (*emm*12.00, *emm*12.19, and *emm*12.49) accounting for 19.5%, 13.8%, and 20.7%, respectively. In total, eight *emm* cluster types were identified, among which the A-C3, A-C4, and E1 types were the most common. Fig. 2 shows the differences among the *emm* clusters according to invasiveness. There were significantly more invasive strains in cluster A-C3 than in the others ($P<0.05$). The A-C3 type included the genotypes *emm*1.0, *emm*1.18, *emm*1.3, and *emm*1.76 (Supplemental Data Table 1, Fig. 2).

ST with goeBURST diagram

The STs are presented in Supplemental Data Table 1. The 87 strains comprised 21 STs with exact loci matched against the PubMLST database. ST36, ST28, and ST39, accounting for 20.7%, 18.4%, and 11.5%, respectively, were the most frequent. There were strong associations of genetic characteristics within the MLST complex. The goeBURST diagram is shown in Fig. 3. There were 17 singletons in the CG analysis, and ST28 showed a clonal distribution of invasive strains in the second analysis. The predominant *emm*1 lineage belonged to ST28 (Fig. 3).

Phylogenetic tree and superantigen gene profiling

The phylogenetic tree based on the periodic comparison showed a sporadic distribution (data not shown). The second analysis revealed the genetic relationships among *emm* genotypes or STs. Superantigen profiling revealed that *speB* was present in all strains. *speZ-speB* and *speZ-speB-speC* profiles were present in 37.9% and 28.7% of the total strains, respectively. We found no significant association between the coexistence of different superantigen genes and invasiveness.

Virulence-associated CDSs

When comparing gene origins by pan-genome orthologous group (POG) analysis, we found *csn1*, *ispE*, *nisK*, and *citC* were more significantly present in invasive strains than in non-invasive strains (all $P<0.05$) (Table 2).

We looked for common virulence-associated CDSs among all

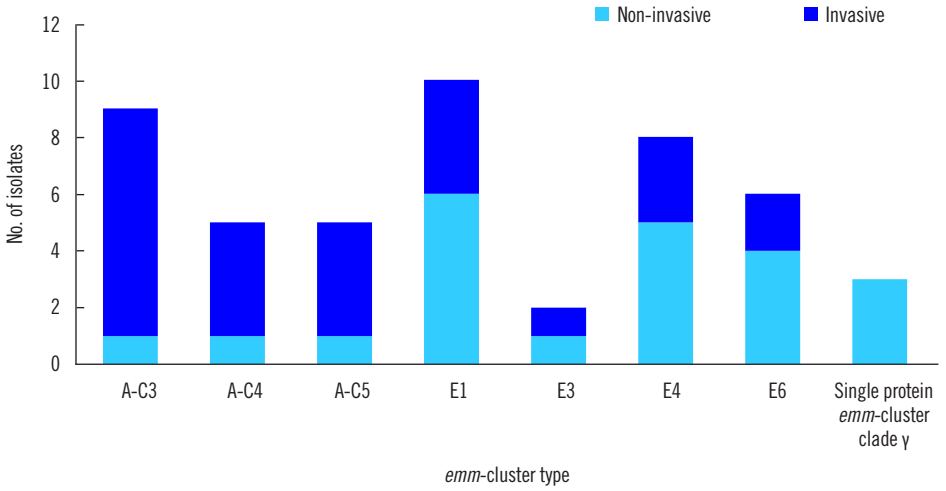

Fig. 2. Distribution of *emm* clusters according to invasiveness (N=48).

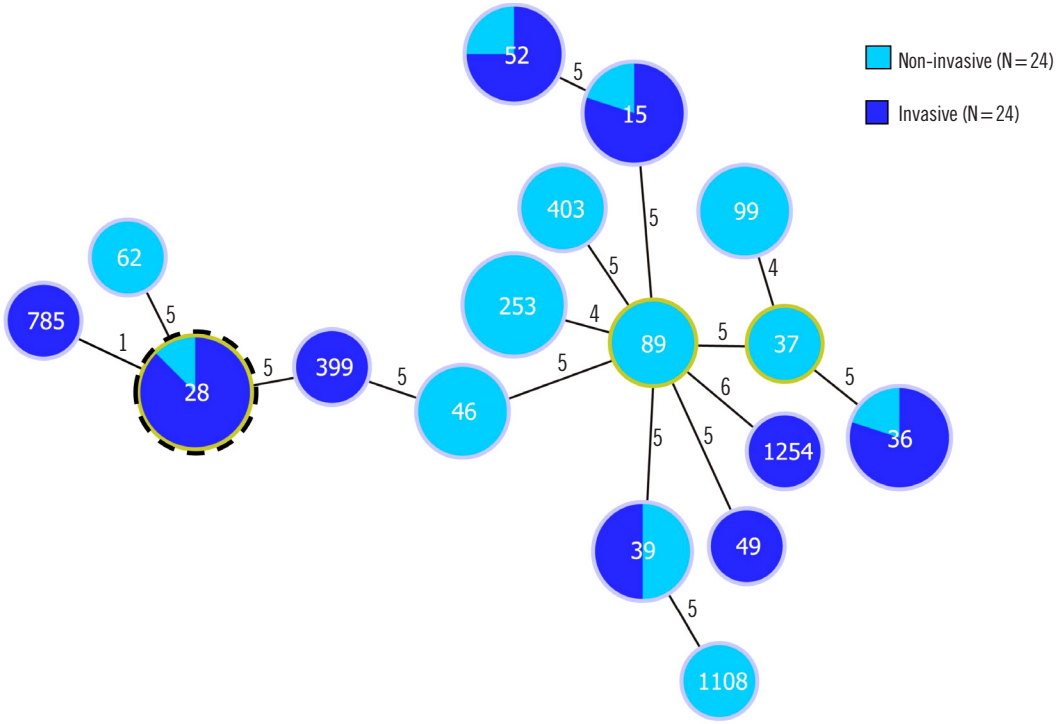

Fig. 3. goeBURST diagram of the relationships among STs according to invasiveness. The numbers in the circles indicate the STs, and the numbers near the lines indicate the number of different alleles between two connected STs. A putative CC is indicated by an outer dotted frame and corresponds to the STs with the highest number of single locus variants. ST785, a single locus variant of ST28, formed CC28.
Abbreviations: ST, sequence type; CC, clonal complex.

87 strains by searching for annotated CDSs based on functional annotation of the whole-genome assemblies. We found 25 CDSs associated with bacterial virulence. Among them, 12 (lactocepin, oleate hydratase, putative glycoslytransferases, capsule biosynthesis protein [CapA], regulatory protein MsrR, internalin-I, deoxyribonuclease, biofilm-regulatory protein, listeriolysin-regulatory protein, streptokinase, C5a peptidase, and M protein) were identified in all strains. CDSs encoding exotoxin A and procollagen-proline 3-dioxygenase were frequently detected in invasive strains (all $P<0.05$) (Table 3).

https://doi.org/10.3343/alm.2022.42.4.438

Table 2. Presence or absence of pan-genome orthologous genes according to invasiveness

Gene	Function	Non-invasive (N=24)	Invasive (N=24)	*P*
IRC3	ATP-binding, helicase, hydrolase, mitochondrion, nucleotide-binding, putative mitochondrial ATP-dependent helicase irc3	Present	Absent	0.0006
recG	DNA helicase	Present	Absent	0.0094
clpP, CLPP	Cytoplasm, hydrolase, protease, serine protease, endopeptidase Clp	Present	Absent	0.0392
topB	DNA-binding, isomerase, magnesium, metal-binding, topoisomerase, DNA topoisomerase	Present	Absent	0.0496
atoD	Transferase, acetate CoA-transferase	Present	Absent	0.0496
K02476	ATP-binding, cell membrane, kinase, membrane, nucleotide-binding, phosphoprotein, transferase, transmembrane, transmembrane helix, two-component regulatory system, histidine kinase	Present	Absent	0.0496
MTHFS	5-Formyltetrahydrofolate cyclo-ligase	Present	Absent	0.0496
csn1, cas9	Antiviral defense, DNA-binding, endonuclease, exonuclease, hydrolase, magnesium, manganese, metal-binding, nuclease, RNA-binding, CRISPR-associated endonuclease Cas9/Csn1	Absent	Present	0.0044
ispE	ATP-binding, isoprene biosynthesis, kinase, nucleotide-binding, transferase, 4-(cytidine 5´-diphospho)-2-C-methyl-D-erythritol kinase	Absent	Present	0.0094
nisK, spaK	ATP-binding, cell membrane, kinase, membrane, nucleotide-binding, phosphoprotein, transferase, transmembrane, transmembrane helix, two-component regulatory system, histidine kinase	Absent	Present	0.0355
citC	ATP-binding, ligase, nucleotide-binding, (citrate [pro-3S]-lyase) ligase	Absent	Present	0.0496

Abbreviation: CRISPR, clustered regularly interspaced short palindromic repeats.

Table 3. Comparison of CDSs among all 87 strains by searching annotated CDSs based on functional annotation pipeline of whole-genome assemblies

CDSs	Non-invasive (N=63)	Invasive (N=24)	*P*
Chitinase	27 (42.9%)	14 (58.3%)	0.293
Exotoxin type A	**13 (20.6%)**	**12 (50.0%)**	**0.015**
Procollagen-proline 3-dioxygenase	**8 (12.7%)**	**8 (33.3%)**	**0.035**
Platelet binding protein GspB	13 (36.1%)	4 (16.7%)	0.771
C protein alpha-antigen	5 (7.9%)	3 (12.5%)	0.679
Glycoprotein-gp2	4 (6.3%)	3 (12.5%)	0.389
N-acetylmuramoyl-L-alanine amidase	1 (1.6%)	0 (0.0%)	1.000
Trehalose transport system permease protein SugB	1 (1.6%)	0 (0.0%)	1.000
Serine-rich adhesin for platelets	1 (1.6%)	0 (0.0%)	1.000
Deoxyribonuclease (Yes/No)	61 (96.8%)	24 (100.0%)	1.000
Hyaluronan.synthase (Yes/No)	49 (77.8%)	22 (91.7%)	0.216
Streptopain (Yes/No)	63 (100.0%)	23 (95.8%)	0.276

The values are presented as N (%). Bold type indicates statistical significance.
Abbreviation: CDS, coding DNA sequence.

DISCUSSION

WGS analyses have proven useful in unraveling the genetic diversity of strains and discriminating between closely related strains. Our study provided information about the genomic characteristics and virulence genes of 87 strains collected in Korea over a 20-year period based on longitudinal analysis of WGS datasets.

Up to 200 *emm* types have been identified, suggesting that the M protein is a polymorphic protein (https://www.cdc.gov/streplab/index.html). A global review of *emm* types revealed a total of 205 *emm* types, including a category of non-typeable strains. The most common *emm* type was *emm*1, which accounted for 18.3% of all strains, followed by *emm*12 (11.1%), *emm*28 (8.5%), *emm*3 (6.9%), and *emm*4 (6.9%) [17]. In Europe, severe clinical manifestations, such as STSS and necrotizing fasciitis, were caused by 45 different types, of which *emm*1 was the most prevalent, accounting for 37% and 31% of cases, respectively [18]. In Korea, *emm*1 was significantly more common among invasive cases, whereas *emm*4, *emm*6, and *emm*12 were dominant in non-invasive cases [19].

Globally, *emm* types influence routine epidemiological surveil-

lance, and MLST is excellent for exotoxin gene profiling [20]. *emm*1 and *emm*3 associated with ST28 have traditionally been associated with invasive *S. pyogenes* strains [19]. In this study, the predominant *emm*1 lineage belonging to ST28 showed a clonal distribution of invasive strains according to the goeBURST results. ST785, a single locus variant of ST28, also belonged to *emm*1. The advantages of the conservative approach used by goeBURST, in which links are shown only between STs that differ at a single locus, have been demonstrated by the analysis of meningococcal CCs using goeBURST, which allowed describing the clonal structures of populations in a quantitative way [21].

By searching for virulence-associated CDSs, we found that four genes, *csn1* (*cas9*), *ispE*, *nisK* (*spaK*), and *citC*, were frequently present in invasive strains. *Cas9* is associated with the clustered regularly interspaced short palindromic repeats (CRISPR) array [22]. The type II-A system of *S. pyogenes* contains four cas genes (*cas9*, *cas1*, *cas2*, and *csn1*) and six CRISPR spacers targeting a phage endopeptidase, superantigen (sepM), methyltransferase, hyaluronidase, hypothetical protein, and an unknown target. *cas9* (previously called *csn1*) and trans-activating CRISPR RNA are essential for all stages of immunity in the type II-A system [23]. In our study, *cas9* was more common in invasive strains. This result indicates the role of *cas9* in *S. pyogenes* pathogenesis and the ability of *S. pyogenes* to counter external stimuli, while playing a direct role in bacterial immunity.

IspE is involved in the isoprenoid (IPP) biosynthesis pathway. IPPs comprise a large, diverse class of naturally occurring organic chemicals essential for cell survival [24]. The IPP pathway is essential for various vital biological functions of bacteria. IPPs are synthesized via the classical mevalonate pathway or the alternative 2C-methyl-D-erythritol 4-phosphate (MEP) pathway. The distribution of the MEP and mevalonate pathways is highly complex, but there is a clear bias towards the former in pathogenic organisms. In our study, *ispE* expression was significantly upregulated in invasive strains, suggesting that the IPP biosynthesis pathway is associated with virulence.

Lactococcal *nisA* is a promoter in the nis cluster that is required for the biosynthesis, immunity, and regulatory systems of *S. pyogenes*; *nisK* and *spaK* also belong to this cluster. The *nisA* promoter is dependent on NisR and NisK, which are important in the survival mechanisms of *S. pyogenes*. The *nisA* promoter allows gene expression modulation in pathogenic streptococci [26]. Bacterial citrate lyase, the key enzyme in citrate fermentation, is encoded by *citC*. Lactic acid bacteria of the genus *Leuconostoc* can produce carbon dioxide and C4 aromatic compounds through lactose heterofermentation and citrate utilization [27]. We confirmed that *nisR* and *citC* are significantly associated with invasive *S. pyogenes*. Their protein products are widely found in *Lactococcus*; therefore, we presume that the genes must have been transmitted via plasmid transfer, allowing efficient control of gene expression by regulatory proteins [28, 29]. The transmitted genes allow *S. pyogenes* to survive in various environments, strengthening its invasiveness [22].

We investigated virulence-associated CDSs among all 87 strains searched from annotated CDSs based on functional annotation of the whole-genome assemblies. The genes encoding exotoxin A and procollagen-proline 3-dioxygenase were frequently present in invasive strains. Streptococcal exotoxin A is encoded by *speA*, which is part of bacteriophage T12 [30]. The presence of *speA* is frequently associated with scarlet fever or rheumatic fever and streptococcal disease [1, 31]. Procollagen-proline 3-dioxygenase catalyzes procollagen L-proline to produce procollagen trans-3-hydroxy-L-proline. This enzyme belongs to the family of oxidoreductases, and its activity has been detected in several strains [32]. A relationship between this enzyme and invasiveness has been rarely observed in *S. pyogenes* [33].

The phylogenetic analysis revealed no significant associations between the superantigen profiles and invasiveness. Moreover, establishing links between longitudinal groups within the phylogenetic tree was difficult. These results indicate the preservation of stable genetic elements over time. The *S. pyogenes* population may have maintained a state of host adaptation by maintaining stable genetic elements over long periods [34].

This study had some limitations. Although our study spanned two decades and was population-based, only 87 strains were included, explaining why we did not observe significant genome changes during the study period. We investigated virulence-associated CDSs among all 87 strains by searching only annotated CDSs based on a functional annotation pipeline of whole-genome assemblies rather than by searching the sequences around specific genes or by PCR simulation. We searched for related articles by entering the search terms "*Streptococcus pyogenes*," "whole genome," or "Korea" into the PubMed database (https://pubmed.ncbi.nlm.nih.gov/). However, there were no hits for related manuscripts as of May 26, 2021. This is probably the first report on WGS datasets of *S. pyogenes* strains from Korea.

In conclusion, this study provided CG characteristics of *S. pyogenes* according to invasiveness over a 20-year period. Genomic dynamics were stable during this time span. Four genes, *csn1*, *ispE*, *nisK*, and *citC*, are candidate virulence-associated CDSs in

https://doi.org/10.3343/alm.2022.42.4.438

host–pathogen interactions of invasive *S. pyogenes* strains. Our results showed considerable agreement with previous epidemiological study results, especially regarding the predominant invasive genotypes, i.e., *emm1*.0, *emm1*.18, *emm1*.3, and *emm1*.76. ST28 showed a clonal distribution of invasive strains. Further epidemiological studies using WGS datasets are needed to better understand and monitor streptococcal virulence.

AUTHOR CONTRIBUTIONS

Kim S, Choi E, and Takahashi T conceptualized the study; Shin H, Choi E, and Lee S collected the data; Shin H, Maeda T, Fukushima Y, Lee S, Kim S, and Takahashi T analyzed the data; Shin H wrote the manuscript; Kim S and Takahashi T reviewed and edited the manuscript; all authors reviewed and approved the manuscript.

CONFLICTS OF INTEREST

None declared.

RESEARCH FUNDING

This work was supported by a grant from the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare (H19C0047), and Bio & Medical Technology Development Program of the National Research Foundation (NRF) (2021M3E5E3080382, 2021R1I1A3044483) by the Korean government. The funders had no role in study design, data collection and interpretation, decision to publish, or preparation of the manuscript.

ORCID

Hyoshim Shin https://orcid.org/0000-0001-9737-2393
Takashi Takahashi https://orcid.org/0000-0003-4131-2062
Seungjun Lee https://orcid.org/0000-0002-3377-4833
Eun Hwa Choi https://orcid.org/0000-0002-5857-0749
Takahiro Maeda https://orcid.org/0000-0003-0899-2860
Yasuto Fukushima https://orcid.org/0000-0003-3284-3056
Sunjoo Kim https://orcid.org/0000-0001-8099-8891

REFERENCES

1. Stevens DL, Tanner MH, Winship J, Swarts R, Ries KM, Schlievert PM, et al. Severe group A streptococcal infections associated with a toxic shock-like syndrome and scarlet fever toxin A. N Engl J Med 1989;321:1-7.
2. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. Lancet Infect Dis 2005;5:685-94.
3. Enright, MC, Spratt BG, Kalia A, Cross JH, Bessen DE. Multilocus sequence typing of *Streptococcus pyogenes* and the relationships between *emm* type and clone. Infect Immun 2001;69:2416-27.
4. Luca AV, Giovanni G, Dezemona P. Pulsed field gel electrophoresis of group A streptococci. Methods Mol Biol. 2015;1301:129-38.
5. Lewis T, Loman NJ, Bingle L, Jumaa P, Weinstock GM, Mortiboy D, et al. High-throughput whole-genome sequencing to dissect the epidemiology of *Acinetobacter baumannii* isolates from a hospital outbreak. J Hosp Infect 2010;75:37-41.
6. Schuchat A, Hilger T, Zell E, Farley MM, Reingold A, Harrison L, et al. Active bacterial core surveillance of the emerging infections program network. Emerg Infect Dis 2001;7:92-9.
7. Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 2010;11:1-11.
8. Chun J, Grim CJ, Hasan NA, Lee JH, Choi SY, Haley BJ, et al. Comparative genomics reveals mechanism for short-term and long-term clonal transitions in pandemic *Vibrio cholerae*. Proc Natl Acad Sci U S A 2009;106:15442-7.
9. Tanizawa Y, Fujisawa T, Nakamura Y. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. Bioinformatics 2018;34:1037-9.
10. Takahashi T, Arai K, Lee DH, Koh EH, Yoshida H, Yano H, et al. Epidemiological study of erythromycin-resistant *Streptococcus pyogenes* from Korea and Japan by *emm* genotyping and multilocus sequence typing. Ann Lab Med 2016;36:9-14.
11. Kim S, Byun JH, Park H, Lee J, Lee HS, Yoshida H, et al. Molecular epidemiological features and antibiotic susceptibility patterns of *Streptococcus dysgalactiae* subsp. *equisimilis* isolates from Korea and Japan. Ann Lab Med 2018;38:212-9.
12. Tamayo E, Montes M, Vicente D, Pérez-Trallero E. *Streptococcus pyogenes* pneumonia in adults: clinical presentation and molecular characterization of isolates 2006-2015. PLoS One 2016;11:e0152640.
13. Sakai T, Taniyama D, Takahashi S, Nakamura M, Takahashi T. Pleural empyema and streptococcal toxic shock syndrome due to *Streptococcus pyogenes* in a healthy Spanish traveler in Japan. IDCases 2017;9:85-8.
14. Takahashi T, Maeda T, Lee S, Lee DH, Kim S. Clonal distribution of clindamycin-resistant erythromycin-susceptible (CRES) *Streptococcus agalactiae* in Korea based on whole genome sequences. Ann Lab Med 2020;40:370-81.
15. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, et al. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 2012;50:1355-61.
16. Nascimento M, Sousa A, Ramirez M, Francisco AP, Carriço JA, Vaz C. PHYLOViZ 2.0: providing scalable data integration and visualization for multiple phylogenetic inference methods. Bioinformatics 2017;33:128-9.
17. Steer AC, Law I, Matatolu L, Beall BW, Carapetis JR. Global *emm* type distribution of group A streptococci: systematic review and implications for vaccine development. Lancet Infect Dis 2009;9:611-6.
18. Luca-Harari B, Darenberg J, Neal S, Siljander T, Strakova L, Tanna A, et al. Clinical and microbiological characteristics of severe *Streptococcus pyogenes* disease in Europe. J Clin Microbiol 2009;47:1155-65.
19. Ekelund K, Darenberg J, Norrby-Teglund A, Hoffmann S, Bang D, Skin-

høj P, et al. Variations in *emm* type among group A streptococcal isolates causing invasive or noninvasive infections in a nationwide study. J Clin Microbiol 2005;43:3101-9.
20. Enright MC, Day NP, Davies CE, Peacock SJ, Spratt BG. Multilocus sequence typing for characterization of methicillin-resistant and methicillin-susceptible clones of *Staphylococcus aureus*. J Clin Microbiol 2000; 38:1008-15.
21. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 2004; 186:1518-30.
22. Le Rhun A, Escalera-Maurer A, Bratovič M, Charpentier E. CRISPR-Cas in *Streptococcus pyogenes*. RNA Biol 2019;16:380-9.
23. Nozawa T, Furukawa N, Aikawa C, Watanabe T, Haobam B, Kurokawa K, et al. CRISPR inhibition of prophage acquisition in *Streptococcus pyogenes*. PLoS One 2011;6:e19543.
24. Heuston S, Begley M, Gahan CGM, Hill C. Isoprenoid biosynthesis in bacterial pathogens. Microbiology (Reading) 2012;158:1389-401.
25. Voynova NE, Rios SE, Miziorko HM. *Staphylococcus aureus* mevalonate kinase: isolation and characterization of an enzyme of the isoprenoid biosynthetic pathway. J Bacteriol 2004;186:61-7.
26. Bekal S, Van Beeumen J, Samyn B, Garmyn D, Henini S, Diviès C, et al. Purification of *Leuconostoc mesenteroides* citrate lyase and cloning and characterization of the citCDEFG gene cluster. J Bacteriol 1998; 180:647-54.
27. Kawada-Matsuo M, Tatsuno I, Arii K, Zendo T, Oogai Y, Noguchi K, et al. Two-component systems involved in susceptibility to nisin A in *Streptococcus pyogenes*. Appl Environ Microbiol 2016;82:5930-9.
28. Quadri LE. Regulation of antimicrobial peptide production by autoinducer-mediated quorum sensing in lactic acid bacteria. Antonie Van Leeuwenhoek 2002;82:133-45.
29. Hu J, Jin K, He ZG, Zhang H. Citrate lyase CitE in *Mycobacterium tuberculosis* contributes to mycobacterial survival under hypoxic conditions. PLoS One 2020;15:e0230786.
30. Weeks CR and Ferretti JJ. Nucleotide sequence of the type A streptococcal exotoxin (erythrogenic toxin) gene from *Streptococcus pyogenes* bacteriophage T12. Infect Immun 1986;52:144-50.
31. Yu CE and Ferretti JJ. Molecular epidemiologic analysis of the type A streptococcal exotoxin (erythrogenic toxin) gene (*speA*) in clinical *Streptococcus pyogenes* strains. Infect Immun 1989;57:3715-9.
32. Shibasaki T, Mori H, Chiba S, Ozaki A. Microbial proline 4-hydroxylase screening and gene cloning. Appl Environ Microbiol 1999;65:4028-31.
33. Mori H, Shibasaki T, Uozaki Y, Ochiai K, Ozaki A. Detection of novel proline 3-hydroxylase activities in *Streptomyces* and *Bacillus* spp. by regio- and streospecific hydroxylation of L-proline. Appl Environ Microbiol 1996;62:1903-7.
34. Park HJ, Gokhale CS, Bertels F. How sequence populations persist inside bacterial genomes. Genetics 2021;217:iyab027.

https://doi.org/10.3343/alm.2022.42.4.438

Supplemental Data Table S1. Isolation year, sample type, *emm* type/cluster, ST, superantigen profile, and GenBank accession number of the 87 *Streptococcus* pyogenes strains

Strain	Isolation year	Sample	*emm* genotype	*emm* cluster	ST	Superantigen profile	Accession number
GCH79	1997	Throat swab	1.76	A-C3	28	SmeZ-SpeB	WSTD00000000
GCH80	1997	Throat swab	1.76	A-C3	28	SmeZ-SpeB-SpeC	WSTE00000000
GCH81	1997	Throat swab	3.10	A-C5	1261	SmeZ-SpeB-ssa	WTFP00000000
GCH82	1997	Throat swab	4.00	E1	39	SmeZ-SpeB-SpeC-ssa	WTFQ00000000
GCH83	1997	Throat swab	4.00	E1	38	SmeZ-SpeB-SpeC-ssa	WWFH00000000
GCH84	1997	Throat swab	12.00	A-C4	36	SmeZ-SpeB-SpeC	WWFI00000000
GCH85	1997	Throat swab	12.00	A-C4	36	SmeZ-SpeB-SpeC	WWFJ00000000
GCH86	1997	Throat swab	12.00	A-C4	36	SmeZ-SpeB-SpeC	WWFK00000000
GCH87	2006	Throat swab	1.00	A-C3	28	SmeZ-SpeB	WWFL00000000
GCH88	2006	Throat swab	1.00	A-C3	28	SmeZ-SpeB	WWFM00000000
GCH89	2006	Throat swab	3.10	A-C5	15	SmeZ-SpeB-ssa	WWFN00000000
GCH90	2006	Throat swab	3.10	A-C5	15	SmeZ-SpeB	WWFO00000000
GCH91	2006	Throat swab	3.10	A-C5	15	SmeZ-SpeB-ssa	JAAAMW000000000
GCH92	2006	Throat swab	4.00	E1	39	SmeZ-SpeB-SpeC-ssa	JAAAMX000000000
GCH93	2006	Throat swab	4.00	E1	39	SmeZ-SpeB-SpeC-ssa	JAAAMY000000000
GCH94	2006	Throat swab	1.00	A-C3	28	SmeZ-SpeB	JAAAMZ000000000
GCH95	2006	Throat swab	11.00	E6	403	SmeZ-SpeB-SpeC	JAAANA000000000
GCH96	2006	Throat swab	28.00	E4	52	SmeZ-SpeB-SpeC	JAAANB000000000
GCH97	2017	Throat swab	1.30	A-C3	28	SmeZ-SpeB	JAAANC000000000
GCH98	2017	Throat swab	1.18	A-C3	28	SmeZ-SpeB	JAAAND000000000
GCH100	2017	Throat swab	4.00	E1	39	SmeZ-SpeB-SpeC-ssa	JAAANE000000000
GCH101	2017	Throat swab	4.00	E1	39	SmeZ-SpeB-SpeC-ssa	JAAANF000000000
GCH102	2017	Throat swab	4.00	E1	39	SmeZ-SpeB-SpeC-ssa	JAAANG000000000
GCH103	2017	Throat swab	12.00	A-C4	36	SmeZ-SpeB	JAAANH000000000
GCH104	2017	Throat swab	12.00	A-C4	36	SmeZ-SpeB	JAAANI000000000
GCH105	2017	Throat swab	12.00	A-C4	36	SmeZ-SpeB	JAAANJ000000000
GCH106	2017	Throat swab	28.00	E4	52	SmeZ-SpeB-SpeC	JAAANK000000000
GCH107	2017	Throat swab	4.00	E1	39	SmeZ-SpeB-SpeC-ssa	JAAANL000000000
GCH108	2017	Throat swab	12.00	A-C4	36	SmeZ-SpeB	JAAANM000000000
GCH109	**1997**	**Blood**	**94.00**	**E6**	**399**	**SmeZ-SpeB**	**JAAANN000000000**
GCH110	**1997**	**Blood**	**49.00**	**E3**	**1254**	**SpeB**	**JAAANO000000000**
GCH111	**2002**	**Blood**	**3.10**	**A-C5**	**15**	**SmeZ-SpeB-ssa**	**JAAANP000000000**
GCH112	**2004**	**Blood**	**3.10**	**A-C5**	**15**	**SmeZ-SpeB-ssa**	**JAAANQ000000000**
GCH113	**2010**	**Joint fluid**	**12.49**	**A-C4**	**36**	**SmeZ-SpeB**	**JAAANR000000000**
GCH114	**2011**	**Blood**	**1.00**	**A-C3**	**28**	**SmeZ-SpeB**	**JAAANS000000000**
GCH115	**2011**	**Joint fluid**	**3.10**	**A-C5**	**15**	**SmeZ-SpeB-ssa**	**JAAANT000000000**
GCH116	**2014**	**Blood**	**4.00**	**E1**	**39**	**SmeZ-SpeB-SpeC-ssa**	**JAAANU000000000**
GCH117	**2016**	**Blood**	**1.30**	**A-C3**	**28**	**SmeZ-SpeB**	**JAAANV000000000**
GCH118	**2017**	**Joint fluid**	**12.00**	**A-C4**	**36**	**SmeZ-SpeB**	**JAAANW000000000**
GCH128	**2011**	**Blood**	**12.00**	**A-C4**	**36**	**SmeZ-SpeB**	**JABUOL000000000**

(*Continued to the next page*)

Supplemental Data Table S1. Continued

Strain	Isolation year	Sample	*emm* genotype	*emm* cluster	ST	Superantigen profile	Accession number
GCH129	**2012**	**Blood**	**12.00**	**A-C4**	**36**	**SmeZ-SpeB**	**JABLSL000000000**
GCH130	**2012**	**Blood**	**1.00**	**A-C3**	**28**	**SmeZ-SpeB**	**JABUOM000000000**
GCH131	**2012**	**Blood**	**1.00**	**A-C3**	**28**	**SmeZ-SpeB**	**JABUON000000000**
GCH132	**2013**	**Blood**	**75.00**	**E6**	**49**	**SmeZ-SpeB-SpeC**	**JABUOO000000000**
GCH133	**2015**	**Blood**	**1.00**	**A-C3**	**785**	**SmeZ-SpeB**	**JABUOP000000000**
GCH134	**2015**	**Blood**	**1.00**	**A-C3**	**28**	**SmeZ-SpeB**	**JABUOQ000000000**
GCH135	**2016**	**Blood**	**28.00**	**E4**	**52**	**SmeZ-SpeB-SpeC**	**JABUJB000000000**
GCH136	**2016**	**Blood**	**1.00**	**A-C3**	**28**	**SmeZ-SpeB**	**JABUJC000000000**
GCH137	**2017**	**Blood**	**3.10**	**A-C5**	**15**	**SmeZ-SpeB-ssa**	**JABUJD000000000**
GCH138	**2017**	**Blood**	**4.00**	**E1**	**39**	**SmeZ-SpeB-SpeC-ssa**	**JABUJE000000000**
GCH139	**2017**	**Blood**	**28.00**	**E4**	**52**	**SmeZ-SpeB-SpeC**	**JABUJF000000000**
GCH140	**2018**	**Blood**	**1.00**	**A-C3**	**28**	**SmeZ-SpeB**	**JABUJG000000000**
GCH141	**2018**	**Blood**	**28.00**	**E4**	**52**	**SmeZ-SpeB-SpeC**	**JABUJH000000000**
GCH142	2006	Throat swab	12.00	A-C4	36	SmeZ-SpeB	JABUJI000000000
GCH143	2006	Throat swab	78.30	E1	253	SmeZ-SpeB-SpeC	JABUJJ000000000
GCH144	2006	Throat swab	6.90	Single protein M cluster clade Y	37	SmeZ-SpeB	JABUJK000000000
GCH145	2006	Throat swab	87.10	E3	62	SmeZ-SpeB	JABUJL000000000
GCH146	2006	Throat swab	114.60	E4	1,108	SmeZ-SpeB	JABUJM000000000
GCH147	2006	Throat swab	22.00	E4	46	SmeZ-SpeB-SpeC-ssa	JABUJN000000000
GCH148	2006	Throat swab	5.14	Single protein M cluster clade Y	99	SmeZ-SpeB-SpeC-ssa	JABUJO000000000
GCH149	2006	Throat swab	11.00	E6	403	SmeZ-SpeB	JABUJP000000000
GCH150	2006	Throat swab	22.00	E4	46	SmeZ-SpeB-SpeC-ssa	JABUJQ000000000
GCH151	2005	Throat swab	78.30	E1	253	SmeZ-SpeB-SpeC	JABUJR000000000
GCH152	2005	Throat swab	22.00	E4	46	SmeZ-SpeB-SpeC-ssa	JABUJS000000000
GCH153	2005	Throat swab	78.30	E1	253	SmeZ-SpeB-SpeC	JABUJT000000000
GCH154	2005	Throat swab	78.30	E1	253	SmeZ-SpeB-SpeC	JABUJU000000000
GCH155	2005	Throat swab	94.00	E6	89	SmeZ-SpeB-SpeC	JABUJV000000000
GCH156	2005	Throat swab	94.00	E6	89	SmeZ-SpeB-SpeC	JABUAZ000000000
GCH157	2005	Throat swab	5.14	Single protein M cluster clade Y	99	SmeZ-SpeB-SpeC-ssa	JABUBA000000000
GCH158	2005	Throat swab	5.14	Single protein M cluster clade Y	99	SmeZ-SpeB-SpeC-ssa	JABUBB000000000
GCH159	2005	Throat swab	78.30	E1	253	SmeZ-SpeB-SpeC	JABUAT000000000
GCH160	2006	Throat swab	22.00	E4	46	SmeZ-SpeB-SpeC-ssa	JABUAP000000000
GCH161	2006	Throat swab	22.00	E4	46	SmeZ-SpeB-SpeC-ssa	JABUAQ000000000
GCH163	2006	Throat swab	44.00	E3	25	SmeZ-SpeB-ssa	JABUBC000000000
GCH164	1997	Throat swab	12.00	A-C4	36	SmeZ-SpeB-SpeC	JABUAU000000000
GCH165	1997	Throat swab	12.00	A-C4	36	SmeZ-SpeB-SpeC	JABUAV000000000
GCH166	1997	Throat swab	12.19	A-C4	36	SmeZ-SpeB-SpeC	JABUAW000000000
GCH167	1997	Throat swab	12.00	A-C4	36	SmeZ-SpeB-SpeC	JABUAX000000000

(*Continued to the next page*)

https://doi.org/10.3343/alm.2022.42.4.438

Supplemental Data Table S1. Continued

Strain	Isolation year	Sample	*emm* genotype	*emm* cluster	ST	Superantigen profile	Accession number
GCH168	1997	Throat swab	4.00	E1	38	SmeZ-SpeB-SpeC-ssa	JABUAY000000000
GCH169	1997	Throat swab	1.30	A-C3	28	SmeZ-SpeB	JABUAR000000000
GCH170	2006	Throat swab	12.00	A-C4	36	SmeZ-SpeB	JABUAS000000000
GCH171	2006	Throat swab	12.00	A-C4	36	SmeZ-SpeB	JABUAJ000000000
GCH172	2017	Throat swab	28.00	E4	458	SmeZ-SpeB-SpeC	JABUAK000000000
GCH173	2017	Throat swab	28.00	E4	52	SmeZ-SpeB-SpeC	JABUAL000000000
GCH174	2017	Throat swab	1.00	A-C3	28	SmeZ-SpeB	JABUAM000000000
GCH175	2017	Throat swab	28.00	E4	52	SmeZ-SpeB-SpeC	JABUAN000000000
GCH178	2017	Throat swab	4.00	E1	39	SmeZ-SpeB-SpeC-ssa	JABUAO000000000

Bold type indicates invasiveness.
Abbreviation: ST, sequence type.