WHOLE GENOME SEQUENCE-BASED CHARACTERIZATION OF GROUP A STREPTOCOCCI (GAS; *STREPTOCOCCUS PYOGENES*) ISOLATES CAUSING INVASIVE DISEASE IN SOUTH AFRICA, 2019 - 2020

A Dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Medicine
(Clinical Microbiology and Infectious Diseases)

Submitted by

Sibusisiwe Thobile Zulu
(1304805)
Johannesburg, 2024

DECLARATION

I, Sibusisiwe Thobile Zulu, declare that this dissertation is my own work. Experiments described here were conducted under the supervision of Professor Anne von Gottberg and Dr. Sibongile Walaza at the Centre for Respiratory Diseases and Meningitis, National Institute for Communicable Diseases, a division of the National Health Laboratory Service, Johannesburg. It is being submitted for the degree of Master of Science in Medicine to the Faculty of Health Sciences at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination to this or any other university.

_____

Sibusisiwe Thobile Zulu

Date: 30 October 2024

FOR ME

"'I want to thank me for believing in me, I want to thank me for doing all this hard work. I want to thank me for having no days off. I want to thank me for never quitting. I want to thank me for always being a giver and trying to give more than I receive. I want to thank me for trying to do more right than wrong. I want to thank me for being me at all times"- Calvin Cordozar "Snoop dogg" Broadus Jr.

PRESENTATIONS

Zulu S. Laboratory-based surveillance for group A streptococci (GAS; *Streptococcus pyogenes*) isolates causing invasive disease in South Africa, 2019-2020. GERMS-SA principal investigators meeting. Virtual. 15 June 2021. (Oral)

Zulu S. Laboratory-based surveillance of group A streptococcus (GAS) and group B streptococcus (GBS) isolates causing invasive disease in South Africa. GERMS-SA surveillance officers meeting. Birchwood Hotel, Boksburg, Johannesburg, Gauteng, South Africa, 02-05 August 2022. (Oral)

Zulu S. Whole genome sequence-based characterization of invasive group A streptococci (GAS, *Streptococcus pyogenes*) isolates causing invasive disease in South Africa. NICD 2022 Annual Influenza and Respiratory Disease Surveillance Symposium (AIRSS). Poliomyelitis Research Foundation Auditorium, NICD, Johannesburg, Gauteng, South Africa, 17-18 August 2022. (Oral)

Zulu S. *Streptococcus pyogenes* (Group A streptococcus; GAS) isolates causing invasive disease in South Africa. NICD Scientific Research Forum. Virtual. 29 November 2023. (Oral)

ABSTRACT

Invasive group A streptococcus (iGAS) is a major cause of morbidity. Penicillin and erythromycin are recommended for treatment. A 30-valent M-protein-based vaccine is being developed. We describe the molecular epidemiology of iGAS disease in South Africa, 2019 - 2020.

iGAS episodes (cultured from normally sterile sites or from necrotising fasciitis wounds, among all ages) were detected through active, national, laboratory-based surveillance. At the national reference laboratory, isolates were confirmed phenotypically based on colony morphology, β-haemolysis and sensitivity to bacitracin; antimicrobial susceptibility testing was done by disk diffusion and broth microdilution; and isolates were further characterized by whole-genome sequencing. A SNP-based phylogenetic analysis determined genetic clusters.

Of 1487 iGAS cases that were reported, 618 were characterised phenotypically (viable isolates available) and 577 genotypically. Incidence was 1.7 cases/100,000 persons in 2019 and 0.9/100,000 in 2020, with peaks in the extremes of age. All isolates were susceptible to the β-lactam antibiotics. Resistance was detected against tetracycline (10.5%;65/618), erythromycin (3.6%; 22/618), chloramphenicol (0.2%;1/618) and levofloxacin (0.2%;1/618). *tetM* and *mef* genes were present among isolates resistant to tetracycline and erythromycin, respectively. Overall, 56 *emm* types were identified. *emm*76 (33/577,5.7%), *emm*92 (30/577,5.2%) and *emm*82 (29/577,5.0%) were the most prevalent. Globally, *emm*1 is the most frequently occurring *emm* type in high-income countries, yet it was only detected in 3.7% of the isolates in the present study. *emm*76 and *emm*95 types accounted for most of the tetracycline- and erythromycin-resistant isolates, respectively. *emm* types in the 30-valent M-protein-based vaccine accounted for 44.6% (244/547) of isolates. Phylogenetically, identical *emm* types clustered together. Among genetic clusters, there was no geographical and temporal aggregation of cases.

iGAS treatment with β- lactams remains appropriate. The 30-valent M-protein-based vaccine offers coverage to less than half of the isolates. No unidentified outbreaks were detected.

ACKNOWLEDGEMENTS

Glory be to GOD! Thank you Father for giving me the strength and a sound mind to complete my Masters. When I first started on this journey, I did not think I was capable, but You have yet again shown me that “I can do all things through Christ who strengthens me’’ Philippians 4:13. None of this would have been possible had You not walked ahead of me and paved my way.

Academic

Prof. Anne von Gottberg and Dr. Sibongile Walaza – Thank you for your patience throughout this journey. Thank you for all the time you took to review my work and answer all my questions. The gentle and affirming manner in which you both supervised my work has taught me to trust myself and the work that I present. Often I felt as though I was moving at a much slower pace than all my other colleagues, but you both always re-affirmed that this was my race and I had to work at a pace that was manageable so that I could still enjoy my research. I truly appreciated that.

Sopio Chochua, Benjamin Metcalf and Yuan Li – Thank you all for your assistance with setting up the pipeline used for my analysis. This work would have not been possible without you.

Prof. Scott Hazelhurst – Thank you for spending months assisting us with setting up the pipeline on the Wits server. This work would have not been possible without your valuable contribution.

Buhle Ntozini, Rito Mikhari, Nosihle Msomi and Cebile Lekhuleni – I can’t think of a better bunch to have been on this journey with. It didn’t matter how stressed we were – or if we had just cried in the office ☺ - we always found time to joke and laugh together. Thank you for the late nights and weekends we all worked together. Sharing an office with such bright and hardworking individuals pushed me to also work harder.

Nthabeleng Makakole and Keabetswe Motsepe – Thank you both for being wonderful colleagues. Balancing full time employment and school was not easy at all

and I thank you both for the times you covered for me while I worked on my dissertation.

Scientists, Technologists and Technicians at CRDM – Thank you all for your contribution.

Linda de Gouveia – Thank you for all the database drawdowns and answering all the questions I had about the database and GERM-SA study.

GERMS-SA Team – This work would not have been possible without all the beautiful data each individual helped collect. It may not seem like it but the work that you do makes a big difference in public health.

Personal

My parents – Thank you both for all the hard work and sacrifices you have made to ensure that I have a bright future. You weren't completely sure what exactly I was doing, but you always made it a point to ask me how I was coping.

My daughter – My sweet beautiful baby ☺ You were a beautiful surprise halfway through this journey. The final year of my masters was challenging. Having to juggle work, school and being a new mom was not easy but you remained my top priority throughout. I have and will always put you first. You gave me a needed extra push to finish off this race so that I could spend all of my time after work with you ☺. Thank you for your beautiful smiles that always cheered me up regardless of how challenging the day was. Mommy loves you so much!

Bruce Ndlovu – Brucey! You have been there from the start of this journey. Your idea of supporting me was saying you would stay up with me while I work but then 5 minutes later I would hear you snoring on the couch ☺. Thank you for always encouraging me with your words and re-affirming that I could do this.

Research Funding

This work was supported in part by a Fogarty International Center Global Infectious Disease research training grant, National Institutes of Health, to the University of

Pittsburgh and National Institute for Communicable Diseases (D43TW011255) (PI's: Dr Lee Harrison and Prof Anne von Gottberg).

TABLE OF CONTENTS

FIGURES

TABLES

ABBREVIATIONS AND SYMBOLS

ABCs	Active Bacterial Core surveillance
AMR	Antimicrobial resistance
APSGN	Acute post streptococcal glomerulonephritis
ARF	Acute rheumatic fever
AST	Antimicrobial susceptibility testing
BHI	Brain heart infusion
BLT	Bead-linked transposome
BP	Base pair
CC	Clonal complex
CDC	Centers for Disease Control and Prevention
CLSI	Clinical and Laboratory Standards Institute
CNS	Central nervous system
COVID-19	Coronavirus Disease 2019
CO_2	Carbon dioxide
CRDM	Centre for Respiratory Diseases and Meningitis
CRF	Case report form
DoH	Department of Health
DNA	Deoxyribonucleic acid
ECM	Extracellular matrix
ERM	Erythromycin ribosome methylase
FbaA	Fibronectin binding protein of group A streptococci type A
FbaB	Fibronectin binding protein of group A streptococci type B
Fbp54	Fibronectin binding protein 54
GAPDH	Glyceraldehyde-3- phosphate dehydrogenase
GAS	Group A streptococcus
GBS	Group B streptococcus
GERMS-SA	Group for Enteric, Respiratory and Meningeal Disease Surveillance in South Africa
GTR	Generalized Time Reversible
HIV	Human Immunodeficiency Virus
HPCC	High performance computing cluster

HREC	Human Research Ethics Committee
iGAS	Invasive group A streptococcus
iTOL	Interactive Tree of Life
Lbp	Laminin binding protein
LIS	Laboratory information system
lnu	Lincosamide nucleotidyltransferase
LRTI	Lower respiratory tract infection
LTA	Lipoteichoic acid
MALDI-TOF	Matrix- Assisted Laser Desorption Ionization-Time Of Fight
MEF	Macrolide efflux
MIC	Minimum inhibitory concentration
ML	Maximum-likelihood
MLST	Multilocus sequence typing
Mrp	M related protein
NaOH	Sodium hydroxide
NHLS	National Health Laboratory Services
NICD	National Institute for Communicable Diseases
NMC	Notifiable medical conditions
NML	National Microbiology Laboratory
NPI	Non-pharmaceutical interventions
NT	Non-typable
O_2	Oxygen
Pbp2x	Penicillin-binding protein 2x
PCR	Polymerase chain reaction
PFBP	*S. pyogenes* fibronectin binding protein
PFGE	Pulse-field gel electrophoresis
PHAC	Public Health Agency of Canada
PHE	Public Health England
PI	Principal investigator
POCT	Point-of-care test
PrtF1	Protein F1
PrtF2	Protein F2
QC	Quality check

qPCR	Real-time polymerase chain reaction
RADT	Rapid antigen detection test
RHD	Rheumatic heart disease
SAg	Streptococcal antigen gene
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
Scl1	Streptococcal collagen-like surface protein 1
Scl2	Streptococcal collagen-like surface protein 2
SDA1	Streptodornase D
SfbI	Streptococcal fibronectin binding protein I
SfbII	*S. pyogenes* fibronectin binding protein II
SfbX	*S. pyogenes* fibronectin binding protein X
Shr	Streptococcal haemoprotein receptor
SIC	Streptococcal inhibitor of complement
SLO	Streptolysin O
SLS	Streptolysin S
SmeZ	Streptococcal mitogenic exotoxin Z
SMHA	Mueller Hinton with 5% sheep blood agar
SNP	Single nucleotide polymorphism
SOF	Serum opacity factor
Spa	Streptococcal protective agent
Spe	Streptococcal pyrogenic exotoxin
Ssa	Streptococcal superantigen A
SSTI	Skin and soft tissue infection
ST	Sequence type
STSS	Streptococcal toxic shock syndrome
TAT	Turnaround time
TB1	Tagmentation buffer
TET	Tetracycline
TSB	Tagment stop buffer
TWB	Tagment wash buffer
UK	United Kingdom
UKHSA	United Kingdom Health Security Agency
URTI	Upper respiratory tract infection

US	United States
UTI	Urinary tract infection

1 INTRODUCTION/LITERATURE REVIEW

1.1 *Streptococcus pyogenes* (group A streptococcus)

1.1.1 History of *Streptococcus pyogenes* (*S. pyogenes*)

Streptococci were first observed in 1874 by Billbroth in patients presenting with erysipelas and wound infections (1). Five years after this, Pasteur observed the same pathogen in the blood of patients presenting with puerperal sepsis and in 1883, Fehleisen cultured streptococci from erysipelas lesions (1). In 1884, Rosenbach named this organism *Streptococcus pyogenes*, which translates to 'pus-forming chained round-shaped bacteria' (1).

1.1.2 Bacteriology of *S. pyogenes*

Streptococcus pyogenes (*S. pyogenes*) are classified as group A streptococcus (GAS) based on the Lancefield grouping system (2). The Lancefield grouping system was created by Rebecca Lancefield and it classifies catalase-negative Gram-positive cocci based on the antigen of the bacterial cell wall (2). There are currently 20 groups, designated group A-V (excluding I and J) (3). GAS grow in pairs as spherical or ovoid cells or as small to moderate sized chains in clinical specimens (1). When grown on blood agar, GAS appear as white to grey colonies surrounded by zones of complete β- haemolysis (**Figure 1**) (1). The haemolysis is due to the presence of the virulence factor streptolysin S, present in most GAS strains (1). GAS exclusively colonize humans as their host (4).

Figure 1: Growth of *Streptococcus pyogenes* on blood agar showing β-hemolytic colonies and inhibition by bacitracin disc after 24-hour incubation

1.1.3 Pathogenesis of *S. pyogenes*

The primary site of GAS colonization are epithelial surfaces, mainly the throat and skin (5). The throat and skin are the main sites where GAS are maintained and are the main source of GAS transmission (4). Transmission occurs from person-to-person via respiratory droplets from an infected person or through contact with contaminated skin (6).

Colonization by GAS is a precursor for developing disease and various GAS surface structures are involved in host colonization (4,5). The initial attachment of GAS to human host cells and extracellular matrix (ECM) is thought to be a process that occurs in two stages, namely (i) weak/long-range binding followed by (ii) strong high-affinity binding (7). A membrane-bound amphiphilic polymer of glycerol phosphate containing glucose and D-alamine termed lipoteichoic acid (LTA) is hypothesized to play a role in facilitating the first stage of GAS attachment to the host by establishing weak hydrophobic bonds between bacterial cells and host components (7). This weak interaction is a precursor step that allows longer-distance first-attachment events to occur with the assistance of GAS long-surface appendages such as pili (4). This is then followed by the second stage of adherence that involves numerous high affinity binding events between GAS and host components (4). These high affinity

binding events include GAS surface protein-to-host-cell-protein binding or lectin-to-carbohydrate binding events (4). There are numerous GAS surface proteins that enable GAS to interact with multiple host components, thus allowing GAS to establish themselves in various tissue sites in the human host (4). These GAS surface proteins include fibronectin/laminin binding proteins such as PrtF1/SfbI (Protein F1/Streptococcal fibronectin binding protein I), PrtF2/PFBP (Protein F2/*S. pyogenes* fibronectin binding protein), FbaB (Fibronectin binding protein of group A streptococci type B), SOF/Sfb II (Serum opacity factor/*S. pyogenes* fibronectin binding protein II), SfbX (*S. pyogenes* fibronectin binding protein X), Fbp54 (Fibronectin binding protein 54), FbaA (Fibronectin binding protein of group A streptococci type A), Scl1 (Streptococcal collagen-like surface protein 1), Scl2 (Streptococcal collagen-like surface protein 2), Lbp (laminin binding protein), Shr (Streptococcal hemoprotein receptor) and GAPDH (Glyceraldehyde-3-phosphate dehydrogenase) (4,8).

1.1.4 Virulence of *S. pyogenes*

Once GAS has colonized its host, it may cause non-invasive disease or penetrate deeper into epithelial barriers in order to spread into blood vessels and deeper organs, thus causing invasive disease. For GAS to cause disease, it needs to evade the hosts innate immune system (9). GAS have a plethora of surfaced-expressed and secreted virulence factors that help it evade the host immune system (9,10). In addition to their role in host colonization, the surface proteins PrtF1/SfbI, FbaA, Scl1 and GAPDH assist GAS in evading the host's innate immune system by inhibiting the ability of the host antibodies and phagocytic cells from removing microbes and damaged cells from the host i.e. Inhibiting complement activity at the site of infection (8).

Most GAS strains possess a hyaluronic acid capsule which is an accessory virulence factor (1). The hyaluronic acid capsule shields GAS from phagocytosis by polymorphonuclear leukocytes and macrophages of the human host (1). When cultured on blood agar, GAS strains that produce high amounts of the hyaluronic acid capsule have a mucoid appearance (1).

The M-protein, a surface protein and an important virulence determinant of GAS, was first described in 1927 by Rebecca Lancefield who demonstrated that the immune response was targeted to this antigen (9,11). The M-protein presents itself as a long α-helical coiled coil structure that is anchored in the bacterial membrane and extends from the surface of the cell (**Figure 2**) (9). The M-protein consists of a C-terminal region that is highly conserved and a surface exposed N-terminal region that is hypervariable and associated with serotype identity (4). The serotypes are referred to as *emm* types (4). This feature offers protection through diversity as novel *emm* types (i.e. *emm* types that have not previously colonized the host) evade antibodies raised by the host in response to a previous infection (4,11). The M-protein also protects GAS from opsonization and phagocytosis by polymorphonuclear leukocytes of the host and strains that lack the M-protein have been shown to be avirulent (1). The M-protein has also been shown to induce thrombosis in the host through coagulation activation (12–14). Some M-proteins are capable of binding to human fibrinogen leading to platelet activation which can result in the activation of neutrophils and monocytes, thus accelerating inflammation in the host (14).

In addition to the surface-expressed virulence factors discussed above, GAS produce various secreted virulence factors including two types of hemolysins, namely streptolysin O (SLO) and streptolysin S (SLS) (1). Most GAS strains produce both hemolysins and they are both toxic to numerous host cells and cell fractions (1). SLS is responsible for the β- hemolysis of GAS observed when cultured on blood agar (10). SLS is produced only by GAS strains grown in the presence of serum, serum albumin, α-lipoprotein, ribonucleic acid or detergents (1). Antibodies raised against SLO are used in serology testing as an indicator of a recent streptococcal infection (15).

Some GAS strains produce streptococcal superantigen exotoxins (10). The term ‘superantigen’ was coined to describe the strong primary T cell response these exotoxins illicit (10). The excessive T cell response results in the release of pro-inflammatory cytokines which can lead to toxic shock syndrome (10). These exotoxins were historically known as erythrogenic or scarlet fever toxins because of the red rash they cause in patients presenting with scarlet fever (10). The term streptococcal pyrogenic exotoxin (Spe) was later proposed (16). There are currently 14 *S. pyogenes* exotoxins that have been described and they are Spe_A, Spe_C,

Spe_G, Spe_H, Spe_I, Spe_J, Spe_K, Spe_L, Spe_M, Spe_N, Spe_O and Spe_P as well as the streptococcal superantigen SS_A and the streptococcal mitogenic exotoxin Z (Sme_Z) (10). Different GAS strains will encode different combinations of 3 - 6 streptococcal pyrogenic exotoxin genes (10).

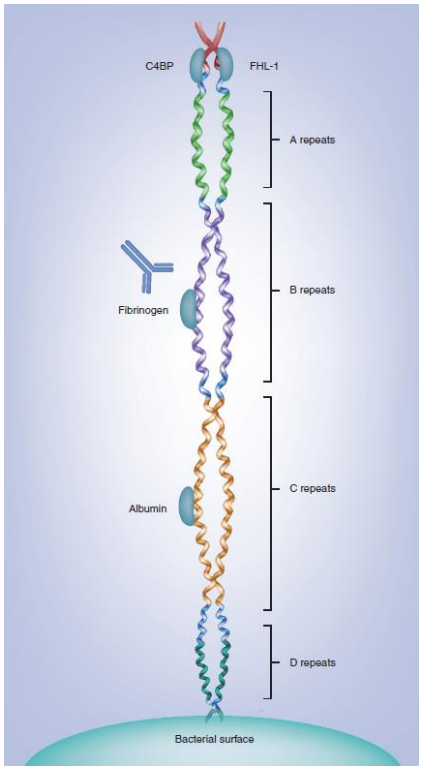


Figure 2: The M-protein structure of group A streptococcus. Adapted from (17).

1.2 Clinical manifestation of GAS disease

GAS can be asymptomatically carried in the throat, which is more common in children aged 5-15 years (18). GAS can cause disease and are responsible for various non-invasive diseases (e.g., pharyngitis, scarlet fever and superficial skin infections) and are also capable of penetrating epithelial barriers and mucous membranes, thus causing more serious invasive disease (e.g., bacteremia, cellulitis, puerperal sepsis, pneumonia, septic shock syndrome, soft-tissue infections and necrotizing fasciitis) (5,18,19). GAS infections that are not treated properly may trigger post-streptococcal autoimmune sequelae, such as acute rheumatic fever (ARF) (20) and acute post streptococcal glomerulonephritis (APSGN) (21). Repeated episodes of ARF may ultimately lead to rheumatic heart disease (RHD) (20,22).

1.2.1 Non-invasive GAS disease

1.2.1.1 Pharyngitis

Streptococcal pharyngitis is an infection of the pharynx and is the most common illness caused by GAS, with an estimated 600 million cases occurring annually (19,21). Transmission, which occurs from person-to-person via the respiratory route (4), is favored by crowded environments such as military camps and schools (1). GAS causes 20-30% of all pharyngitis cases in children and 5-15% of cases in adults (21). The most vulnerable group are school-aged children aged 5-15 years (21) and peak incidence is seen during the early years of school (1). The most common symptoms of pharyngitis include fever, sore throat, headache, nausea, abdominal pain and malaise (21). Uncomplicated GAS pharyngitis is often self-limiting and symptoms normally resolve within a week (21). Suppurative (e.g., otitis media, sinusitis, bacteremia), non suppurative (e.g., ARF) or toxin mediated (e.g., streptococcal toxic shock syndrome) complications may arise during or after a GAS pharyngitis episode resulting in more severe disease (21).

1.2.1.2 Superficial skin infections

GAS can cause various skin infections including impetigo and pyoderma (21). Impetigo is the most common skin infection caused by GAS, with an estimated annual global disease burden of 111 million cases (19). Individuals with impetigo present with reddish sores normally around the nose, mouth, hands and feet (21). A few days following initial appearance of the sores, these sores burst and form honey-colored crusts (21).Transmission occurs through contact with infected skin and most commonly affects young children in low-income countries, residing in crowded areas that experience tropical and subtropical climates (4,21).

1.2.1.3 Scarlet fever

Scarlet fever manifests as a result of an infection with a GAS strain that produces streptococcal pyrogenic exotoxins (Spe) (1). Scarlet fever is recognized by a distinct rash (scarlatinal rash) (1). At first, the rash appears over the upper part of the chest and spreads lower to the trunk and higher to the neck before moving to the extremities (1). Additional distinct scarlet fever symptoms include (i) scattered petechiae on the body; (ii) covering of the tongue by a white coat with red papillae, giving it a white strawberry tongue appearance and (iii) a beefy red tongue (red strawberry tongue) that appears when the white coat disappears (1).

1.2.2 Invasive GAS disease

It is estimated that globally, approximately 163 000 deaths occur annually as a result of invasive GAS (iGAS) disease (19). Invasive disease caused by GAS includes severe skin and soft tissue infections, pneumonia, meningitis, bacteremia, peritonitis and streptococcal toxic shock syndrome.

1.2.2.1 Severe skin and soft tissue infections and streptococcal toxic shock syndrome

The most common cause of severe skin and soft tissue infections (SSTI) are *Staphylococcus aureus* and GAS (1). Common manifestations of SSTI include erysipelas, cellulitis, necrotizing fasciitis (often polymicrobial) and Fournier's gangrene (1). The most common portal of GAS entry which may result in a SSTI is a break in the host's skin (1). Necrotizing fasciitis, which is the most severe SSTI, is an infection of the deeper subcutaneous tissue and fascia and it is distinguished by broad and rapidly spreading necrosis and gangrene of the skin and underlying structures (1). Even with appropriate treatment, mortality rates for necrotizing fasciitis are extremely high (20%-70%) (1).

Streptococcal toxic shock syndrome (STSS) occurs when an individual is infected with a GAS strain that possesses a streptococcal superantigen exotoxin (10). The exotoxins exert a strong T-cell response which results in the release of pro-

inflammatory cytokines causing the sudden onset of shock and organ failure (1). STSS manifests in three phases namely: (i) fever, chills, nausea, vomiting and diarrhea that is accompanied by low blood pressure within 24-48 hours of initial symptom presentation; (ii) increased heart rate (>100 beats per minute), abnormal breathing and severe pain at the site of infection; and (iii) the sudden onset of shock and organ failure (1).

1.2.3 Post-streptococcal sequelae

1.2.3.1 Acute rheumatic fever/Rheumatic heart disease

Acute rheumatic fever (ARF) is an autoimmune disease that may develop after pharyngitis or a skin infection caused by GAS strains that are rheumatogenic i.e. elicit an autoimmune response (23,24). After an infection with GAS occurs, an immune response is triggered in which neutrophils, macrophages and dendritic cells initiate phagocytosis of the bacteria and present it to the T cells (23). T cells respond by initiating T cell activation which together with B cells produces antibodies (23). In individuals who develop ARF, an autoimmune reaction against various host tissues (heart, brain, joints etc.) is triggered by both the antibodies and T cells through a process termed molecular mimicry as the antibodies and T cell epitopes are shared between the pathogen and those human host tissues (23). Symptoms of ARF include polyarthritis, Sydenham's chorea, skin manifestations (such as erythema marginatum and subcutaneous nodules on skin) and inflammation of the heart valves and myocardium (23). Repeated episodes of ARF may lead to permanent damage of the heart valves, thus resulting in rheumatic heart disease (RHD) (23). A 2005 systematic review estimated that globally, approximately 471 000 cases of ARF occurred annually and most of these cases (336 000; 71%) occurred in children aged 5-14 years (19). Deaths due to ARF or RHD were estimated at 350 000 annually, with the majority of these deaths occurring in low- and middle-income countries (19). A South African study reported that during 2006-2007, the incidence of newly diagnosed RHD cases among adults (>14 years old) in an urban community was 23.5 cases /100 000 per annum (25).

1.2.3.2 Acute post-streptococcal glomerulonephritis

Acute post-streptococcal glomerulonephritis (APSGN) is an immune-mediated sequelae of pharyngitis or skin infections caused by GAS (4). APSGN is not caused by a direct GAS infection in the kidneys but rather it is a kidney disease that develops when the host immune system, while combating the GAS infection, mistakenly targets and damages healthy renal tissue (4). Symptoms of APSGN may develop as soon as 10 days after GAS pharyngitis or may take 3- 6 weeks to develop after skin infections (21). The clinical presentation of APGSN includes malaise, fatigue, headache, edema, hypertension, blood in urine and elevated levels of protein in the urine (21). Globally, more than 470 00 cases of APSGN occur annually resulting in approximately 5000 deaths, the majority of which occur in low-income countries (19).

1.3 Diagnosis

It is important to rapidly diagnose GAS disease to initiate appropriate patient management for the prevention of serious complications and to prevent potential outbreaks from occurring (26). Clinical suspicion based on symptom presentation alone is not enough to accurately diagnose a patient with GAS disease as the signs and symptoms of a GAS infection may overlap with the signs and symptoms of other infectious causes (27). Therefore, a laboratory test may be necessary to accurately diagnose a patient with GAS disease.

The gold standard for the diagnosis of GAS infection is culture (26). Briefly, an appropriate specimen type is collected from an individual suspected to have GAS disease and the specimen is inoculated on horse blood agar and incubated in aerobic conditions for at least 48 hours (26). Resulting colonies are identified as GAS based on colony morphology, the presence of β-hemolysis, Gram-positive staining, a negative oxidase and catalase test and a sensitivity to bacitracin (5,26,27). Matrix-Assisted Laser Desorption/Ionization-Time Of Flight (MALDI-TOF) mass spectrometry and latex agglutination can also aid in the detection of GAS (26). One of the limitations of using culture is the long turn-around-times (TAT's) as it may take 2-3 days to obtain a result which is not ideal in emergencies. Another limitation

is that patients may be treated empirically using antibiotics prior to obtaining a specimen for testing and in such cases, obtaining a positive culture result may not be possible.

Serology may also be used for the diagnosis of disease caused by GAS infection. Serology testing is normally utilized when a prior infection with GAS is suspected in a patient that presents with symptoms suggestive of post-streptococcal sequelae, such as ARF and APSGN (15). Serology is the preferred method for diagnosis in such cases as the post-streptococcal sequelae may only be evident long after the initial infection with GAS has cleared, thus yielding a negative culture result (15). Serology may also be utilized to discriminate between carriage and active infection with GAS (28). Streptolysin O and deoxyribonuclease B antigens are one of the main targets of streptococcal serological assays (15). Similar to culture, a limitation of using serology testing to diagnose GAS disease is the long TAT's as it may take up to 4 weeks to mount an immune response that can be detected and this may not be useful for diagnosis in patients who present with acute GAS disease (15).

As mentioned above, the use of culture and serology to diagnose a patient with GAS disease can result in long TAT's and this is why over the past two decades, other techniques have been developed in an effort to shorten TAT's. One such technique is the use of immunochromatographic rapid antigen detection tests (RADTs) that detect the polysaccharide C cell-wall antigen of the bacteria (26). These rapid tests are often used for patients that present with acute pharyngitis in order to determine the etiological agent so to make a decision on the need for antibiotics. RADT's offer the advantage of quick test results and testing can be conducted either within the laboratory or at the physician's office as a point-of-care test (POCT) (27). One of the major short-comings of RADT's is that interpretation of results relies heavily on the user, which may introduce user variability and subjectivity (27). Nucleic acid amplification tests (26,27) including real-time polymerase chain reaction (qPCR) have also been developed for the detection of GAS in specimens by targeting specific conserved regions in GAS such as the *spy* gene (29).

The diagnosis of ARF is made based on fulfilling specific clinical criteria, referred to as the Jones Criteria (**Table 1**) (30). Initially established in 1944 (31), the Jones Criteria has undergone several updates with the latest update in 2015 (30). The

Jones Criteria is divided into major and minor manifestations. A patient diagnosed with ARF for the first time should have laboratory results indicating a preceding GAS infection and in addition, present with two major manifestations or one major manifestation accompanied by at least two other minor manifestations (30). Should a patient present only with Sydenham's chorea, they are exempt from fulfilling these criteria as this manifestation may only become evident long after GAS infection and thus a laboratory test result may not yield a positive result (32). For patients with a history of ARF or RHD, diagnosis is based on laboratory results indicating a preceding GAS infection and two or one major manifestations together with two or three minor manifestations (30).

Table 1: The Jones Criteria 2015 used for the diagnosis of acute rheumatic fever. Adapted from (30).

Criteria	Population	Manifestations
Major	*Low-risk populations	+Carditis (Clinical and/or subclinical)
		Arthritis (Polyarthritis only)
		Chorea
		Erythema marginatum
		Subcutaneous nodules
	Moderate-and high-risk populations	Carditits (Clinical and/or subclinical)
		Arthritis (Monoarthritis, polyarthritis or polyarthralgia)
		Sydenham's chorea
		Erythema marginatum
		Subcutaneous nodules
Minor	*Low-risk populations	§Polyarthralgia
		Fever (≥38.5°C)
		ESR ≥60 mm/hr and/or CRP≥ 3.0 mg/dL
		Prolonged PR interval, after accounting for age variability (unless carditis is a major criterion)
	Moderate-and high-risk populations	Monoarthralgia
		Fever (≥38°C)
		ESR ≥30 mm in the first hour and/or CRP≥ 3.0 mg/dL
		Prolonged PR interval, after accounting for age variability (unless carditis is a major criterion)

ESR- Erythrocyte sedimentation rate; CRP- C-reactive protein

Erythema marginatum and subcutaneous nodules rarely occur in isolation as major criteria.

In a single patient, joint manifestations can only be regarded as either a major or a minor criterion

CRP value must be greater than the normal laboratory upper limit

ESR values may change during the course of ARF and therefore, peak ESR values should be used

*Low-risk populations are defined as those with an annual ARF incidence ≤2 per 100 000 school-aged children or an annual RHD prevalence of ≤1 per 1000 population in all ages

+Subclinical carditis is defined as echocardiographic valvulitis

§ Polyarthralgia is considered a major manifestation only in moderate-to high-risk populations after other causes have been ruled out

1.4 Prevention and treatment of GAS disease

1.4.1 Prevention

Currently, the only available treatment strategy for GAS disease is antibiotics (18). To prevent ARF and RHD, in South Africa there are two strategies that are recommended: primary prevention and secondary continuous prophylaxis (33). As a primary prevention strategy, it is recommended that children aged 3 -15 years presenting with a sore throat are treated with penicillin to prevent ARF (33). In addition, it is recommended that children with a history of ARF should continuously be treated with penicillin as prophylaxis, either intramuscularly once every four weeks or twice daily orally until they reach 21 years old to prevent recurrent ARF (33). Any individual with a history of RHD should receive prophylaxis until they are 35 years old (33).

1.4.2 Treatment

Antimicrobial resistance patterns of GAS to various antibiotics vary globally, but using phenotypic and genotypic techniques, various studies have reported that GAS causing invasive and non-invasive disease continue to be universally susceptible to β-lactams such as penicillin (18,34–40). Penicillin and penicillin-based antibiotics are therefore the first-line agents for the treatment of invasive and non-invasive GAS disease and the prevention of ARF/RHD (18). Macrolides (erythromycin) are used as an alternative for treating patients who are allergic to penicillin (21,41). Some patients may experience gastrointestinal side effects when taking erythromycin and therefore other antibiotics (e.g., clindamycin, azithromycin) have been approved for the treatment of GAS disease (21). Vancomycin or linezolid can also be used to treat patients who are intolerant to penicillin (41).

For severe iGAS disease such as necrotizing fasciitis, meningitis, pneumonia and streptococcal toxic shock syndrome (STSS), a combination of clindamycin and penicillin is recommended for treatment (42) as clindamycin has been shown to inhibit protein synthesis at the ribosomal level, thus resulting in the inhibition of the synthesis of GAS virulence factors such as streptococcal pyrogenic exotoxins and the M-protein (39). Although not routinely used to treat GAS disease, particularly in

children due to the negative side effects they have on bone and joint development (43) fluoroquinolones such as levofloxacin can be used as an alternative for adult patients who cannot take penicillin and considering the global increase in macrolide resistance (44). Tetracycline antibiotics are not commonly prescribed for the treatment due to the high rate of resistance among GAS (43).

1.4.3 Antimicrobial non-susceptibility

The antimicrobial resistance rates of GAS to different antibiotics vary globally (45). Globally, GAS remain susceptible to β- lactams. Although clinical resistance to β-lactam antibiotics has never been reported, in 2021 the CDC reported two of the first cases presenting with iGAS disease that exhibited a rare penicillin-binding protein 2x (*pbp2x)* mutation resulting in elevated β-lactam (ampicillin, amoxicillin and cefotaxime) minimum inhibitory concentrations (MIC) (46).

Various studies have demonstrated a resistance rate to macrolides varying between 3% to 5%, with an increase being reported (18,35,37). A 2024 South African study on iGAS and non-iGAS isolates reported a much lower resistance rate of 1% to the macrolides azithromycin and erythromycin (47). Two main mechanisms of macrolide resistance in GAS have been described (48). The first mechanism is the modification of the ribosomal macrolide-binding site, thus preventing the antimicrobial from binding to the ribosome (49). This mechanism of resistance is encoded by the *erm* (erythromycin ribosome methylase) gene family and results in the cross-resistance to all macrolides, lincosamides and streptogramins B, commonly referred to as the MLSB phenotype (48). The second mechanism of macrolide resistance involves the active removal of the antimicrobial from the cytoplasm (48). A membrane-spanning pump (macrolide efflux), encoded by the *mef(A)* and *msr(D)* gene, lowers the intracellular antibiotic concentration by pumping out the antimicrobial, thus reducing its effectiveness (48). This process only results in low to moderate resistance to 14- and 15- membered lactone ring macrolides, termed the M phenotype (48). In addition to the resistance encoded by the *erm* family genes mentioned above, two other mechanisms resulting in resistance to lincosamides (clindamycin) have also been described (45). The two mechanisms are i) drug inactivation which is driven by the *lnuB* (lincosamide nucleotidyltransferase) gene and ii) active efflux by an efflux pump encoded by the *lsa* gene (45).

Resistance to tetracycline is due to ribosomal protection proteins encoded by the *tet* gene family (41). Resistance of GAS to tetracycline is generally higher than the resistance against macrolides, with a 2011 Tunisian study reporting a tetracycline resistance rate as high as 70% amongst their isolates collected from invasive and

non-invasive sites (50). In GAS, resistance to fluoroquinolones is as a result of point mutations that occur within the quinolone resistance determining region of topoisomerase IV ParC and/or topoisomerase II DNA gyrase GyrA (51).

The over-use of antibiotics may drive antibiotic resistance, which is a big global issue as this may lead to the development of resistant GAS strains (46). The over exposure of commensal flora to antibiotics may also result in long-term dysbiosis (52).

1.4.4 Vaccines

A vaccine is currently not available to prevent GAS disease. As mentioned in section 1.2.3.1, the highest burden of the most severe consequence of GAS disease, ARF and RHD, is seen in low-income countries. This may be an indication that ARF is not treated adequately in these settings due to limited access to healthcare, thus resulting in RHD. A vaccine is therefore the best intervention to reduce the incidence of ARF and RHD.

1.4.4.1 M-protein-based vaccines

Various vaccine formulations are currently in development and the leading vaccine target antigen is the M-protein (53,54). Serology testing has shown that the host immunity response targets the M-protein by producing M-protein type-specific antibodies (55,56). These antibodies have been shown to be opsonic and promote the ingestion and killing of GAS by phagocytic cells (55). The leading vaccine currently in development is the 30-valent M-protein based vaccine that is designed based on the epidemiology and *emm* type prevalence of GAS in high-income countries (53–55). As mentioned in section 1.1.4, the M-protein has a hypervariable N-terminal region (a portion of the *emm* gene) that determines the *emm* type of a strain (4). The 30-valent vaccine is a recombinant vaccine consisting of four different recombinant proteins that contain 7 or 8 N-terminal peptides representing 30 *emm* types (**Figure 3**) (54,57). A streptococcal protective agent 18 (Spa18) that is expressed in many GAS strains is also included in this vaccine formulation (53,54,57). The *emm* types included in the 30-valent vaccine are *emm* types (1, 2, 3,

4, 5, 6, 11, 12, 14, 18, 19, 22, 24, 28, 29, 44, 49, 58, 73, 75, 77, 78, 81, 82, 83, 87, 89, 92, 114 and 118) that are important causes of invasive disease and pharyngitis in North America and Europe (53–55). *emm* types that have been historically considered to be rheumatogenic (1, 2, 5, 6, 11, 12, 14, 17, 18, 19, 24, 27 and 29) are also included in this vaccine formulation (9,54). The 30-valent vaccine is currently in phase I testing of clinical trials (54). The biggest challenge with development of an M-protein-based vaccine has been the diversity of *emm* types and that distribution and prevalence of these *emm* types vary across geographic regions (58). This is a major limitation particularly for low- and middle-income countries, where populations are at a greater risk for developing ARF and RHD, as the *emm* type distribution and prevalence in these settings may differ from high-income countries (22,58). An animal study conducted using the 30-valent vaccine revealed that there may be cross-protection against some *emm* types that are not included in the vaccine formulation (57). In this animal study, forty *emm* types were tested of which 24 (8, 9, 15, 30, 33, 36, 40, 48, 51, 52, 59, 65, 66, 68, 76, 79, 85, 94, 97, 103, 105, 109, 111 and 122) showed >40% bactericidal killing activity in the presence of rabbit anti-sera generated after vaccination (57). A study done in Cape Town, where *emm* types were determined for invasive and non-invasive GAS isolates, reported that the 30-valent vaccine would only potentially offer protection in 23% (54/233) of the typed isolates (59). Five of the most prevalent *emm* types detected in this study (76, 81, 80, 43 and 183) were not part of the vaccine formulation (59). When cross-opsonisation was considered, potential vaccine coverage in these isolates rose to 66% (154/233) (59), which is still much lower than the potential vaccine coverage in high-income countries such as the United States where potential vaccine coverage is >90% (60). In a Kenyan study, the potential vaccine coverage was 28% (99/357) amongst iGAS isolates and this rose to 57% (203/357) when cross-opsonisation was considered (61). In a study conducted in Malian school children presenting with pharyngitis, the potential vaccine coverage was 31.0% (139/449) for vaccine serotypes and 75.1% (337/44) with cross opsonization (62).

Disadvantaged individuals of the aboringal population in Australia have one of the highest prevalence of RHD and it is believed that the high prevalence of GAS skin infections in this populations is the cause (63). The proposed 30-valent vaccine may not offer adequate protection as the formulation of this vaccine does not take into account the *emm* types that cause skin infections and these *emm* types differ from

the types that cause invasive disease and pharyngitis (63). In addition, GAS skin infections are caused by a greater diversity of *emm* types in comparison to pharyngitis and iGAS disease (63).

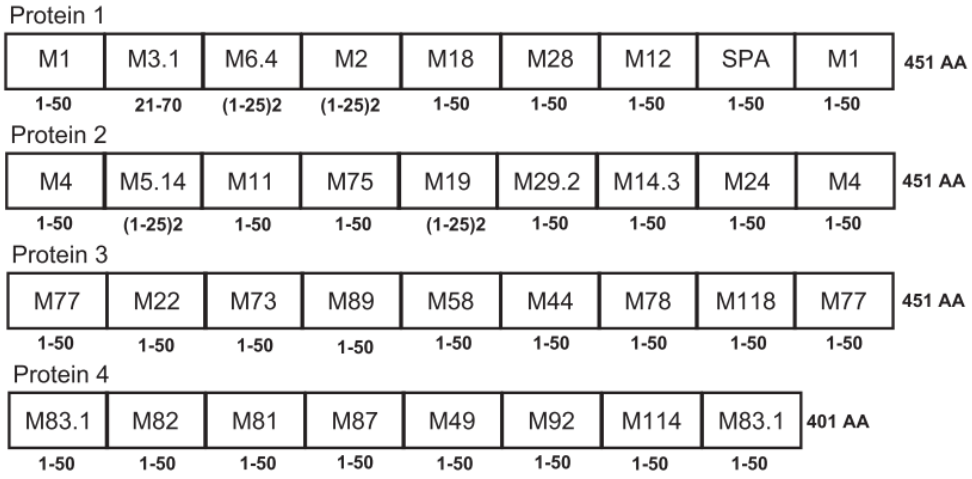


Figure 3: The four proteins containing the 30-valent M-protein-based group A streptococcus vaccine. The numbers in the blocks represent the M-protein type and the numbers below each peptide indicate the amino acids of the mature M-protein included in the vaccine protein. Adapted from (57).

Other GAS antigens that may have protective epitopes and are present in most or all GAS strains are being explored (58,64). One such antigen is the M-related protein (Mrp) that is a member of the M-protein family of GAS, which includes Emm (referred to as M-protein throughout this study) and Enn proteins (65). Some isolates express only the M-protein while others express the M-protein, Mrp and/or Enn protein (65). The Mrp is a virulence factor of GAS and has been shown to evoke an immune response in rabbits (58,65). The N-terminal region of the *mrp* gene that codes for this protein is not hypervariable but rather semi-conserved (58). Mrp is present in most GAS clinical isolates that have been studied thus far (58,65). All Mrp sequenced thus far belong to three structurally similar families, which are MrpI, MrpII and MrpIII (58,65). A trivalent Mrp (trMrp) vaccine formulation (**Figure 4**) that includes Mrp representing each of the three families has been tested in rabbits and an immune response against each component of the trivalent vaccine (i.e. each of the individual Mrp proteins) and the recombinant trMrp vaccine protein was observed (58). The trMrp vaccine sera produced by the rabbits in this study displayed bactericidal

activity against isolates of different *emm* types (some not included in the 30-valent M-protein vaccine) of GAS that expressed Mrp from all three families (58). In this study they also explored the possible outcome of a vaccine construct that is a combination of both the 30-valent M-protein and the trMrp vaccine (58). They did this by mixing 30-valent antiserum and the trMrp vaccine antiserum produced by the rabbits and performed bactericidal assays using this serum (58). Bactericidal activity was higher when the 30-valent and trMrp immune sera were mixed than when each serum was tested on its own (58).

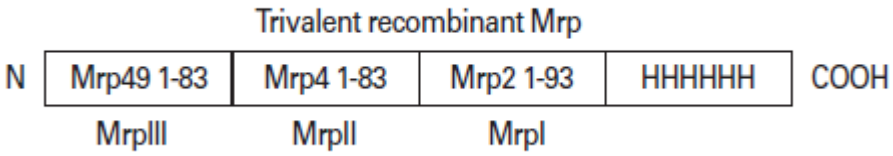


Figure 4: The potential trivalent recombinant M-related protein-based vaccine construct. Adapted from (58).

1.4.4.2 Non M-protein- based vaccines

In addition to the M-protein another antigen of interest is J8, which is a 29 amino acid long peptide found in the conserved C-terminal region of the M-protein (66,67). In the vaccine formulation, the J8 peptide is conjugated with diphtheria toxoid and adsorbed onto aluminium hydroxide (66). This vaccine has reached phase I clinical trial testing (MJ8VAX) (66).

1.5 GAS disease surveillance

A study conducted in 2005 estimated that globally from 1 January 1980, there were 1.78 million new cases of severe GAS disease (ARF, RHD, APSGN and iGAS) occurring each year, resulting in approximately 517 000 deaths annually in individuals of all ages (19). The majority of these deaths were as a result of iGAS disease (31.5%, 163 000) and RHD (45.1%, 233 000) (19). There are limited published studies on invasive and non-invasive GAS disease in Africa and as a

result, it is difficult to accurately estimate the true burden of GAS disease in this region (68).

From 2014, high-income countries reported an increase in the annual incidence of invasive and non-invasive GAS disease (69,70). This prompted various countries to re-assess their surveillance systems in order to efficiently monitor the epidemiology of GAS disease (69–71). England, Wales and Northern Ireland noted a steep rise in scarlet fever cases in 2014 (70). In the United Kingdom (UK), scarlet fever is a notifiable medical condition and suspected cases of scarlet fever are notified to the UK Health Security Agency (UKHSA), formerly known as Public Health England (PHE), by clinicians on the basis of clinical symptoms with or without laboratory confirmation of a GAS infection (70). In addition to conducting syndromic surveillance for scarlet fever, the UKHSA also conducts laboratory-based surveillance of iGAS disease, defined as GAS cultured from a normally sterile site (70). In the United States (US), the Centers for Disease Control and Prevention (CDC) conducts population and laboratory-based surveillance of iGAS disease through their Active Bacterial Core surveillance (ABCs) of emerging infections (60). In this setting, iGAS disease is defined as the isolation of GAS from a typically sterile body site or wound culture, along with the presence of necrotizing fasciitis or streptococcal toxic shock syndrome (60). In Canada, iGAS disease is a notifiable medical condition (72). The Public Health Agency of Canada (PHAC) conducts routine iGAS disease surveillance and this is done through the Canadian notifiable disease surveillance system (72). A case of iGAS disease in this setting is defined as a laboratory-confirmed iGAS diagnosis from a normally sterile body site, bone or joint fluid excluding the middle ear and superficial wound aspirates (e.g. skin and soft tissue abscesses) (72). The National Microbiology Laboratory (NML) in Canada conducts national, laboratory-based surveillance of invasive streptococcal disease in Canada and GAS is one of the pathogens included in this surveillance (72).

There is a paucity of data on the epidemiology of GAS disease in Africa because of the lack of surveillance systems and as an effort to address this, a number of initiatives have been started. Currently, the National Institute for Communicable Diseases (NICD) in South Africa reviews ARF reported through the electronic Notifiable Medical Conditions (NMC) system (73). ARF is listed on the NMC system as a category 1 NMC, necessitating immediate reporting by the most rapid means

available upon diagnosis. Subsequently, a written or electronic notification to the Department of Health (DoH) within 24 hours of diagnosis should be submitted by health care providers, private health laboratories or public health laboratories (73). Although useful data is received through this notification system, a 2006 publication reported that there is underreporting of ARF cases by health care professionals and in addition, there is poor administration of the notification system thus resulting in a decline in the number of reported cases in South Africa between 1990 – 2004 (74). In 2019, the NICD also started conducting active, national, laboratory-based surveillance for iGAS disease through the GERMS-SA (Group for Enteric, Respiratory and Meningeal Disease Surveillance in South Africa) network (75). In GERMS-SA, over 200 microbiological labs (public, private, military and mining) across the country report iGAS cases and submit invasive isolates to the Centre for Respiratory Diseases and Meningitis (CRDM) at the NICD for further phenotypic and genotypic characterization. The African group A Streptococcal infection registry (Afro*Strep*) was established in 2016 and is a collaborative multicenter study to describe the clinical, microbiology, epidemiological and molecular characteristics of invasive and non-invasive GAS infections across Africa (71). In the Afro*Strep* registry study, active surveillance of GAS pharyngitis is conducted in sentinel primary care centers across Africa and passive surveillance of iGAS disease is conducted by collecting isolates from microbiology laboratories across Africa (71). This study will serve as a platform to gain a better understanding of GAS epidemiology across Africa (71).

1.6 Outbreaks of GAS disease

GAS is capable of causing endemic outbreaks. Scarlet fever outbreaks have been described in the literature, occurring in China (76), Hong Kong (77), Spain (78) and the UK (79). iGAS disease outbreaks have also been reported in the US (80), the UK (81–83), Canada (84) and New Zealand (85). Outbreaks may arise from the transmission of a GAS strain to susceptible individuals, either through a shared source (81,84,86) or, particularly in crowded settings, through person-to-person contact (78,87). Specific GAS clones that are more virulent may also drive GAS outbreaks (82,83,85).

1.7 Molecular epidemiology of GAS

1.7.1 GAS typing

Typing/characterization of GAS isolates is an important part of understanding the epidemiology of GAS disease (20,88). Typing can include using methods to study the morphology of an organism (e.g., capsular wall, surface antigens, pili etc.) which provides insight into why certain strains are more virulent and thus more successful than others (20,88). Similarities between strains may also help elucidate transmission patterns and detect outbreaks and clusters (20). GAS are highly diverse organisms and typing methods such as pulse-field gel electrophoresis (PFGE), where the DNA fingerprint of an organism is determined, and multilocus sequence typing (MLST) using 7 housekeeping genes have been utilized to genotype GAS and elaborate on the clonal structure of GAS populations (89–94). Serological methods have also been used to type GAS isolates, mainly by targeting the M-protein of this organism (20). The variants of the M-protein can be identified and typed serologically using the anti-sera of recognized type specific anti-M sera (11).

1.7.1.1 *emm* typing

Although the typing methods mentioned in section 1.7.1 have been useful in the past, they have largely been replaced by the molecular sequence typing of the M-protein, a technique referred to as *emm* typing (20). *emm* typing is the sequence typing of a portion (the hypervariable N-terminal region) of the M-protein (*emm*) gene, which is responsible for coding of the M-protein (20), and is currently the most common technique used to discriminate between different GAS strains. Due to the hypervariability of the N- terminal region, different sequences coding for that region will result in different *emm* types (6). Currently ~243 *emm* types have been described and the prevalence and distribution of circulating *emm* types differ by geographic regions (6,88).

1.7.1.1.1 Prevalence of *emm* types

Table 2 depicts a summary of *emm* studies from different continents conducted between 2011 – 2023. The data presented in the table as well as previous reports on the global distribution of *emm* types (95) illustrates that the distribution of *emm* types differs geographically.

Table 2: Summary of *emm* typing studies published between 2011 - 2023 from six continents

Continent	Country	[a]Year	Specimen type	Sample size	Typed isolates	No. of *emm* types	[b]Prevalent *emm* types	Reference
Africa	South Africa	2014	Non-invasive	157	157	26	48, 89, 4*, 12*, 75*, 1¢, 94¢, 22, 9	(22)
	Mali	2015	Non-invasive	449	396	70	65, 18, 55*, 42*, 81¢, 58¢, 25, 109, 11, 169	(62)
	Kenya	2016	Invasive	357	357	88	44, stg866, 65, 18, 11*, 179*, 8¢, 90¢, 77¢, 92¢	(61)
	South Africa	2021	Invasive and Non-invasive	54	43	15	92, 6, 25*, 75* 52, 3¢, 70¢,st3735.0¢,st2147.0¢	(96)
	The Gambia	2023	Non-invasive	115	107	46	80, 85, 229, *emm*/stG1750	(97)
Europe	France	2011	Invasive	1542	1542	83	1, 28, 89, 4, 3, 12, 75	(98)
	Sweden	2013	Invasive	584	494	NA	1, 89, 28, 3	(99)
	Iceland	2014	Invasive	288	226	25	1, 28*, 89*, 3, 12	(100)
	Finland	2015	Invasive	1165	1122	72	28, 89, 1, 12, 119, 4*, 77*	(101)
	Norway	2016	Invasive	756	756	52	1, 89, 12, 3, 28, 4, 66, 112	(102)
	Germany	2017	Invasive	719	719	46	1, 28, 89, 2, 12	(103)
	Ireland	2018	Invasive	561	473	33	1, 3, 28, 12, 89, 4, 6	(104)
	Portugal	2019	Invasive	381	381	40	1, 89, 3, 12, 6	(105)
	Spain	2019	Invasive	103	93	14	1, 12, 6, 3*, 75*, 4	(34)
	Spain	2020	Invasive	29	26	11	1,3, 5*, 6*, 44*	(106)
	Spain	2021	Invasive	1893	1893	58	1, 89, 3, 4, 12, 6, 87*, 28*, 75¢, 77¢	(107)
	Belgium	2023	Invasive	518	122	37	1, 64, 11, 77, 5*, 89*, 101*	(108)
	United Kingdom	2023	Non-invasive	142	142	23	108, 89, 12, 1, 4	(97)
Asia	Lebanon	2011	Non-invasive	103	103	31	1, 22, 28*, 88*, 4, 97, 12, 110¢, stC5345.1¢, stC6979¢	(90)
	China	2020	Non-invasive	297	297	9	12, 1, 75, 89, 3, 6, 225*, 4*, stg485*	(37)
	Japan	2021	Invasive	621	621	34	1, 89, 12, 3, 6, 4, 28, 76, 77	(109)
North America	Alaska	2016	Invasive	516	422	51	1, 82, 49, 12*, 3*, 89, 108, 28, 92, 41	(110)
	Canada	2018	Invasive	3551	3053	65	1, 28, 59, 12, 82, 3, 89, 101, 11, 41	(111)
	USA	2020	Invasive	3895	3864	71	1, 89, 49, 12, 3, 82, 92, 28, 4	(112)
	USA	2022	Invasive	483	194	38	1, 12, 28, 11°, 4°	(113)
South America	Brazil	2014	Non-invasive	91	86	28	12, 1, 4, 22*, 8*	(114)
	Brazil	2015	Invasive and Non-invasive	229	229	48	1, 87, 22*, 12*, 77¢, 6¢, 89, 33β, 75β, 3β	(115)
Oceania	New Caledonia	2010	Invasive	90	90	31	15, 92, 106, 74, 89*, 109*	(116)
	New Zealand	2014	Invasive	2861	2861	111	1, 49, 81, 75*, 89*, 82, 92, 12, 91, 28	(117)
	Australia	2015	Invasive	128	82	26	81, 197, 113, 101, 44*, 92*	(118)

[a] The year that the study was published; [b] Descending order of prevalence; *° ¢ β Similar number of isolates; NA – Not available

1.7.1.1.2 The *emm* cluster system

As mentioned in section 1.7.1.1, *emm* typing relies on a small portion of the M-protein *emm* gene, the hypervariable N-terminal region. A 2014 study by Sanderson-Smith *et al*., demonstrated that analysis of the complete, and not just a portion, *emm* gene sequence indicated that many *emm* types are highly similar to one another (119). In the above-mentioned study, 1806 isolates representing 175 *emm* types collected from 31 countries were used. The complete *emm* gene was sequenced for these isolates and phylogenetic analysis were conducted. The general organization of the tree obtained from the phylogenetic analysis revealed that most of these isolates belonged to two clades (X and Y) with the exception of isolates that were *emm* type 55, 95, 111, 215, 221 and 222. Clade Y was further divided into two major subclades ,Y1 and Y2 (119). Within clades X and subclades Y1 and Y2, isolates were further divided into 48 *emm* clusters (**Table 3**). In 16 (E1 - E6, D1 - D5 and AC1 - AC-5) of the 48 *emm*- clusters, each cluster had isolates that represented a variety of *emm* types (e.g., *emm* cluster E1 had *emm* types 4, 60, 78, 165 and 176; *emm* cluster E2 had *emm* types 13, 27, 50 (50/62), 66, 68, 76, 90, 92, 96, 104, 106, 110, 117, 166 (st1207), 168 (st1389) etc.). This phylogeny indicated that although different *emm* types, *emm* types in the same *emm* cluster were highly similar based on their complete *emm* gene sequence. The remaining 32/48 *emm* clusters had isolates that expressed only one specific *emm* type [e.g. *emm* cluster 5 had isolates that expressed only the M5 (*emm*5) protein and *emm* cluster 14 had isolates that expressed only the M14 (*emm*14) protein]. The 32 *emm* clusters mentioned above indicated that some *emm* types are highly divergent from others, thus not clustering closely to other *emm* types within the phylogenetic tree.

In addition to conducting the phylogenetic analysis to establish the *emm* cluster system, Sanderson-Smith *et al*., also used a subset of isolates that were a representative *emm* type from each of the dominant *emm* types to assess their binding capability to key host proteins that are known to interact with M-proteins (119). The results indicated that *emm* types within an *emm* cluster displayed homogenous functional profiles (119).

Table3: The clustering of 173 *emm* types into 48 *emm* clusters based on phylogenetic analysis. Adapted from (119).

***emm* types**	**Clade**	**Sub-clade**	***emm*-cluster**
4, 60, 78, 165 (st11014), 176 (st213)	X	N/A	E1
13, 27, 50 (50/62), 66, 68, 76, 90, 92, 96, 104, 106, 110, 117, 166 (st1207), 168 (st1389)	X	N/A	E2
9, 15, 25, 44 (44/61), 49, 58, 79, 82, 87, 103, 107, 113, 118, 144 (stknb1), 180 (st2460), 183 (st2904), 209 (st6735), 219 (st9505), 231 (stNS292)	X	N/A	E3
2, 8, 22, 28, 73, 77, 84, 88, 89, 102, 109, 112, 114, 124, 169 (st1731), 175 (st212), 232 (stNS554)	X	N/A	E4
34, 51, 134 (st2105), 137 (st465), 170 (st1815), 174 (st211), 205 (st5282)	X	N/A	E5
11, 42, 48, 59, 63, 65 (65/69), 67, 75, 81, 85, 94, 99, 139 (st7323), 158 (stxh1), 172 (st2037), 177 (st2147), 182 (st2861UK), 191 (st369)	X	N/A	E6
36, 54, 207 (st6030)	Y	Y1	D1
32, 71, 100, 115, 213 (st7700)	Y	Y1	D2
123, 217 (st809)	Y	Y1	D3
33, 41, 43, 52, 53, 56, 56.2 (st3850), 64, 70, 72, 80, 83, 86, 91, 93, 98, 101, 108, 116, 119, 120, 121, 178 (st22), 186 (st2940), 192 (st3757), 194 (st38), 208 (st62), 223 (stD432), 224 (stD631), 225 (stD633), 230 (stNS1033), 242 (st2926)	Y	Y2	D4
97,157 (stn165), 184 (st2911)	Y	Y2	D5
46, 142 (st818)	Y	Y1	A-C1
30, 197 (st4119)	Y	Y1	A-C2
1, 163 (st412), 227 (stil103), 238 (1-2), 239 (1-4)	Y	Y2	A-C3
12, 39, 193 (st3765), 228 (stil62), 229 (stmd216)	Y	Y2	A-C4
3, 31, 133 (st1692)	Y	Y2	A-C5
164 (st106M), 185 (st2917), 211 (st7406), 236 (sts104) 36,	X	N/A	Single M-protein *emm*-cluster belonging to clade X
5, 6, 14, 17, 18, 19, 23, 24*, 26, 29, 37, 38 (38/40), 47, 57, 74, 105, 122*, 140 (st7395), 179 (st221), 218 (st854), 233 (stNS90), 234 (stpa57)	Y	Y1 or Y2*	Single M-protein *emm*-cluster belonging to clade Y
55, 95, 111, 215 (st804), 221 (stCK249), 222 (stCK401)	Outlier	N/A	Single M-protein *emm* cluster outlier (i.e. not part of clade X or clade Y)

The asterisks (*) mark indicates *emm* types that cluster into sub-clade Y2. Brackets () represent old nomenclature

1.7.1.2 Whole- genome sequencing

Whole-genome sequencing (WGS) is fast becoming the main technique used to study and describe the molecular characteristics of pathogens, as this method offers greater genomic resolution and the development of bioinformatics tools make it feasible to manage and analyze the large amounts of data generated through this technology (120,121). A study in the United States using whole genome sequencing data to describe the molecular epidemiology of GAS strains causing invasive disease, obtained from one year of population-based surveillance, was the first of its kind (122). From the whole genome sequencing data, strain characteristics including *emm* types, MLST's, predictors of antimicrobial resistance and other key features associated with the virulence and adhesion of iGAS strains were extracted (122). There is a paucity of whole genome sequencing GAS data in Africa and studies focusing on this would be beneficial to gain a better understanding of the GAS population in this setting.

2 STUDY AIM AND OBJECTIVES

2.1 Aim

To describe the epidemiology of iGAS and molecular characteristics of iGAS isolates obtained from active, national, laboratory-based surveillance (GERMS-SA) in South Africa from 01 January 2019 – 31 December 2020

2.2 Objectives

(i) Describe the demographic and clinical characteristics of cases with iGAS reported through the GERMS-SA surveillance in South Africa from 1 January 2019 - 31 December 2020

(ii) Document the antimicrobial resistance patterns to relevant antimicrobials of iGAS isolates collected through the GERMS-SA surveillance in South Africa from 1 January 2019 - 31 December 2020

(iii) Perform whole genome sequencing on iGAS isolates collected through the GERMS-SA surveillance in South Africa from 1 January 2019 - 31 December 2020 in order to:

- Assess predictors of antimicrobial resistance
- Describe *emm* type prevalence and determine the potential vaccine coverage of the proposed 30-valent M-protein-based vaccine
- Determine the presence of the M related protein (Mrp) in all isolates
- Determine multilocus sequence type (MLST) based genotypes
- Determine key virulence factors and strain features associated with improved GAS colonization and transmission
- Determine genetic relatedness of isolates in order to detect clusters and outbreaks

3 METHODS AND MATERIAL

3.1 Ethical approval

GERMS-SA (an active, national, laboratory-based surveillance system) ethics has been obtained from the Human Research Ethics Committee (HREC) of the University of the Witwatersrand (Protocol number: M1809107), Johannesburg, South Africa. Project specific ethics was also obtained from the University of Witwatersrand (M210750), Johannesburg, South Africa (**Appendix A**).

3.2 GERMS-SA surveillance and isolate collection

GAS cases were detected through the GERMS-SA (Group for Enteric, Respiratory and Meningeal Disease Surveillance in South Africa) network. The GERMS-SA network was established in 1999 to initially conduct surveillance of invasive disease caused by *Haemophilus influenza*e type b , *Neisseria meningitidis* and *Streptococcus pneumoniae* in South Africa (123). In 2003, surveillance was enhanced at select sentinel sites where dedicated surveillance officers were placed to collect additional clinical and demographic data for each case (124) and in 2019, surveillance of invasive disease caused by *Streptococcus agalactiae* (Group B Streptococcus; GBS) and *Streptococcus pyogenes* (GAS) was added. Presently, GERMS-SA conducts active, national, laboratory-based surveillance in the nine provinces of South Africa, representing 60.6 million persons (https://www.statssa.gov.za/?p=15601 Accessed: 12 May 2023). Approximately 200 clinical microbiology laboratories (including public, private, military and mining sectors) across the country are requested to report iGAS cases and submit available laboratory confirmed isolates collected from patients in all age groups to the National Institute for Communicable Diseases (NICD) Centre for Respiratory Diseases and Meningitis (CRDM) for further phenotypic and genotypic characterization. In addition, routine quarterly retrospective audits of the laboratory information system (LIS) TrakCare [used by the National Health Laboratory Services (NHLS) in the public sector] were done to detect cases that are not recorded on GERMS-SA. These cases were then added to the GERMS-SA database. This study was conducted on iGAS cases recorded between 1 January 2019 - 31 December 2020.

3.3 Case definitions

Individuals with iGAS disease were defined as people with GAS identified from any normally sterile site specimen [e.g., cerebrospinal fluid (CSF), blood, pleural fluid, pericardial fluid, joint fluid, tissue etc.] or from a non-sterile (wound) site specimen in a patient presenting with necrotizing fasciitis or streptococcal toxic shock syndrome (**Appendix B**). The patient clinical presentation is often detailed on the specimen submission forms that are sent together with the specimen to the diagnostic laboratories.

3.4 Patient demographic and clinical data

Isolates received at CRDM were accompanied by a TrakCare or other laboratory report and basic patient demographics (gender and age), date of specimen collection and site of specimen collection were extracted from the report and captured on the GERMS-SA database by the laboratory clerks. For cases identified by audit (representing only the public sector) and not reported to GERMS-SA, these data were extracted directly from TrakCare and also captured on the GERMS-SA database. The place of residence (by province) for each of the cases was extrapolated from the location of the testing laboratory.

From 2020, at 25 sentinel site hospitals across all 9 provinces, surveillance officers tracked all laboratory-confirmed GAS cases and an electronic case report form (CRF) was completed with additional data including race and final outcome of hospitalization (**Appendix C**). The CRF's were captured on the GERMS-SA database.

If the site of specimen collection suggested that the case might not fit the case definition, surveillance officers at enhanced sites were contacted and requested to verify true cases by checking the relevant patient hospital records. For non-enhanced sites, the reason for submission was reviewed on TrakCare to assess if non-sterile site isolates were associated with necrotizing fasciitis. Cases that could not be verified from non-enhanced sites were not included in our study.

If more than one isolates were received from one patient and the isolates were isolated within 21 days of one another, this was considered a single case and only

one isolate was included in the downstream analysis. Multiple isolates from one patient that were isolated more than 21 days apart were considered separate cases and all isolates were included in the downstream analysis.

The descriptive analyses for this study included all cases on the GERMS-SA database from 1 January 2019 - 31 December 2020. Place of residence, gender, race, age, date of specimen collection, site of infection and final outcome of hospitalization were described for each case. Incidence was stratified by age category (<1 year, 1-4 years, 5-9 years, 10-14 years, 15-24 years, 25-44 years, 45-64 years and ≥ 65 years) and calculated as follows: Incidence= (New cases) / (Population x Timeframe). Incidence is expressed as cases per 100 000 persons and was calculated using mid-year population denominators obtained from Stats SA's population projections for 2019 and 2020 (https://www.statssa.gov.za/). Cases were stratified by specimen collection date to illustrate the monthly distribution of cases.

3.5 Laboratory testing

3.5.1 Phenotypic

3.5.1.1 Bacterial culture

Upon receipt at CRDM, isolates were sub-cultured on 5% horse blood agar (Diagnostic Media Products, Johannesburg, South Africa) and incubated at 37°C in 5% CO_2 for 18-24hrs, together with a bacitracin (0.04 units) disc (Mast group, Bootle, UK). β-hemolytic pure colonies that appeared with a zone of inhibition around the bacitracin disc were presumptively identified as GAS. β-hemolytic colonies that did not exhibit a zone of inhibition around the bacitracin disc were confirmed as GAS using Matrix-assisted Laser Desorption Ionization-Time of Flight Mass Spectrophotometry (MALDI- TOF MS). Presumptive GAS colonies were cryo-preserved in glycerol skimmed milk at -70°C for further phenotypic and genotypic characterization.

3.5.1.2 Antimicrobial susceptibility testing

3.5.1.2.1 Kirby-Bauer disc diffusion method

Antimicrobial susceptibility testing (AST) was done on viable isolates. Isolates were initially screened for resistance by the direct Kirby-Bauer disk diffusion method. Preparing the culture for AST by Kirby-Bauer disc diffusion method and preparing the culture to confirm organism identification was done simultaneously. Briefly, a primary culture received on Dorset transport media was sub-cultured on 5% horse blood agar and incubated at 37°C in 5%
for 18-24hrs to check for purity. An inoculum was prepared by taking a light sweep of the growth with a sterile swab and placing it in 0.9% normal saline. A 0.5 McFarland standard was used to compare for turbidity of the suspension. The suspension was inoculated on two fresh Mueller Hinton with 5% sheep blood agar (SMHA) plates and using forceps, the following antibiotics were placed on the surface of one of the SMHA: a systematic Gram positive ring (MASTRINGS-S®, Mast Group, UK) containing oxacillin (1µg), chloramphenicol (30µg), tetracycline (30µg), erythromycin (15µg), clindamycin (2µg) and rifampicin (5µg). The penicillin disk is not stable on Mueller Hinton with 5% sheep blood agar and as a result, the oxacillin disk is used instead with the assumption that AST interpretation is identical for the two antibiotics. On the second plate, a disk for ampicillin (10 µg) was placed. The plates were incubated at 37°C in 5% CO_2 for 18-24hrs. Zones of inhibition around each disk were measured (to the nearest mm) after overnight incubation and the isolate was interpreted as ‘’sensitive’’, ‘’intermediate’’ or ‘’resistant’’ using the Clinical and Laboratory Standards Institute (CLSI) M100 guidelines (125).

3.5.1.2.2 Broth microdilution method

The Kirby-Bauer disc diffusion method was merely used for initial screening, but final AST interpretation was based on the minimum inhibitory concertation (MIC) results from the broth microdilution method. All viable GAS isolates were tested for MIC’s irrespective of zone diameters. MIC’s were determined by broth microdilution using Sensititre® HPB susceptibility plates (TREK Diagnostics Systems, West Sussex,

UK). Briefly, a 0.5 McFarland standard of each isolate was prepared in Mueller-Hinton broth (with TES buffer). A 100µl of the suspension was combined with Mueller-Hinton broth containing TES/lysed horse blood. The mixture was then poured into a sterile trough and 100 µl of the resulting mixture was transferred into each well of the Sensititre® HPB susceptibility plates coated with the following antibiotics: penicillin, ceftriaxone, erythromycin, clindamycin, chloramphenicol, tetracycline, rifampicin, cotrimoxazole, linezolid, vancomycin, ceftaroline, quinupristin-dalfopristin, meropenem, amoxicillin and levofloxacin. MIC breakpoints were interpreted as ‘’sensitive’’, ‘’intermediate’’ or ‘’resistant’’ using the Clinical and Laboratory Standards Institute (CLSI) M100 guidelines (125). Resistance breakpoints for penicillin, ceftriaxone and ceftaroline are not available in the CLSI guidelines as resistance against these antibiotics has never been clinically observed. For the purpose of this study, elevated MICs against these antibiotics were defined as 0.12µg/ml for penicillin and 0.5 for ceftriaxone and ceftaroline. These are the susceptibility breakpoints for these antibiotics according to the CLSI guidelines. Breakpoints for rifampicin, cotrimoxazole and amoxicillin are not available in the CLSI guidelines and therefore, these antibiotics were excluded from all analysis.

3.5.1.2.3 E-test method

In an effort to resolve discrepancies between the broth microdilution MIC results and genotypic antimicrobial resistance predictions, these discrepant isolates were tested using the gradient E-test method® (BioMérieux). Briefly, a 0.5 McFarland suspension of each isolate was sub-cultured on SMHA. The required antimicrobial E-test strip (containing two-fold dilutions of a predefined antibiotic on an impervious plastic strip) was applied to the agar surface. The antibiotic diffuses into the agar and after overnight incubation creates an elliptical zone of inhibition along the length of the strip. MIC interpretation was based on the CLSI guidelines (125).

3.5.2 Genotypic

3.5.2.1 Preparation of DNA extracts for whole genome sequencing

3.5.2.1.1 Pre-lysis

Isolates were retrieved from -70°C storage and a sub-culture of each isolate was prepared on 5% horse blood agar and incubated at 37°C in 5% CO_2 for 18-24hr. A sweep of the β-hemolytic colonies were suspended in Brain Heart Infusion (BHI) broth (Diagnostic Media Products, Johannesburg, South Africa) and incubated at 37°C in 5% CO_2 for 18-24hr. Roughly 1.2-1.3 ml of the broth was poured into a sterile 1.5ml Eppendorf tube and centrifuged (Hettich® Rotofix 32) for 2-3 minutes at 13 000 rpm to collect the bacterial cells at the bottom of the tube. The supernatant was discarded and the remaining bacterial pellet was re-suspended in 200µl of 10mg/ml of freshly prepared lysozyme (Sigma-Aldrich, St. Louis, MO, USA), followed by a 30 to 60-minute incubation at 37°C.

3.5.2.1.2 DNA extraction

Genomic GAS DNA for short-read whole genome sequencing was extracted manually using the QIAamp® DNA Mini kit (QIAGEN, Hilden, Germany) following the manufacturer's instructions. After the 30-60-minute incubation, 50µl of AL lysis buffer and 20µl of proteinase K was added into each sample tube and incubated at 56°C for 1 hour. 150 µl AL lysis buffer was again added into each sample tube and incubated for 10 minutes at room temperature. 250µl of 100% molecular grade ethanol was added into each sample and 700µl of this mix was transferred to spin column. Each sample was spun in a centrifuge (Hettich ® Rotofix 32) at 8000rpm for 1 minute. The collection tubes were discarded and the spin column was transferred to a new collection tube. 500µl of buffer AW1 was added into each sample tube which was spun at 8000rpm for 1 minute. The collection tubes were again discarded and the columns were transferred to a new collection tube. 500µl of buffer AW2 was added into each sample tube followed by another spin at 14 000 rpm for 3 minutes. The column tubes were discarded and the columns were transferred to a sterile 1.5 ml Eppendorf tube. 60µl of pre-warmed (56°C) elution buffer was added into each sample tube and left to incubate at room temperature for 5 minutes. Each tube was spun at 12 000rpm for 2 minutes and 60µl of eluted DNA was retained.

3.5.2.1.3 DNA quantification

Following manufacturer's instructions, the Qubit® 2.0 fluorometer (Invitrogen, Carlsbad, CA, USA) and the Qubit™ dsDNA BR assay kit (Invitrogen, Carlsbad, CA, USA) were used to quantify DNA concentrations. Briefly, 199µl of BR buffer and 1µl of BR reagent was mixed to prepare a cocktail mixture. 195µl of the cocktail was mixed with 5µl genomic DNA and incubated for a minimum of 2 minutes at room temperature. The DNA concentration of each sample was quantified using the Qubit® 2.0 fluorometer. All DNA samples with a concentration ≥1 ng/µL were sent for whole genome sequencing, whereas DNA samples with a concentration <1ng/µL were re-extracted. 55µl of the extracted genomic DNA was used for DNA library preparation, which was followed by whole genome sequencing.

3.5.2.2 Whole-genome sequencing

3.5.2.2.1 DNA library preparation

All DNA samples with a concentration between 1 - 4ng/µL were processed as is, whereas DNA samples with a concentration >4ng/µL were brought to a starting concentration of 4ng/µL by dilution with distilled water.
Using the Illumina Nextera DNA Flex Library preparation kit (Illumina, Inc., San Diego, United States) following manufacturer's instructions, each DNA sample was fragmented and tagged in order to produce adapter-ligated DNA fragments. Briefly, 7.5 µl of each DNA sample was pipetted into individual wells in a MicroAmp® 96-well PCR plate (Bio-Rad, Hercules, California, United States) and mixed with a 5µl cocktail of the bead-linked transposome (BLT) complex and tagmentation buffer (TB1). The plate was then incubated in a thermal cycler at 55°C for 15 minutes.
Next, the adapter-tagged DNA fragments underwent a washing process. Briefly, the tagmentation reaction was halted by adding 2.5µl of Tagment Stop Buffer (TSB) into each well and the plate was incubated in a thermal cycler at 37°C for 15 minutes. The plate was then placed on a magnetic stand until the solution was clear and the supernatant was removed by pipette and discarded. 25µl of Tagment Wash Buffer (TWB) was then added into each well to wash and the supernatant was again pipetted and discarded. This washing step was repeated twice.
Subsequently, the tagmented DNA fragments were amplified using a limited-cycle PCR (8 cycles) program in order to attach Illumina sequencing adapters and also create a dual indexed library for each sample. Sample purification beads were used to purify the amplified libraries (Illumina, Inc.). After the purification step, the concentration of the DNA libraries was measured using the Qubit® 4 fluorometer with the Qubit dsDNA high sensitivity assay kit. Additionally, the size of the library preparations was confirmed using the Agilent 4200 TapeStation system (Agilent Technologies, Santa Clara, California, United States).
To ensure, the quality of the DNA library preparation, a gBlock® gene fragment that also undergone Nextera DNA flex library preparation was added into the DNA library pool.

3.5.2.2.2 Next-generation sequencing

Tris-Cl (ph8.5) containing 0.1% Tween 20 was utilized to normalize DNA libraries to a concentration of 2nM. Subsequently, the DNA library samples were pooled and denatured using 0.2N sodium hydroxide (NaOH) to create single strands before loading onto the reagent cartridge. To ensure the quality of the Illumina sequencing runs, a PhiX control V3 (Illumina, Inc., San Diego, United States) which is an adapter-ligated library, was used. Whole-genome sequencing was carried out on the Illumina MiSeq® (Illumina, Inc., San Diego, United States), Illumina NextSeq® 550 (Illumina, Inc., San Diego, United States) or Illumina NextSeq 1000/2000 (Illumina, Inc., San Diego, United States) instrument. For sequencing done on the Illumina MiSeq, a 600-cycle MiSeq Reagent v3 Kit (Illumina, Inc., San Diego, United States) was used to generate paired-end reads (2 x 300 base pair) with 100X coverage. The NextSeq™ 550 v2 300 cycles kit and NextSeq 1000/2000 kit (300 Cycles) was used for sequencing runs that were done on the Illumina NextSeq® instruments to generate paired-end reads (2 X 150 base pair) with 100X coverage.

3.6 Genome analysis

3.6.1 Bioinformatics analysis

The resulting sequences were processed using a validated CDC (Centre for Disease Control and Prevention) GAS bioinformatics pipeline (122) in order to: quality check (QC) the raw reads; trim the raw reads; obtain whole genome assemblies; identify *emm* subtypes; identify multilocus sequence types (MLST); determine the presence of the *mrp* gene; determine the presence of genes commonly associated with antimicrobial resistance; and to determine strain features that are important for virulence. The CDC GAS bioinformatics pipeline used in the present study was installed and set up in a high performance computing cluster (HPCC) belonging to the University of the Witwatersrand. The original bioinformatics scripts and associated databases are available on GitHub at https://github.com/BenJamesMetcalf (created by Dr. Benjamin Metcalf from the CDC). The nextflow implementation (used in the current study) of the original

pipeline is also available on GitHub https://github.com/shaze/GS (nextflow implementation by Prof. Scott Hazelhurst from the University of Witwatersrand).

3.6.1.1 Quality check (QC), trimming and assembly

The tools used for QC, trimming and assembly are incorporated into the CDC GAS pipeline.
To check the quality of the raw reads, the pipeline used FastQC and MultiQC. Genome size (~1.8MB) and GC content (~38%) and the number of contigs (<500) were taken into consideration when evaluating the quality of the sequences. The raw reads were trimmed using Cutadapt v1.6 (126). Cutadapt removed low quality bases at the end of the read using the '-q 20 option', removed the adapters using the '-b' option and also filtered out reads that were less than 50 bases in length using the '—minimum length' option (126). VelvetOptimiser v1.2.10 was used to carry out *de novo* assembly (126).

3.6.1.2 Identification of *emm* subtypes/types,

The CDC *emm* subtype database is incorporated into the CDC GAS pipeline to screen for *emm* subtypes (122).
The database (https://ftp.cdc.gov/pub/infectious_diseases/biotech/tsemm/) consists of defined 180 base pair (bp) sequences representing different *emm* subtypes. As mentioned above, the pipeline utilizes *de novo* assembly and queries sequences in the M-type specific region (*emm* gene) against sequences in the CDC *emm* subtype database (122). Queries that obtained a perfect 180/180 match to an existing sequence on the database were assigned that same *emm* subtype. The integer part of the *emm* subtype represents the *emm* type e.g. *emm* subtype 18.21 is *emm* type 18. In the present study, *emm* types are reported. Sequences that did not obtain a perfect 180/180 match to existing *emm* subtypes, but showed high similarities to an existing *emm* subtype were reported as 'novel *emm* subtypes' and are to be submitted to the CDC database curator for quality checks and *emm* subtype assignment.

In instances where the pipeline could not confidently call for *emm* subtype, which may be due to the read quality at the specific region, the isolates were re-extracted and sequencing was repeated on the fresh extract. If the problem persisted, these isolates were designated NT (non-typable) by WGS for *emm* subtype/type as per discussions with the CDC. Targeted PCR or Sanger sequencing were not explored to resolve this.

An isolate was considered covered by the 30-valent M-protein-based vaccine if the isolate belonged to one of the 30 *emm* types (1, 2, 3, 4, 5, 6, 11, 12, 14, 18, 19, 22, 24, 28, 29, 44, 49, 58, 73, 75, 77, 78, 81, 82, 83, 87, 89, 92, 114 and 118) included in the vaccine formulation.

Cross-protection by the 30-valent vaccine was considered if the isolate belonged to one of the 24 *emm* types (8, 9, 15, 30, 33, 36, 40, 48, 51, 52, 59, 65, 66, 68, 76, 79, 85, 94, 97, 103, 105, 109, 111 and 122) shown to be cross-opsonized (>40% bactericidal killing activity) in previous animal studies (57).

emm types were assigned to *emm* clusters based on the Sanderson-smith *et al*., 2014 publication (119).

3.6.1.3 Detection of M related protein (Mrp)

Presence of the Mrp was screened for by the CDC GAS pipeline by detecting the presence of the *mrp* gene. An isolate was considered to be covered by the theoretical Mrp based vaccine candidate if the *mrp* gene was detected.

3.6.1.4 Identification of multilocus sequence types (MLST)

The sequence type (ST) of each isolate was determined by the allelic profile of seven housekeeping genes (*gki*, *gtr*, *murI*, *mutS*, *recP*, *xpt*, and *yqiL*). The short read typing tool SRST2 (127) and the PubMLST database, both incorporated into the CDC GAS pipeline, were used to assign ST's. This was done by extracting the sequences of the 7 housekeeping genes and comparing them to previously described ST's on the GAS MLST web server (http://pubmlst.org/spyogenes/). Newly identified allelic profiles were submitted to the MLST GAS database curator for approval and ST assignment. Sequencing was repeated for isolates that could not

be assigned an ST by the pipeline as a result of poor quality during base-calling on one of the 7 housekeeping genes. If the problem persisted, these isolates were designated NT for sequence type. PHYLOViZ v2.0 was used to determine the genetic relatedness among isolates based on ST and construct a minimum spanning tree (128).

3.6.1.5 Detection of antimicrobial resistance (AMR) determinants

Common streptococcal antimicrobial resistance determinants were screened for by the CDC GAS pipeline. A CDC in-house sequence database that detects 21 targets (122), the ResFinder sequence database that detects 1913 targets (129) and the ARG-ANNOT sequence database that detects 1653 targets (130) are incorporated into the pipeline for this. For this study, the four common antimicrobial resistance genes analysed were *erm* (presence predicts macrolide and clindamycin resistance); *mef* (presence predicts resistance to macrolides); *tet* (presence predicts resistance to tetracycline) and mutations in the *parC/gyrA* region (predicts differing levels of resistance to fluoroquinolones).

A pbp2x typing scheme which detects substitutions within the transpeptidase region of pbp2x is incorporated into the pipeline (122,126). In order to extract the pbp2x sequence fragments from the samples paired-end sequence files, a query file containing representative fragments of the pbp2x gene was passed through the lo_trac gene program (126). The pbp2x gene fragments were extracted and blasted against a CDC in-house database (126). An existing allele number was assigned to the sequence should there be a perfect match (percent identity= 100%, alignment length= query length) to the CDC in-house reference database. A new allele was designated ‘New’ and will be shared with the CDC for quality checks and assignment. pbp2x sequences that could not be extracted due to poor quality at the region were designated ‘Not determined’ (126).

The genomic antimicrobial resistance prediction results were compared to the phenotypic antibiotic susceptibility patterns. For samples with discordant results, due to time constraints, only phenotypic testing was repeated for available antimicrobials.

3.6.1.6 Detection of strain features important for virulence

The CDC GAS pipeline screened for the presence/absence of genes encoding features important for virulence: the hyaluronic acid capsule (*hasA*); enn protein (*enn*); four fibronectin-binding domain repeat proteins [Fibronectin-binding protein of group A streptococci type A (*fbaA*), Protein F2/*S. pyogenes* Fn-binding protein (*prtF2*), *S. pyogenes* Fn-binding protein I(*sfb1*), Serum opacity factor/*S. pyogenes* Fn-binding protein II(*sof*)]; R28 protein (*spr28*); streptodornase 1 (*sda1*); two mutations in the regulatory protein RocA; fourteen exotoxin genes (*speA*, *speC*, *speG*, *speH*, *speI*, *speJ*, *speK*, *speL*, *speM*, *speN*, *speO*, *speP*, *ssa* and *smeZ*); streptococcal inhibitor of complement (*sic*) and an upregulated *nga* operon that encodes NADase and streptolysin O (*pnga*3).

3.6.2 Phylogenetic analysis

Sequences with >500 contigs were excluded to ensure that only high quality assemblies were included (n=577) to construct the phylogenetic trees. kSNPv3 (http://ksnp.sourceforge.net/) was used to perform a single nucleotide polymorphisms (SNP) analysis. kSNP, an alignment-free tool that does not require a reference genome, is used to detect SNPs and construct phylogenies of complete microbial genome sequences based on k-mer analysis (131). In the present study, assemblies were used to generate a core SNPs matrix and the detected SNPs were then used to construct the phylogenies. A k-mer length of 19 was used and 100% of the genomes were required to possess a nucleotide at a given SNP in order for the SNP to be considered a core. Phylogenetic trees were constructed by the RAxML (132) tool using 100 bootstrap, with the maximum-likelihood (ML) method based upon the general time-reversible (GTR) substitution model and gamma distribution. Phylogenetic trees along with the metadata were visualized using the Interactive Tree of Life (iTOL) online (https://itol.embl.de/) software (iTOLv6) (133). The phylogenetic analysis determined genetic clusters.

3.7 Data analysis

Descriptive statistics were used to summarize patients' characteristics: province, gender, race, age, hospitalization final outcome and site of infection. Percentages were used for quantitative data. This was all performed using formulas on Microsoft Excel.

The proportion of *emm*95 isolates resistant to tetracycline were compared to the proportion of *emm*95 isolates resistant to erythromycin using the Two-sample proportion test on Stata v18.0

4 RESULTS

4.1 Invasive Group A streptococcal disease in South Africa 2019 – 2020

From 1 January 2019 - 31 December 2020, 1487 iGAS disease cases were recorded (**Figure 5**) through our active, national, laboratory-based surveillance, GERMS-SA. Of the 1487 cases, isolates were not received for 847 (57.0%) cases. Cases detected through routine quarterly retrospective audits of the laboratory information system (LIS) TrakCare accounted for 838 (98.9%) of the cases without isolates. These cases are not expected to have associated isolates. Nine cases (9/847; 1.1%) were recorded as 'missing' on the GERMS-SA database as the sending laboratories communicated at a later stage that an isolate was obtained and should have been sent but was unfortunately not. These isolates could no longer be located on their side.

At the NICD's CRDM, 640/1487 (43.0%) isolates were received for confirmation and further phenotypic and genotypic characterization. Of the 640 isolates, 622 (97.2%) were viable on culture and presumptively identified as GAS by phenotypic methods and 18 (2.8%) isolates were non-viable when attempts were made to culture them for further characterization. Whole genome-sequencing was done on 596 (95.8%) of the 622 viable isolates. Twenty-six (4.2%) isolates were not sequenced due to time constraints. Only specimens that passed the quality check (577/596; 97.0%) were included in the downstream genome analysis.

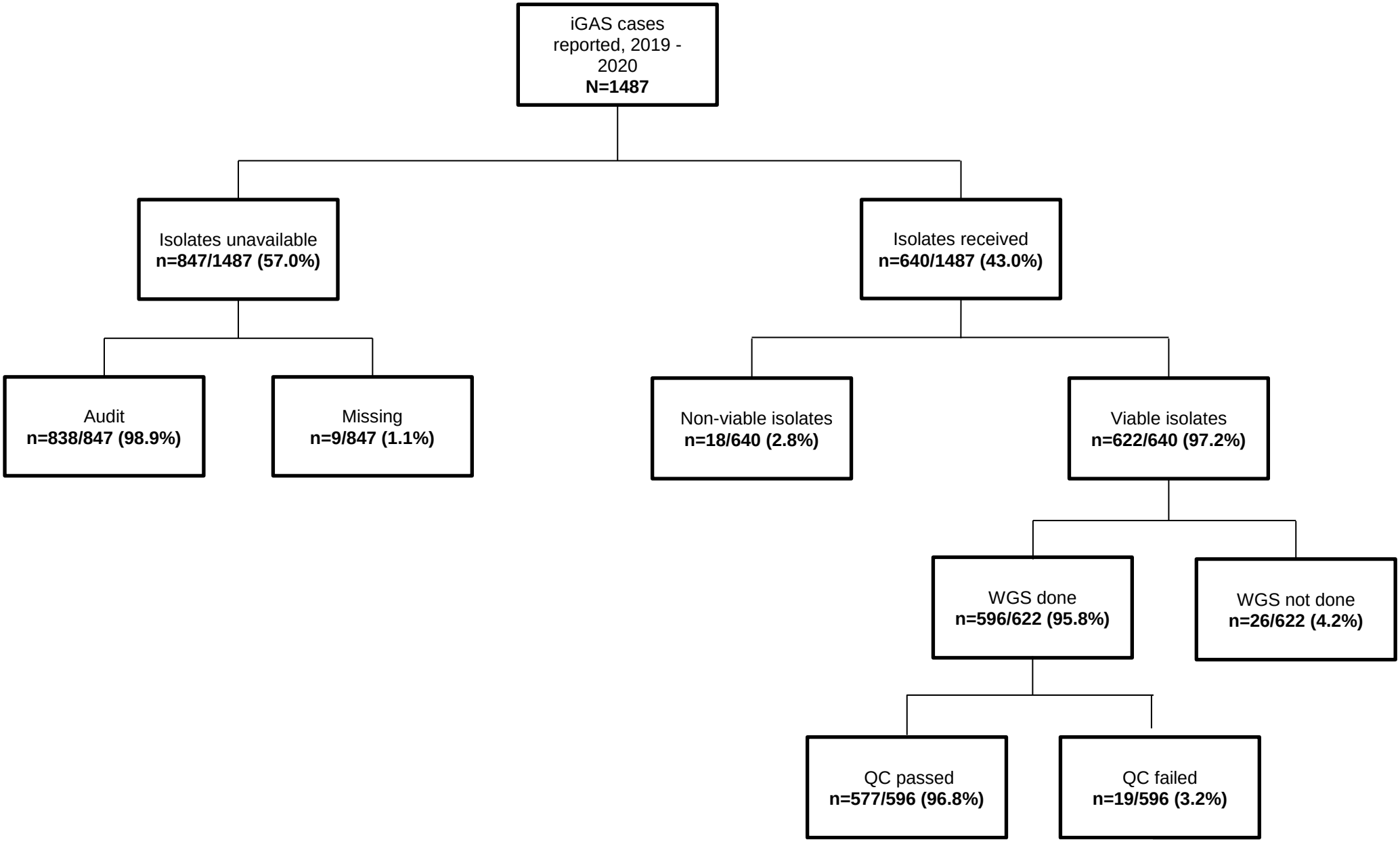


Figure 5: Invasive group A streptococcal disease cases recorded in South Africa, GERMS-SA, 2019-2020 (N=1487)

4.2 Demographic and clinical characteristics of cases presenting with invasive GAS disease

Patient characteristics are illustrated in **Table 4**. Gauteng, Kwa-Zulu Natal, Western Cape and the Eastern Cape recorded 94.0% (1398/1487) of the total cases. The Western Cape reported 45.7% (679/1487) of the total cases. Where gender was known, 53.6% (785/1464) of iGAS disease occurred in males. For cases where age was known (n=1408), the median age was 36 years and age ranged from 0 – 94 years. Thirty-four percent (499/1408) of cases occurred in individuals aged 25 - 44 years old, 22.8% (339/1408) of the cases were reported in individuals 45 – 64 years old and young children aged 0 – 4 years old accounted for 12.9% (193/1408) of the cases. **Figure 6** illustrates that in both 2019 and 2020, in individuals where age was known, the incidence of iGAS disease was highest in infants aged <1 years old (4.4 cases/100,000 persons in 2019 and 2.1 cases/100,000 persons in 2020) and the elderly aged ≥65 years old (3.0 cases/100,000 persons in 2019 and 1.9 cases/100,000 persons in 2020). Race data was available for only 105 cases and black individuals accounted for 73.3% (77/105) of these cases. Black individuals in Gauteng (36.4%; 28/77) made up most of these cases, followed by black individuals in Kwa-Zulu Natal (16.9%; 14/77), Eastern Cape (16.9%; 14/77) and the Western Cape (16.9%; 13/77). Coloured individuals accounted for 21.9% (23/105) of the cases and most of these cases were recorded in the Western Cape (52.2%; 12/23) and the Eastern Cape (43.5%; 10/23). Where final in-hospital outcome data was available, 76.4% (107/140) of patients recovered from iGAS disease, 20.7% (29/140) of the cases died and 0.3% (4/140) of patients refused hospital treatment and thus final outcome was unknown.

Table 4: Demographic and clinical characteristics of case patients presenting with invasive group A streptococcus disease in South Africa, GERMS-SA, 2019-2020 (N=1487)

Characteristic	2019 (n=971)		2020 (n=516)		Total (N=1487)	
	n	%	n	%	n	%
Province						
Gauteng	198	20.4	96	18.6	294	19.8
Kwa-Zulu Natal	162	16.7	49	9.5	211	14.2
Western Cape	415	42.7	264	51.2	679	45.7
Mpumalanga	11	1.1	11	2.1	22	1.5
Limpopo	10	1.0	5	1.0	15	1.0
Northern Cape	7	0.7	7	1.4	14	0.9
Free State	22	2.3	11	2.1	33	2.2
North West	2	0.2	3	0.6	5	0.3
Eastern Cape	144	14.8	70	13.6	214	14.4
Gender						
Male	490	50.5	295	57.2	785	52.8
Female	462	47.6	217	42.1	679	45.7
Unknown	19	2.0	4	0.8	23	1.5
[a]Race						
Black	3	0.3	74	14.3	77	5.2
White	0	0	3	0.6	3	0.2
Coloured	1	0.1	22	4.3	23	1.5
Asian	0	0	2	0.4	2	0.1
Unknown	967	99.6	415	80.4	1382	92.9
Age category (years)						
<1	50	5.1	25	4.8	75	5.0
1-4	90	9.3	28	5.4	118	7.9
5-9	33	3.4	10	1.9	43	2.9
10-14	20	2.1	10	1.9	30	2.0
15-24	81	8.3	49	9.5	130	8.7
25-44	328	33.8	171	33.1	499	33.6
45-64	233	24.0	106	20.5	339	22.8
≥65	104	10.7	70	13.6	174	11.7
Unknown	32	3.3	47	9.1	79	5.3
[b] Final in-hospital outcome						
Died	1	0.1	28	5.4	29	2.0
Recovered	3	0.2	104	20.2	107	7.2
Refused hospital treatment	0	0	4	0.8	4	0.3
Unknown	967	99.6	380	73.6	1347	90.6

[a]Data on race was available for only 4 cases that occurred in 2019 and 101 cases that occurred in 2020.

[b] Data on final in-hospital outcome was available for only 4 cases that occurred in 2019 and 136 cases that occurred in 2020. The collection of additional data using the case report form (CRF) was only implemented in 2020, hence the improvement of data collection compared to 2019.

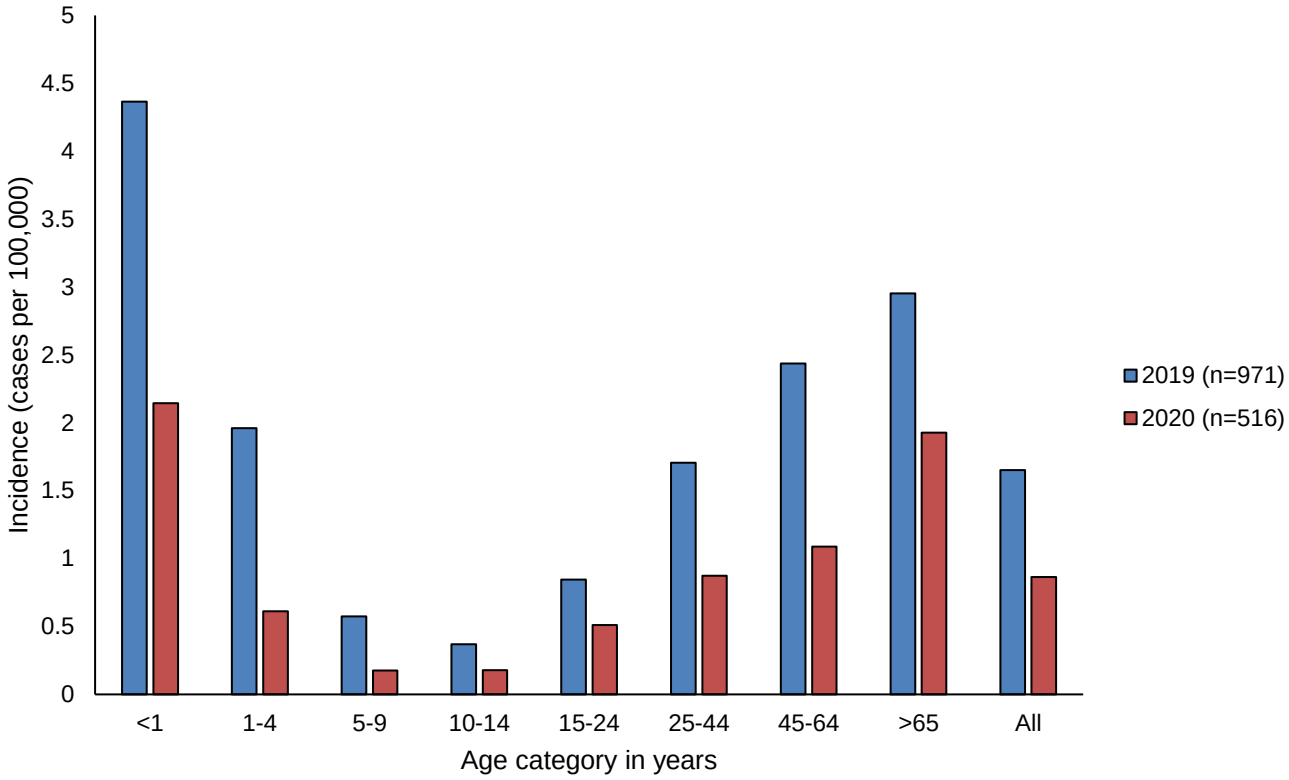


Figure 6: Incidence of invasive group A streptococcus disease by age-category, South Africa, GERMS-SA, 2019 – 2020 (N=1408).
"All" bar also includes all individuals including where age was not known (N=1487)

Table 5 illustrates that blood (51.9%, 770/1487) was the most common specimen type where iGAS was detected. Other body sites where iGAS was detected included joint/joint fluid (11.0%, 163/1487), bone (4.2%, 62/1487) and pleural fluid (4.0%, 59/1487). Fine needle aspirate/biopsy/tissue specimens collected from various body sites accounted for 18.1% (267/1487) of the cases.

Table 5: Site of specimen collection in patients presenting with invasive group A streptococcus disease in South Africa, GERMS-SA, 2019-2020 (N=1487)

Site of infection	n	%
Blood	770	51.8
[a]Fine needle aspirate/ [b]Biopsy/ [c]Tissue	267	18.0
Joint/Joint fluid	163	11.0
Bone	62	4.2
Pleural fluid	59	4.0
[d]Necrotic wound/tissue (necrotizing fasciitis)	40	2.7
Central nervous system	34	2.3
Peritoneum/Ascitic fluid	31	2.1
Lung	18	1.2
Endocervical	12	0.8
Lymph node	11	0.7
[e]Deep-seated abscess	10	0.7
[f]Septic wound	5	0.3
Product of conception	2	0.1
[g]Sterile cavity	2	0.1
Pericardial fluid/Heart	1	0.1

[a]Anatomical site: Neck= 3; Unknown=14

[b]Anatomical site: Unknown=8

[c]Anatomical site: Abdomen=2; Back=1; Breast=8; Chest=1; Foot=1; Lower limb=22; Upper limb=18; Groin=2; Kidney=1; Neck=1; Ocular=1; Penis=2; Scrotum=1; Shoulder=1; Upper respiratory tract=2; Unknown=178

[d]Anatomical site: Fournier's gangrene=1; Breast=1; Buttock=1; Cutaneous=2; Face=1; Lower limb=10; Ocular=1; Penis=1; Unknown=13; Upper limb=9

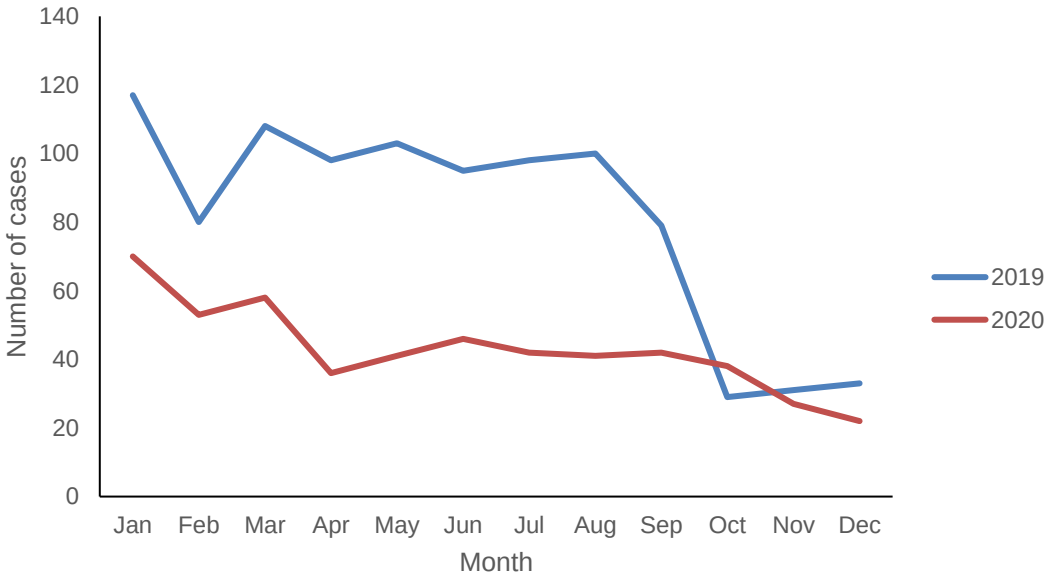
[e]Anatomical site: Breast=1; Lower limb=1; Neck=1; Unknown=7

[f]Anatomical site: Lower limb= 1 Unknown=4

[g]Anatomical site: Unknown=1

4.3 Distribution of invasive GAS disease cases

In 2019, more iGAS disease cases were reported (65.3%, 971/1487) in comparison to 2020 (34.7%, 516/1487) (**Table 4, Figure 7**). Six provinces (excluding Mpumalanga, Northern Cape and North West) reported fewer cases in 2020 than what was reported in 2019 (data not shown). The biggest decline in the number of cases was seen in the Western Cape, Kwa-Zulu Natal, Gauteng and the Eastern Cape which reported 151, 113, 102 and 74 fewer cases, respectively, than the preceding year. **Figure 7** illustrates the number of iGAS cases reported for each month in 2019 and 2020. The highest number of cases were reported in January 2019 (117 cases), March 2019 (108 cases), May 2019 (103) and August 2019 (100 cases). The least number of cases were reported in December 2020 (22 cases). There was a sharp decline in the number of cases reported from October 2019 in comparison to the preceding months of that year. The number of cases remained relatively stable throughout 2020. Most cases (866 cases; 444 in autumn and 422 in winter) occurred in the winter and autumn months (March – August).



Jan- January; Feb- February; Mar- March; Apr- April; Jun- June; Jul- July; Aug- August; Sep- September; Oct- October; Nov- November; Dec- December

Figure 7: Distribution of invasive group A streptococcus cases by month in South Africa, GERMS-SA, 2019-2020 (N=1487)

4.4 Antimicrobial susceptibility

4.4.1 Phenotypic antimicrobial susceptibility patterns

Phenotypic antimicrobial susceptibility testing (AST) results were available for 618/622 (99.4%) of the viable isolates. One isolate could not be located when the samples were retrieved from storage for testing. Penicillin and ceftriaxone results were pending for three isolates and thus these isolates were excluded. The AST results are presented in **Table 6**. Based on the CLSI susceptibility breakpoints, all of the study isolates tested phenotypically susceptible to penicillin, ceftaroline, ceftriaxone, clindamycin, linezolid, vancomycin and quinupristin-dalfopristin. The highest phenotypic resistance was seen against tetracycline (10.5%, 65/618) followed by erythromycin (3.6%, 22/618). Only one (0.2%) isolate was resistant to chloramphenicol and another one (0.2%) was resistant to levofloxacin. Eight isolates displayed resistance against two antibiotics as follows: one isolate displayed resistance against both erythromycin and chloramphenicol and seven isolates displayed resistance against erythromycin and tetracycline.

Table 6: Antimicrobial susceptibility patterns of invasive group A streptococcus isolates based on phenotypic minimum inhibitory concentrations (MIC), South Africa, GERMS-SA, 2019-2020 (n=618)

Antimicrobial agent	Number of isolates (%)		
	Sensitive	Intermediate	Resistant
Penicillin	618 (100.0)	0 (0.0)	0 (0.0)
Ceftaroline	618 (100.0)	0 (0.0)	0 (0.0)
Ceftriaxone	618 (100.0)	0 (0.0)	0 (0.0)
Erythromycin	596 (96.4)	0 (0.0)	22 (3.6)
Clindamycin	618 (100.0)	0 (0.0)	0 (0.0)
Chloramphenicol	617 (99.8)	0 (0.0)	1 (0.2)
Tetracycline	548 (88.7)	5 (0.8)	65 (10.5)
Linezolid	618 (100.0)	0 (0.0)	0 (0.0)
Vancomycin	618 (100.0)	0 (0.0)	0 (0.0)
Quinupristin-dalfopristin	618 (100.0)	0 (0.0)	0 (0.0)
Levofloxacin	617 (99.8)	0 (0.0)	1 (0.2)

4.4.2 Major antimicrobial resistance determinants

Of the 622 viable isolates received at CRDM, only 577 (92.8%) isolates were included in the downstream genome analysis. The 45 (7.2%) isolates that were excluded from this analysis were due to the whole genome-sequencing data not being available as a result of time constraints (26 isolates) or the resulting sequences not passing quality checks (19 isolates). The four isolates excluded in the phenotypic AST results in section 4.4.1 were also excluded from the current analysis as they fell into one of the two exclusion criteria mentioned above.

Four common antimicrobial resistance (AMR) genes/mutations (*erm*, *mef*, *tet* and mutations in the *parC_gyrA* region) were analysed in this study and the results are presented in **Figure 8** and **Table 7**. Amongst the 577 isolates that were screened for

these AMR genes, the most frequently detected resistance determinant was the *tet* family gene which was detected in 59 (10.2%) isolates, followed by the *mef* gene detected in 14 (2.4%) isolates and finally the *erm* family gene which was detected in 6 (1.0%) isolates. Mutations in the *parC* region were detected in 7 (1.2%) isolates and no isolates carried mutations in the *gyrA* region. More than one AMR determinant was detected in seven isolates as follows: both the *erm* and *tet* genes were detected in three isolates, both the *mef* and *tet* genes were detected in three isolates and the *mef* gene and a mutation in the *parC* region was detected in one isolate. The *tet* family genes detected in the study isolates were *tetM* (n=54), *tetO* (n=2), a combination of *tetL*: *tetM* (n=2) and unassigned *tet* (n=1). The *erm* family genes detected included *ermTR* (n=4), *ermT* (n=1) and unassigned *erm* (n=1). The most common mutation in the *parC* region was the *parC*-S6A substitution (n=4). No isolates carried mutations in both the *parC_gyrA* region.

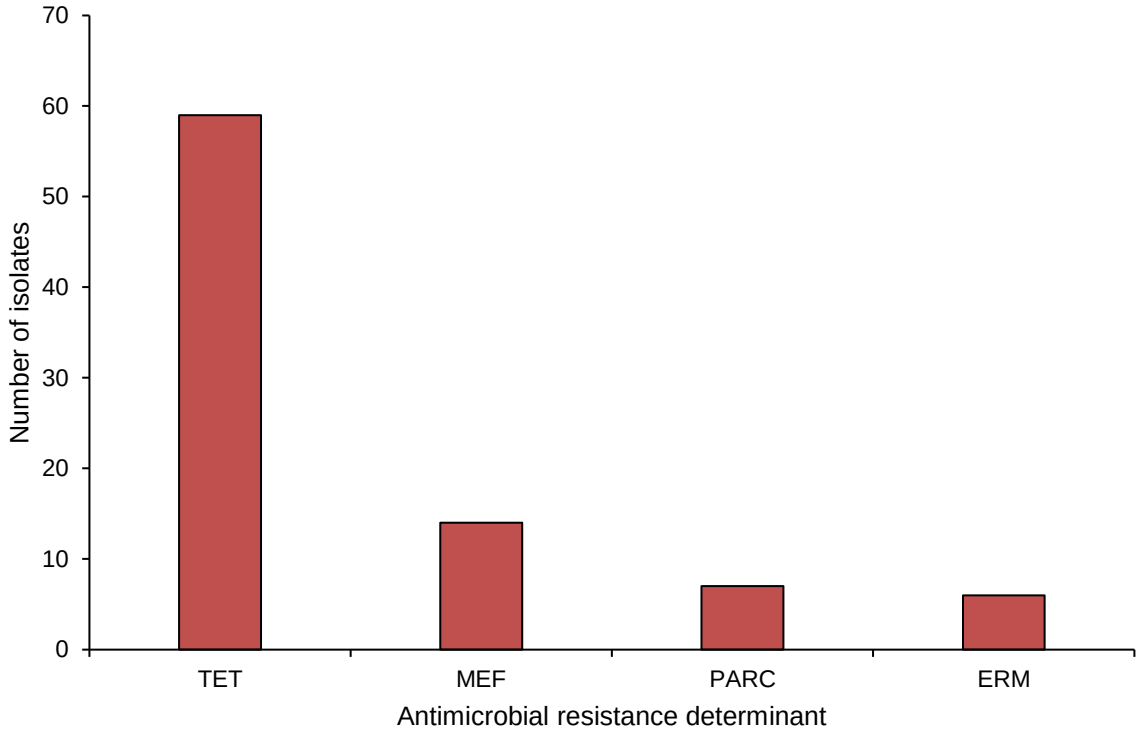


Figure 8: Group A streptococcus invasive isolates positive for the ERM, MEF, TET targets or mutations in the parC_gyrA targets associated with macrolide, tetracycline and fluoroquinolone resistance. South Africa, GERMS-SA, 2019-2020 (n=577)

There were some discordant results between the phenotypic and genotypic testing and therefore, the isolates (excluding those with discrepant fluoroquinolone results as the laboratory had a shortage of the levofloxacin gradient diffusion strips at the time) were tested using the E-test and re-tested using the Kirby- Bauer disc diffusion method. The repeat results correlated with the initial phenotypic results, but repeat sequencing is still pending. We have presented the current discrepancies while awaiting final resolution. The results are presented in **Table 7**. The *erm* gene was detected in one isolate that was phenotypically sensitive to erythromycin and clindamycin. The *erm*/*mef* gene was not detected in two isolates that tested phenotypically resistant to erythromycin. The *tet* gene was detected in two isolates that were phenotypically sensitive to tetracycline whereas this gene was not detected in four isolates that displayed resistance to tetracycline. Mutations in the *parC* region were seen in seven isolates (MIC range: 1 - 2 µg.ml) that all displayed sensitivity to levofloxacin: five isolates had an amino acid alteration in site 6, which included S6F (n=1) and S6A (n=4); one isolate had an alteration as site 10 (D10N) and another isolate had an amino acid alteration at site 5 (D5N). One isolate was resistant to levofloxacin but no mutation in the *parC_gyrA* region was detected.

Table 7: Common antimicrobial resistance determinants associated with macrolide, tetracycline and fluoroquinolone resistance detected among invasive group A streptococcus isolates in South Africa, GERMS-SA, 2019-2020 (n=577)

AMR determinant	Antibiotic								
	Erythromycin			Tetracycline			Levofloxacin		
	S (n=556)	I (n=0)	R (n=21)	S (n=516)	I (n=4)	R (n=57)	S (n=576)	I (n=0)	R (n=1)
Macrolides									
erm	**1**	0	0						
ermT	0	0	1						
ermTR	0	0	4						
Mef	0	0	14						
None	555	0	**2**						
Tetracycline									
Tet				0	0	1			
TetM				**2**	4	48			
TetO				0	0	2			
TetL:TetM				0	0	2			
None				514	0	**4**			
Fluoroquinolones									
ParC-S6f							**1**	0	0
ParC-D10N							**1**	0	0
ParC-S6A							**4**	0	0
ParC-D5N							**1**	0	0
None							569	0	**1**

Bold writing indicates discrepancy between phenotypic and genotypic results

S=Sensitive; I=Intermediate; R=Resistant

Although all isolates tested phenotypically susceptible to the β- lactam antibiotics penicillin, ceftriaxone and ceftaroline, various pbp2x sequence types were detected amongst these isolates (**Table 8**). The most common pbpx sequence type detected in 501/577 (86.8%) isolates was the wild type pbp2x-1, followed by pbp2x-8 detected in 19 (3.3%) isolates and pbp2x-10 detected in 14 (2.4%) isolates. Five isolates that were sequence type pbp2x-1 phenotypically displayed elevated MICs to penicillin and ceftriaxone. All five isolates had an MIC of 0.12µg/ml for penicillin. One isolate had an MIC of 0.12µg/ml for penicillin and an MIC of 0.5µg/ml for ceftriaxone. These MIC's were at the cusp of susceptibility for penicillin and ceftriaxone. One isolate that was pbp2x-10 also phenotypically displayed an MIC of 0.12µg/ml for penicillin. Twenty-two isolates (3.8%) represented unassigned (potentially new) pbp2x sequence types that require submission to the CDC for further quality checks and sequence type assignment. The MIC ranges of these isolates for penicillin and ceftriaxone were 0.006 - 0.023µg/ml and 0.015 – 0.06 µg/ml, respectively. All the isolates had an MIC of 0.03 µg/ml for ceftaroline.

The pbp2x sequence type could not be determined for 10 isolates (1.7%) due to unsatisfactory sequence quality in the pbp2x region. The MIC ranges of these isolates for penicillin and ceftriaxone were 0.004 -0.015 µg/ml and 0.03 – 0.06 µg/ml, respective. All the isolates had an MIC of 0.03 µg/ml for ceftaroline.

Table 8: PBP2x sequence types detected among invasive group A streptococcus isolates in South Africa, GERMS-SA, 2019-2020 (n=577)

PBP2x type	No. of isolates (%)	[a]MIC range (µg/ml)	[b]MIC50 (µg/ml)	PBP2x substitution
1	501 (86.8)	0.004 – 0.12	0.015	Wild type
8	19 (3.3)	0.008 – 1.015	0.008	I502V, P676S
10	14 (2.4)	0.008 - 0.12	0.015	**M593T**
9	6 (1.0)	0.004 – 0.015	0.008	S562T, P676S
13	2 (0.3)	0.008 – 0.015	-	D353A
2	2 (0.3)	0.008 – 0.015	-	D353A, P676S
3	1 (0.2)	0.015	-	S562T
*NEW	22 (3.8)	0.006 – 0.023	0.008	-
°Not determined	10 (1.7)	0.004 – 0.015	0.015	-

Boldface indicates pbp2x substitutions previously reported by the CDC to be associated with elevated MIC values in comparison to their baseline wild-type MICs (134)

[a]MIC range against penicillin

[b]MIC50 against penicillin

*NEW – Unassigned pbp2x sequence types to be submitted to the CDC

°Not determined – Unassigned pbp2x sequence types due to unsatisfactory sequence quality at the pbp2x region

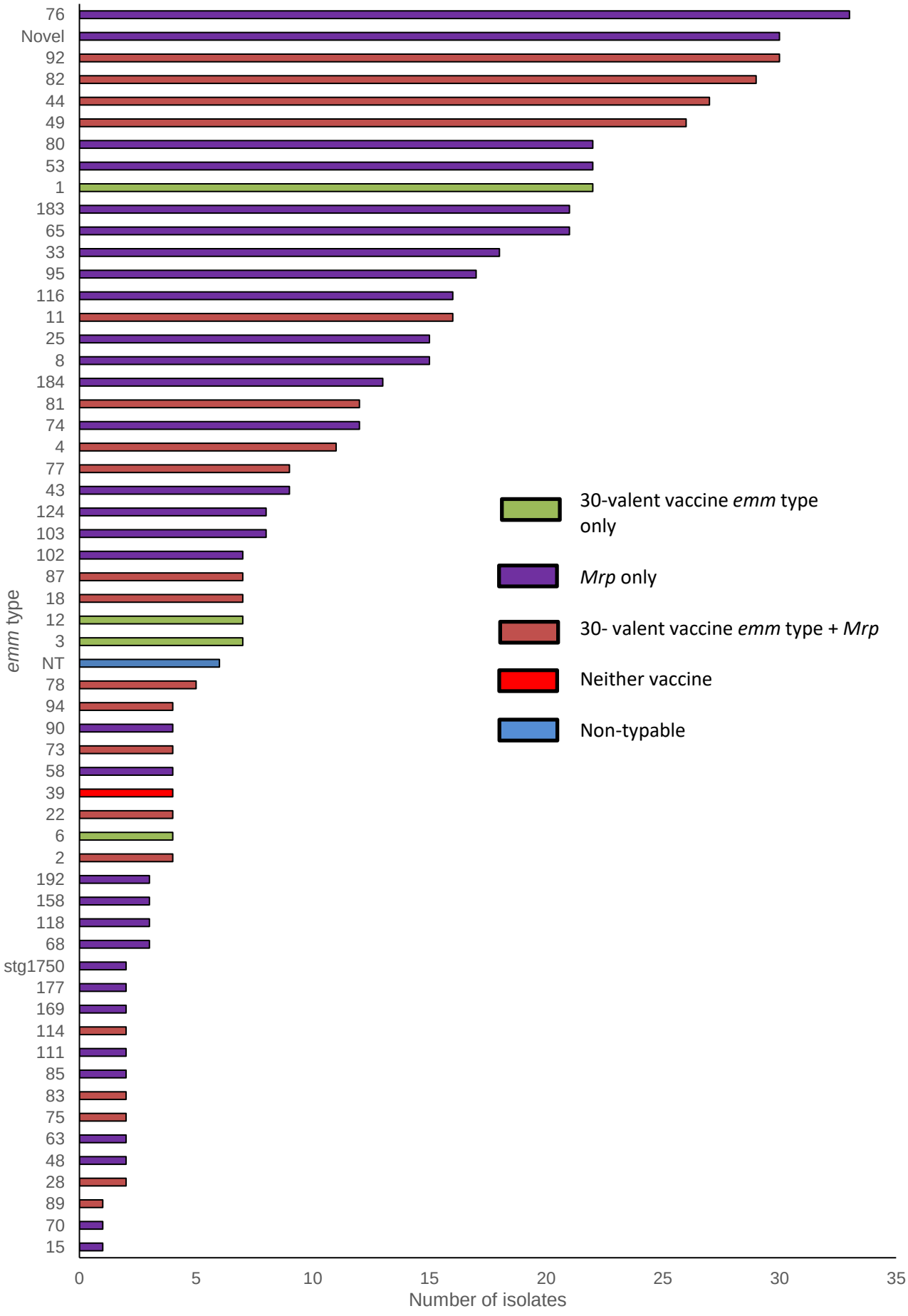
4.5 Potential coverage of a combined 30-valent M-protein-based vaccine and theoretical Mrp protein vaccine

From the 577 invasive isolates that were typed in this study, 70 *emm* subtypes were identified and from this, 56 *emm* types were assigned (**Figure 9A**). For 30 isolates, the *emm* subtype could not be determined as the sequences were not a perfect 180/180 match to any of the existing *emm* subtypes in the CDC database. Instead, the pipeline assigned these isolates an *emm* subtype it closely resembled with an asterisk (e.g. 58.3*). This meant that the isolate in question was similar to *emm* subtype 58.3* but not identical and thus may represent a novel *emm* subtype as no other hits were found. In this document, these isolates are referred to as potentially novel *emm* subtype/type and were not included when calculating the potential vaccine coverage as the *emm* subtype must be known in order to assign an *emm* type. The sequences for these 30 isolates will be submitted to the CDC for further quality checks and *emm* subtype assignment. Six isolates did not type even when

the sequencing was repeated for these isolates. After investigations and discussions with the CDC, we were advised to report these isolates as non-typable.

The number of isolates within each *emm* type ranged from 1 – 33 isolates. The 10 most common *emm* types detected in this study were (in descending order): *emm*76 (33/577, 5.7%), *emm*92 (30/577, 5.2%), *emm*82 (29/577, 5.0%), *emm*44 (27/577, 4.7%), *emm*49 (26/577, 4.5%), *emm*80 (22/577, 3.8%), *emm*53 (22/577, 3.8%), *emm*1 (22/577, 3.8%), *emm*183 (21/577, 3.6%) and *emm*65 (21/577, 3.6%). These *emm* types accounted for 43.8% (253/577) of the isolate collection. Two isolates belonged to *emm*/*st*G1750, a sequence type that has been previously isolated from a group G streptococcal isolate, but in this case the isolates were *S. pyogenes* as they were positive for the group A carbohydrate (97,116). Only 24 *emm* types in this study collection (*emm*92, *emm*44, *emm*82, *emm*49, *emm*1, *emm*11, *emm*4, *emm*81, *emm*77, *emm*3, *emm*12, *emm*87, *emm*18, *emm*94, *emm*22, *emm*73, *emm*2, *emm*6, *emm*78, *emm*28, *emm*75, *emm*83, *emm*114 and *emm*89) which accounted for 244 isolates are included in the experimental 30-valent M-protein-based vaccine. This implies that the 30-valent vaccine would theoretically cover 44.6% (244/547) of the isolates in this study. An additional 108 isolates in this study are *emm* types that have displayed cross-protection in animal studies, which further implies that potential coverage by the 30-valent vaccine could increase to 64.4% (352/547). As depicted in **Figure 9B**, Mrp only was detected in 323/577 (56.0%) isolates, a 30-valent vaccine *emm* type only was detected in 42/577 (7.3%) isolates and 205/577 (35.5%) isolates had both Mrp and a 30-valent vaccine *emm* type identified. Neither Mrp nor a 30-valent vaccine *emm* type was detected in 7/577 (1.2%) isolates. A theoretical vaccine that targets both components (M + Mrp) would theoretically cover 98.8% (570/577) of the isolates in this collection.

A


76
Novel
92
82
44
49
80
53
1
183
65
33
95
116
11
25
8
184
81
74
4
77
43
124
103
102
87
18
12
3
NT
78
94
90
73
58
39
22
6
2
192
158
118
68
stg1750
177
169
114
111
85
83
75
63
48
28
89
70
15
emm type
0
5
10
15
20
25
30
35
Number of isolates
30-valent vaccine emm type only
Mrp only
30- valent vaccine emm type + Mrp
Neither vaccine
Non-typable

B

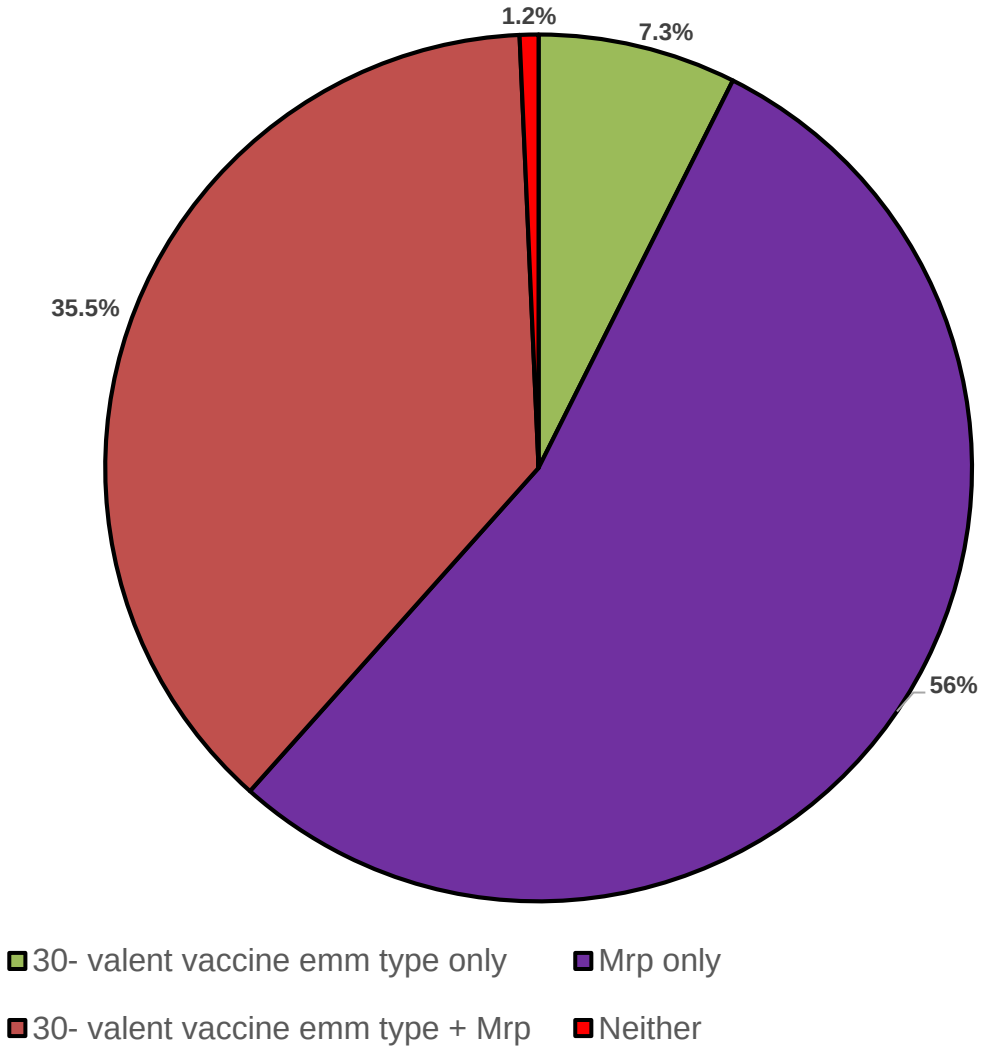


Figure 9: A. The *emm* type distribution among 577 invasive group A streptococcus isolates collected through GERMS-SA, 2019 - 2020

NT: Non-typable

Blue bar indicates NT isolates where 4/6 carried MRP only

B. The distribution of the M and M-related protein amongst 577 invasive group A streptococcus isolates collected through GERMS-SA, 2019 - 2020

Table 9 illustrates the distribution of study isolate *emm* types into *emm* clusters. Thirty-six isolates are not included in the table as these isolates did not *emm* subtype (novel *emm* subtypes and NT's). The isolates in this study collection belonged to 15 different *emm* clusters (E1, E2, E3, E4, E6, A-C3, A-C4, A-C5, D4, D5, *emm*6, *emm*18, *emm*74, *emm*95 and *emm*111) and >75% (428/541) of the isolates clustered into 5 *emm* clusters (E2, E3, E4, E6 and D4). Five *emm* types (*emm*6, *emm*18, *emm*74, *emm*95 and *emm*111) belonged to single M-protein *emm* clusters and two isolates (two *st*G1750 isolates) did not belong to any of the previously described 16 *emm* clusters since these *emm* type were not part of the 173 *emm* types that were tested to create the *emm* cluster system (119).

Table 9: The clustering of 541 invasive group A streptococcus isolates into *emm* clusters. South Africa, GERMS-SA, 2019-2020

Clade	***emm* cluster**	***emm* type (No. of isolates)**	**No. of isolates (%)**
X	E1	4 (11), 78 (5)	16 (3.0)
X	E2	68 (3), 76 (33), 90 (4), 92 (30)	70 (12.9)
X	E3	15 (1), 25 (15), 44 (27), 49 (26), 58 (4), 82 (29), 87 (7), 103 (8), 118 (3), 183 (21)	141 (26.1)
X	E4	2 (4), 8 (15), 22 (4), 28 (2), 73 (4), 77 (9), 89 (1), 102 (7), 114 (2), 124 (8), 169 (2)	58 (10.7)
X	E6	11 (16), 48 (2), 63 (2), 65 (21), 75 (2), 81 (12), 85 (2), 94 (4), 158 (3), 177 (2)	66 (12.2)
Y	A-C3	1 (22)	22 (4.1)
Y	A-C4	12 (7), 39 (4)	11 (2.0)
Y	A-C5	3 (7)	7 (1.3)
Y	D4	33 (18), 43 (9), 53 (22), 70 (1), 80 (22), 83 (2), 116 (16), 192 (3)	93 (17.2)
Y	D5	184 (13)	13 (2.4)
Y	Single M-protein *emm* cluster	6 (4), 18 (7), 74 (12)	23 (4.3)
Outlier	Single M-protein *emm* cluster outlier	95 (17), 111 (2)	19 (3.5)
N/A	Not assigned	Stg1750 (2)	2 (0.4)

In blood specimens, 53 of the 56 *emm* types were detected. The three *emm* types that were not detected in blood specimens seem to be rare *emm* types as each of these *emm* types were detected in only two other specimen types. Eleven *emm* types were unique to blood specimens i.e. only detected in blood specimens and not any other specimen type. The distribution of *emm* types by site of specimen collection is presented in **Figure 10**. Only 20 of the most prevalent *emm* types that accounted for >10 isolates are illustrated (n = 398). The specimen types where the most *emm* types were isolated were blood, joint/joint fluid, tissue/fine needle aspirate, bone and pleural fluid where 20, 18, 15, 11 and 10 *emm* types were isolated respectively. *emm* types did not appear to aggregate by specimen type i.e. *emm* types were not restricted to certain specimen types.

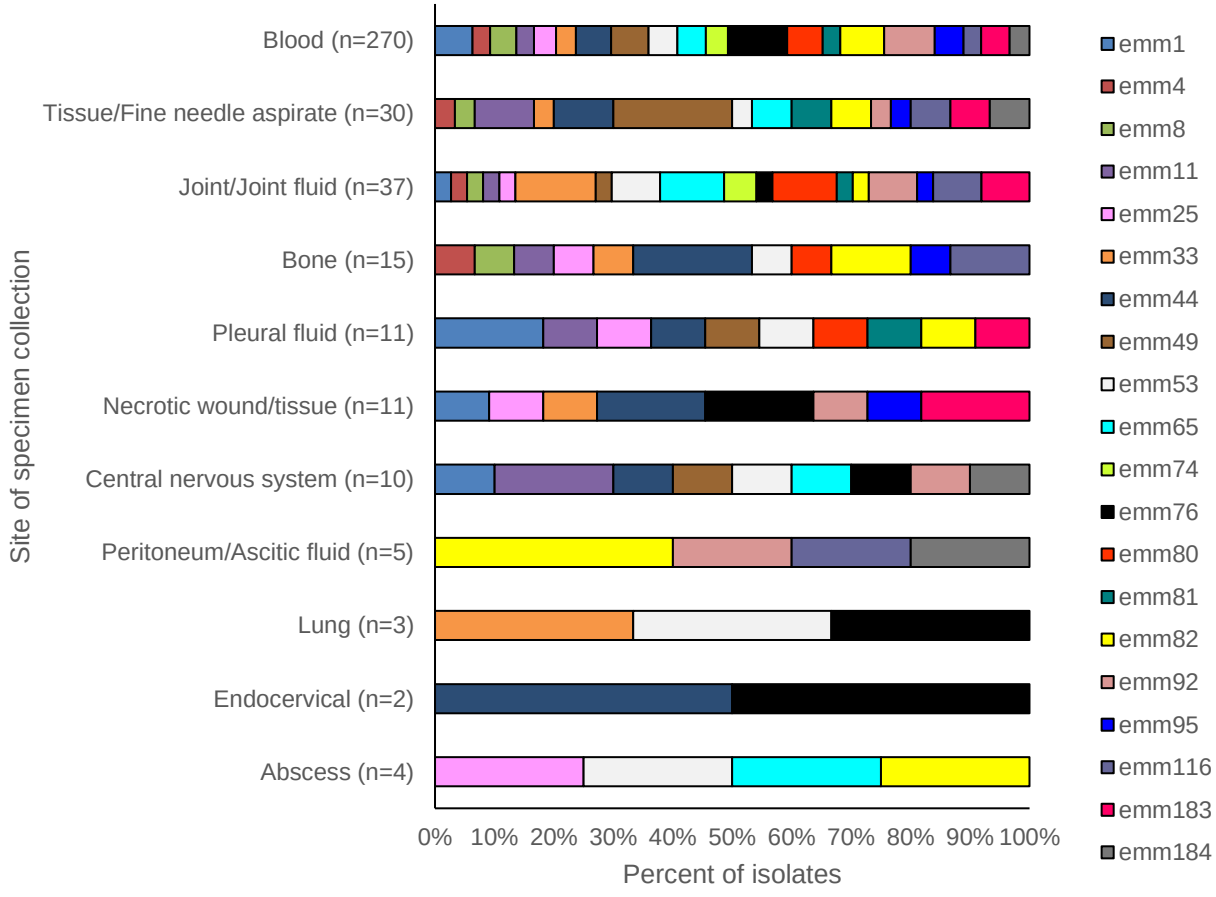


Figure 10: *emm* type distribution among invasive group A streptococcus isolates stratified by site of specimen collection. South Africa, GERMS, 2019-2020 (n = 398). Each bar represents the site of specimen collection and is colored based on *emm* type

Of the 63 isolates that tested phenotypically resistant to tetracycline, an *emm* type was available for only 57 isolates. Six isolates were excluded because they were part of the isolates that failed quality checks (**Figure 5**; 2 isolates) and isolates that were not sequenced due to time constraints (**Figure 5**; 4 isolates). The 57 isolates belonged to 14 *emm* types and 2 isolates were non-typable (**Figure 11**). *emm*76 was the most dominant *emm* type amongst the tetracycline-resistant isolates (52.6%; 30/57). There were no erythromycin-resistant *emm76* isolates. Twenty-two isolates tested phenotypically resistant to erythromycin but an *emm* type was available for only 21 isolates as 1 isolate was excluded due to poor sequence quality. The 21 isolates belonged to 6 *emm* types with *emm*95 isolates accounting for most of the erythromycin-resistant isolates (52.4%; 11/21). The difference in the proportion of *emm*95 isolates that were resistant to tetracycline and *emm*95 isolates resistant to erythromycin was statistically significant (p<0.001)

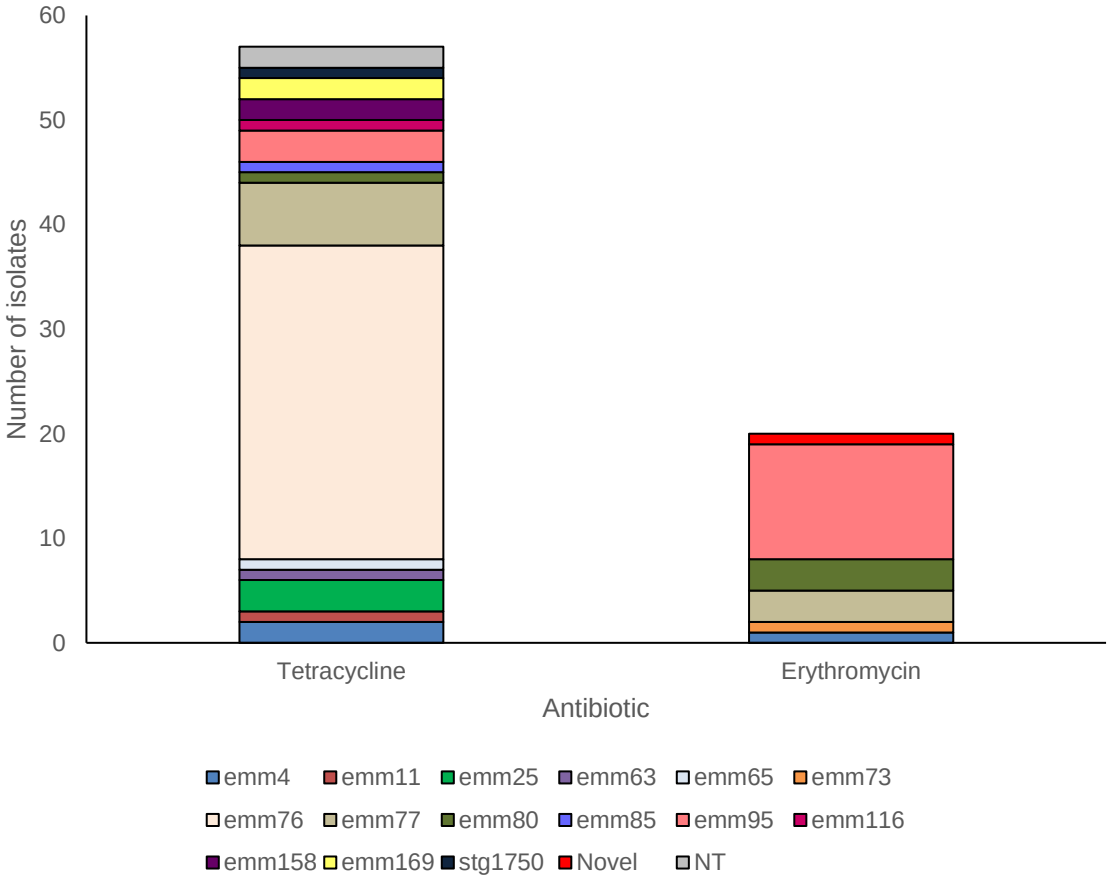


Figure 11: *emm* type distribution among invasive group A streptococcus isolates resistant to tetracycline and erythromycin. South Africa, GERMS-SA, 2019-2020 (n = 78). Each bar represents an antibiotic and is colored based on *emm* type

4.6 Strain diversity assessed by MLST

A total of 90 sequence types (ST's) were assigned for 577 isolates (**Figure 12**). New ST's were assigned to 142 isolates for the first time after the submission of unique allelic profiles of the 7 housekeeping genes to PubMLST. The 10 most common ST's (in descending order) were ST378 (34/577, 5.9%), ST82 (29/577, 5.0%), ST371 (26/577, 4.5%), ST1142 (23/577, 4.0%), ST11 (22/577, 3.8%), ST1439 (21/577, 3.6%), ST28 (21/577, 3.6%), ST400 (18/577, 3.1%), ST3 (17/577, 2.9%), ST1325 (16/577, 2.8%) which accounted for 39.3% (227/577) of the study isolates. Seventeen isolates could not be assigned an ST as an allele for the housekeeping gene *yqiL* could not be determined for these isolates, thus resulting in an incomplete profile. Sixteen of these isolates were *emm*11 and one isolate was *emm*82. The relationships of the ST's are illustrated in the minimum spanning tree below (**figure 12**) and the 17 isolates that could not be assigned an ST were excluded. Based on the distance labels observed on the diagram below, the ST's were highly diverse from one another. Based on our knowledge at the time of this analysis, clonal complexes (CC) for this pathogen have not been described.

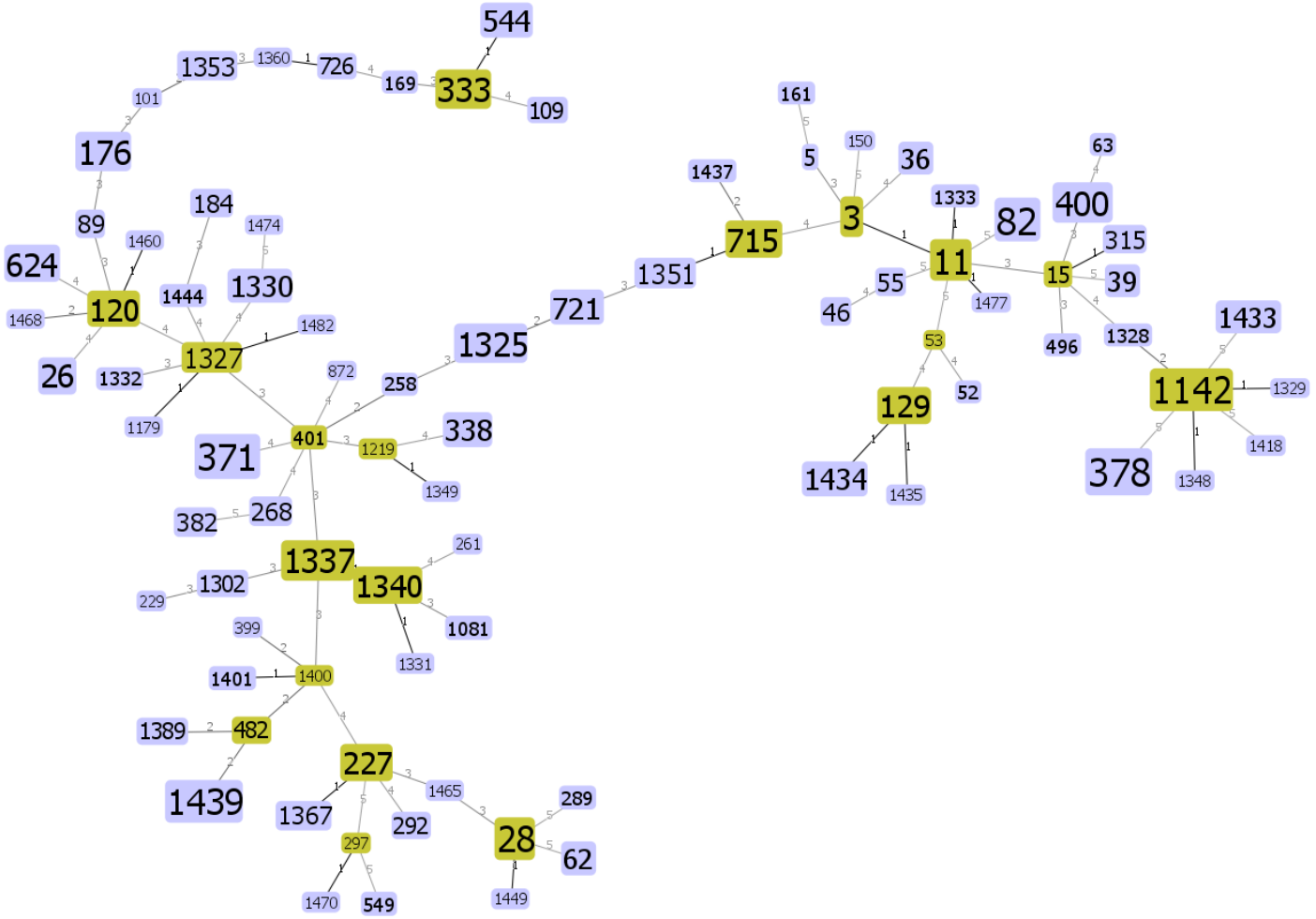


Figure 12: Minimum spanning tree showing the sequence type (ST) relationships among 560 invasive group A streptococcus isolates. South Africa. GERMS-SA, 2019-2020. The numbers in the rectangles indicate the ST and the numbers on the lines are distance labels indicating the number of different alleles between connected ST's

Thirty ST's (ST378, ST82, ST371, ST1142, ST11, ST28, ST1439, ST400, ST3, ST1325, ST333, ST1337, ST624, ST176, ST624, ST715, ST26, ST227, ST129, ST120, ST1340, ST544, ST721, ST228, ST1434, ST1330, ST39, ST36, ST1433 and ST1327) with >5 isolates each were used to assess whether a given ST was represented by a single *emm* type. **Figure 13** illustrates that for isolates with a known *emm* type, the majority of the 30 ST's were generally represented by a single *emm* type: ST378/*emm*76, ST82/*emm*92, ST371/*emm*49, ST1142/*emm*44, ST11/*emm*53, ST28/*emm*1, ST1439/*emm*183, ST400/*emm*95, ST3/*emm*33, ST1325/*emm*184, ST333/*emm*8, ST176/*emm*2, ST624/*emm*81, ST715/*emm*80, ST227/*emm*116, ST120/*emm*74, ST1340/*emm*25, ST544/*emm*124, ST74/*emm*43, ST338/*emm*68, ST1330/*emm*103, ST62/*emm*87, ST39/*emm*4, ST36/*emm*12, ST1433/*emm*102 and ST1327/*emm*18). Both ST1337 and ST26 isolates were all *emm*82 and both ST129 and ST1434 isolates were all *emm*65. No other ST's besides these two shared *emm* types. In contrast, based on the 30 most prevalent *emm* types, only 15 *emm* types were represented by a single ST whereas 14 *emm* types were represented by more than one ST (data not shown). Of the *emm* types represented by only 2 ST's, the allelic profiles of the 2 ST's differed by only one housekeeping gene. The most diverse *emm* type based on ST was *emm*77 which was represented by 5 (ST63, ST482, ST399, ST1400 and ST1401) different ST's.

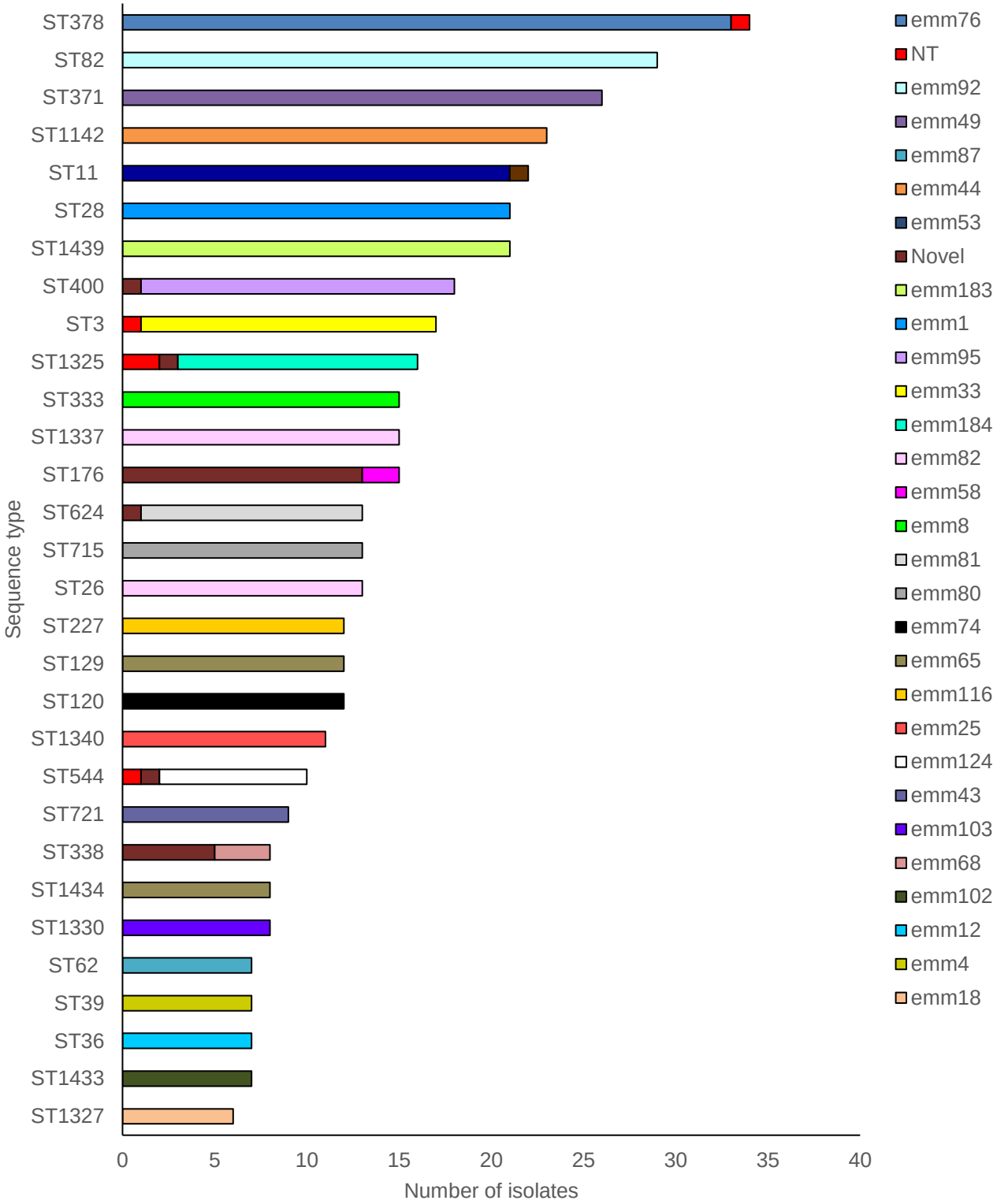


Figure 13: Distribution of multilocus sequence types and *emm* types among invasive group A streptococcus isolates collected in South Africa. GERMS-SA, 2019 - 2020 (n=435). Each bar represents a sequence type and is colored according to *emm* types that share that sequence type.

4.7 Virulence genes identified among invasive group A Streptococci

Group A streptococci genes that are commonly associated with virulence features are presented in **Table 10**. The prevalence of these features ranged from 1.7% (10/577) for the *rocA* null mutations to 97.6% (563/577) for the streptococcal pyrogenic exotoxin G (*spe*G) gene. The *rocA* null mutations [previously associated with *emm*18 (*rocA*M18) isolates and *emm*3 (*rocA*M3) isolates] were predicted in 10 isolates that were *emm*3 (n=7), *emm*25 (n=1), *emm*48 (n=1) and *emm*158 (n=1). The *sic* gene (responsible for streptococcal inhibitor of complement) was only detected in two *emm* types (*emm*1 and *emm*12) and it was detected in all the isolates of these *emm* types. Ninety-five percent (549/577) of the study isolates were predicted to produce a hyaluronic acid capsule due to the presence of the *hasA* gene. The gene targets of four fibronectin-binding proteins [fibronectin-binding protein of group A streptococci type A (*fbaA*), Protein F2/*S. pyogenes* Fn-binding protein (*prtF2*), Serum opacity factor/*S. pyogenes* Fn-binding protein II(*sof*)] and *S. pyogenes* Fn-binding protein I(*sfb1*)] were detected in >60.0% of the study isolates.

Eleven streptococcal pyrogenic exotoxin genes were detected in the study isolates and they (in descending order of prevalence) were *spe*G (97.6%; 563/577); *sme*Z (83.9%; 484/577); *spe*J (42.5%; 245/577); *spe*C (40.6; 234/577); *spe*H (26.7%; 154/577); *spe*A (22.9%; 132/577); *ssa* (22.0%; 127/577); *spe*M (19.9%; 115/577); *spe*I (18.0%; 104/577); *spe*K (12.7%; 73/577) and *spe*I (10.7%; 62/577). The 3 exotoxin genes *spe*N, *spe*O and *spe*P were not detected in the present study. *spe*G was detected in all *emm* types with the exception of *emm*63 isolates. The most exotoxin genes detected (7; *spe*A, speC, speG, speJ, speL, speM and smeZ) occurred in one *emm*92 isolate and one *emm*81 isolate had only one exotoxin gene (*spe*G) detected. One specimen had no exotoxin genes that were detected and this isolate was *emm* non-typable.

Table 10: The prevalence of virulence related features among invasive group A streptococcus isolates, South Africa. GERMS-SA, 2019 - 2020 (n=577)

Virulence feature	**[a]Number of isolates (%)**
Streptococcal pyrogenic exotoxin G (SPE_G)	563 (97.6)
Fibronectin-binding protein of group A streptococci type A (FBAA)	549 (95.1)
Hyaluronic acid capsule	549 (95.1)
ENN	529 (91.7)
MRP	528 (91.5)
Streptococcal mitogenic exotoxin (SME_Z)	484 (83.9)
S. pyogenes fibronectin-binding protein (PRTF2)	448 (77.6)
Serum opacity factor- (SOF)	353 (61.2)
S. pyogenes fibronectin-binding protein I (SFB1)	347 (60.1)
NADase D300G substitution	324 (56.2)
Streptococcal pyrogenic exotoxin J (SPE_J)	245 (42.5)
Streptococcal pyrogenic exotoxin C (SPE_C)	234 (40.6)
Streptococcal pyrogenic exotoxin H (SPE_H)	154 (26.7)
Streptococcal pyrogenic exotoxin A (SPE_A)	132 (22.9)
Streptococcal superantigen A (SS_A)	127 (22.0)
Streptococcal pyrogenic exotoxin M (SPE_M)	115 (19.9)
Streptococcal pyrogenic exotoxin L (SPE_L)	104 (18.0)
Streptococcal pyrogenic exotoxin K (SPE_K)	73 (12.7)
Streptococcal pyrogenic exotoxin I (SPE_I)	62 (10.7)
Pnga 3-Clade 3 up-regulated promoter of the nga operon	60 (10.4)
Streptodornase D1 (Virulence associated DNase)	34 (5.9)
Streptococcal inhibitor of complement (SIC)	29 (5.0)
R28 protein	10 (1.7)
rocA null mutation	10 (1.7)

[a]Number of isolates where the virulence strain feature is detected.

Table 11 illustrates the streptococcal pyrogenic exotoxin gene profiles detected in 577 iGAS isolates. Seventy-two distinct exotoxin gene profiles with unique combinations of superantigen genes (SAg) were identified. The 72 SAg profiles contained 1- 7 genes per profile. Five SAg profiles (SAg6, SAg31, SAg65, SAg70 and SAg72) contained only 1 or 2 exotoxin genes. SAg70 was made up of only 1 (*spe*G) exotoxin gene and this profile was detected in only one *emm*81 isolate. The most prevalent SAg profile was SAg1 (C: G: Z), detected in *emm*11, *emm*48, *emm*65, *emm*89, *emm*95, *st*G1750 and as well as one 'novel' isolates. To observe whether *emm* types possessed a specific SAg profile, *emm* types with >20 isolates

were considered. Eighty- percent (24/30) of *emm*92 isolates displayed a SAg5 (C: G: J: Z) profile and 75.8% (25/33) of *emm*76 isolates displayed a SAg8 (G: J: K: S: Z) profile. More than half (63.0%; 17/27) of *emm*44 isolates were a SAg11 (A: G: H: J) profile whilst no other *emm* types displayed this profile. All *emm*49 isolates (100%; 26/26) were SAg7 (G: H: I: L: M). More than half (59.1%; 13/22) of *emm*80 isolates were SAg10 (G: H: Z). The SAg2 (C: G: L: M: Z) profile was dominated by *emm*53 isolates and these isolates represented 95.2% (21/22) of the total *emm*53 isolates in this study. Seventy-one percent (15/21) of *emm*65 isolates displayed a SAg6 (G: Z) profile. *emm*82, *emm*1 and *emm*183 isolates did not possess a specific SAg profile but instead, a majority of these isolates displayed two SAg profiles.

Table 11: Streptococcal antigen gene (SAg) profiles by *emm* type among 577 invasive group A streptococcus isolates collected in South Africa, GERMS-SA, 2019 - 2020

SAg profile no.	Presence of SAg	*emm* type (no. of isolates)	No. of isolates (%)
1	C:G:Z	11 (12); 48 (2); 65 (5); 89 (1); 95 (15); stG1750 (1); novel (18)	54 (9.4)
2	C:G:L:M:Z	11 (1); 53 (21); 75 (1); 158 (3); 184 (8); novel (1); NT (2)	37 (6.4)
3	A:G:J:Z	1 (10); 80 (13); 116 (3); 183 (9)	35 (6.1)
4	G:J:Z	4 (2); 18 (1); 25 (2); 70 (1); 75 (1); 76 (1); 80 (6); 124 (7); 183 (11); novel (1)	33 (5.7)
5	C:G:J:Z	25 (1); 28 (2); 85 (2); 92 (24); 124 (1); novel (1); NT (1)	32 (5.5)
6	G:Z	65 (15); 68 (3); 73 (1); 77 (6); 114 (2); 118 (2), 183 (1)	30 (5.2)
7	G:H:I:L:M	49 (26)	26 (4.5)
8	G:J:K:S:Z	76 (25)	25 (4.3)
9	G:H:J	44 (5); 81 (3); 103 (1)	19 (3.3)
10	G:H:Z	25 (1); 82 (13); 94 (4)	18 (3.1)
11	A:G:H:J	44 (17)	17 (2.9)
12	G:H:S:Z	33 (15); NT (1)	16 (2.8)
13	C:G:H:I:Z	12 (5); 82 (10)	15 (2.6)
14	G:K:S:Z	8 (10); 76 (3); 177 (2)	15 (2.6)
15	A:C:G:J:Z	1 (12); 92 (2)	14 (2.4)
16	G:H:J:Z	25 (10); 65 (1); 82 (1); 103 (2)	14 (2.4)
17	A:C:G:Z	11 (3); 74 (5); 92 (1); 15 (1); novel (2)	12 (2.1)
18	A:G:J:S:Z	116 (10); 18 (1)	11 (1.9)
19	G:J:S:Z	18 (5); 87 (1); 111 (2); 116 (2)	10 (1.7)
20	C:G:K:Z	43 (9)	9 (1.6)
21	A:G:Z	73 (2); 78 (5); novel (1)	8 (1.4)
22	C:S:Z	4 (7)	7 (1.2)
23	G:H:I:Z	12 (1); 25 (1); 73 (1); 82 (4)	7 (1.2)
24	A:C:G:K:Z	3 (1); 6 (4); 74 (1)	6 (1.0)
25	A:G:M:Z	74 (6)	6 (1.0)
26	C:G:J:S:Z	87 (6)	6 (1.0)
27	C:G:S:Z	22 (4); 95 (1); novel (1)	6 (1.0)
28	A:C:G:L:M:Z	184 (5)	5 (0.9)
29	C:G:L:M	2 (4); 53 (1)	5 (0.9)
30	G:H:I:J:L:M	102 (5)	5 (0.9)
31	G:J	81 (4); novel (1)	5 (0.9)
32	G:L:M:S:Z	39 (4); novel (1)	5 (0.9)
33	A:G:K:S:Z	8 (5)	4 (0.7)
34	G:H:J:L:M	44 (4)	4 (0.7)
35	G:M:Z	90 (4)	4 (0.7)
36	A:C:G:H:I:Z	12 (1); stG1750 (1); novel (1)	3 (0.5)
37	C:G:H:Z	58 (2); 103 (1)	3 (0.5)

SAg profile no.	Presence of SAg	*emm* type (no. of isolates)	No. of isolates (%)
38	C:G:J	81 (1); 92 (2)	3 (0.5)
39	C:G:J:K:S:Z	76 (3)	3 (0.5)
40	G:L:M:Z	77 (1); 83 (1); novel (1)	3 (0.5)
41	A:C:G:K	3 (2)	2 (0.3)
42	A:G:S	3 (2)	2 (0.3)
43	C:G:H:J:Z	103 (2)	2 (0.3)
44	G:H:I:J:L:M:Z	102 (2)	2 (0.3)
45	G:H:I:J:Z	118 (1); 192 (1)	2 (0.3)
46	G:H:S	33 (2)	2 (0.3)
47	G:J:K:S	76 (1); NT (1)	2 (0.3)
48	G:J:L:M:S:Z	116 (1); 192 (1)	2 (0.3)
49	G:J:L:M:Z	80 (2)	2 (0.3)
50	G:K:Z	58 (2)	2 (0.3)
51	G:S:Z	95 (1); 165 (1)	2 (0.3)
52	K:S:Z	4 (2)	2 (0.3)
53	L:M:Z	77 (2)	2 (0.3)
54	A:C:G:J:L:M:Z	92 (1)	1 (0.2)
55	A:C:G:K:S	3 (1)	1 (0.2)
56	A:C:G:S	3 (1)	1 (0.2)
57	A:G:H:I:Z	103 (1)	1 (0.2)
58	A:G:H:J:Z	103 (1)	1 (0.2)
59	A:G:S:Z	169 (1)	1 (0.2)
60	C:G:H:J	81 (1)	1 (0.2)
61	C:G:H:J:L:M	44 (1)	1 (0.2)
62	C:G:H:S:Z	33 (1)	1 (0.2)
63	C:G:J:L:M:Z	192 (1)	1 (0.2)
64	C:H:I	63 (1)	1 (0.2)
65	C:M	63 (1)	1 (0.2)
66	G:H:J:S	81 (1)	1 (0.2)
67	G:H:L:M:Z	82 (1)	1 (0.2)
68	G:J:K:Z	80 (1)	1 (0.2)
69	G:L:M	83 (1)	1 (0.2)
70	G	81 (1)	1 (0.2)
71	C:G:L:M:S:Z	novel (1)	1 (0.2)
72	G:H	81 (1)	1 (0.2)

1532

1533

4.8 Phylogeny

A maximum-likelihood tree based on SNP's (single nucleotide polymorphism) in the core genome of 577 iGAS isolates in this study is depicted in **Figure 14**. The rings of the phylogenetic tree are coloured based on year, province and *emm* type.

Based on the organization of the isolates within the phylogenetic tree, isolates of different *emm* types appeared genetically distinct from one another. We identified genetic clusters as isolates of the same *emm* type grouped closely to one another within the tree. Although this was the general pattern, there were some exceptions where isolates of the same *emm* type formed more than one cluster. *emm*82, *emm*77, *emm*80 and emm58 isolates formed two distinct clusters, whereas *emm*4 isolates formed three distinct clusters. Of the three *emm*118 isolates in this study, 2 isolates clustered together and one isolate was an outlier and of the four *emm*73 isolates detected in this study, 3 isolates clustered together whereas one isolate was an outlier. No geographical and temporal aggregation of isolates sharing the same *emm* type was observed.

As mentioned in section 4.5, there were 30 isolates that were not a perfect match to existing *emm* subtypes in the CDC database, but on the phylogenetic tree, these isolates clustered with other isolates of a known *emm* subtype. Of the 30 isolates, 27 clustered with other isolates of known *emm* subtypes/types as follows: 13 isolates that were almost identical to *emm* subtype 58.2 according to the pipeline clustered with other *emm*58 isolates; 5 isolates that were almost identical to *emm* subtype 68.4 according to the pipeline clustered with other *emm*64 isolates and finally, in at least one isolate, similarities to *emm* subtypes 53.0, 95.0, 83.1, 74.0, 81.0, 85.0, 124.0, 18.7 and 184.0 were detected by the pipeline and these isolates clustered with isolates that were *emm*53, *emm*95, *emm*83, *emm*74, *emm*81, *emm*85, *emm*124, *emm*18 and *emm*184, respectively.

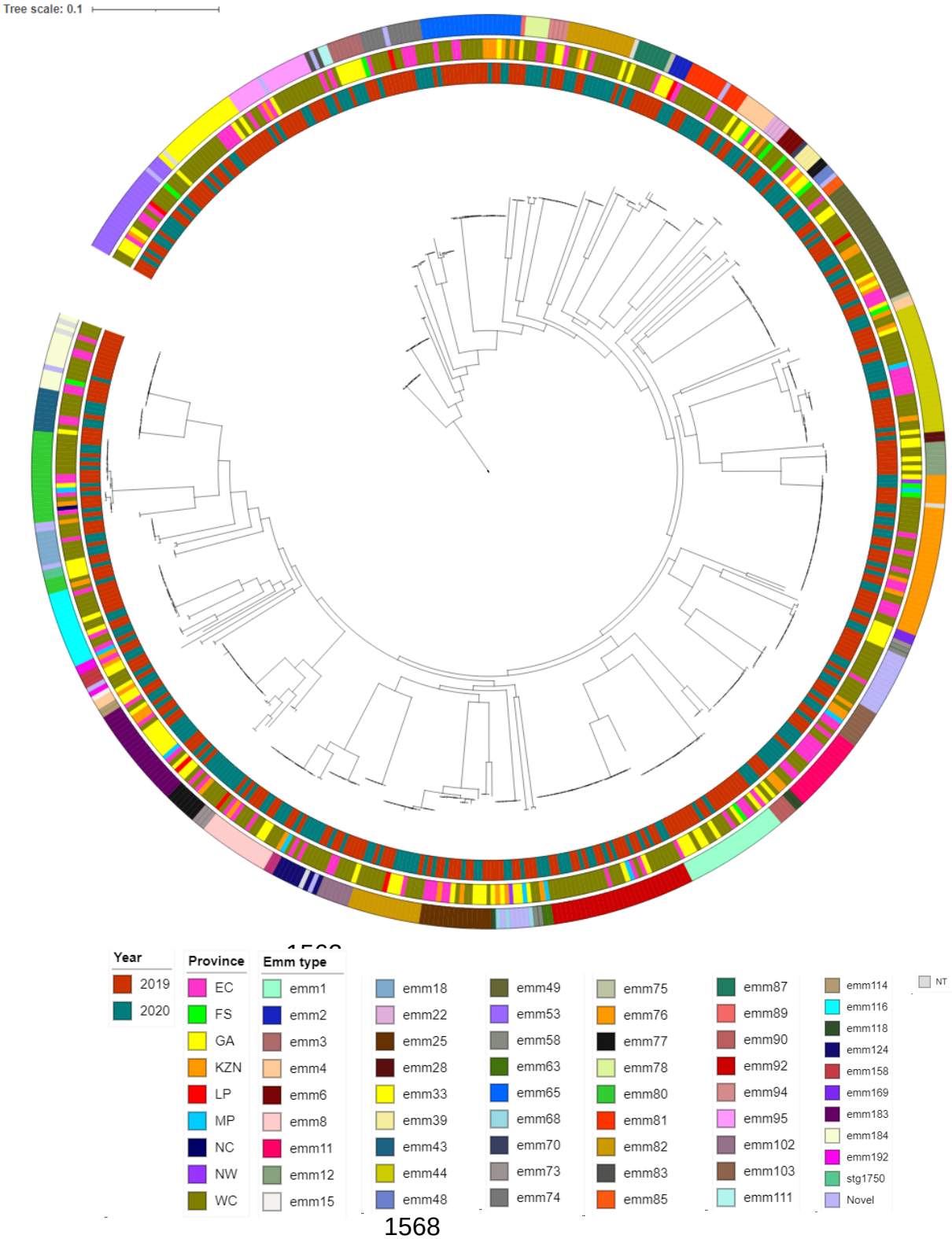


Figure 14: A core gene maximum-likelihood phylogenetic tree of 577 invasive group A streptococcus isolates recovered through GERMS-SA between 2019 – 2020. Inner colour ring going outwards are as follows: Year, Province and *emm* type

5 DISCUSSION

In this study, we aimed to describe the epidemiology and contribute genomic data of *S. pyogenes* isolates causing invasive disease in South Africa, a middle-income country where data of this nature is currently lacking, by whole-genome sequencing of isolates collected in our active, national laboratory-based surveillance. A total of 1487 cases of iGAS disease were detected between 2019 – 2020 and the isolates were a diverse collection.

5.1 Demographic and clinical characteristics

The number of iGAS cases reported in 2019 were relatively stable between January – September. There was a sharp decline in the cases reported in October and the cases remained relatively low until 31 December 2019. The reason for this decline is not clear. Cases slightly increased in January 2020, declined again in April and remained relatively stable until 31 December 2020. From 26 March 2020 – 30 April 2020, level 5 lock down was implemented in South Africa in response to the Coronavirus disease 2019 (COVID-19) pandemic. Various non-pharmaceutical interventions (NPI) including but not limited to mask wearing, hand sanitization and social distancing were implemented to curb the transmission of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (135). The lockdown levels varied between level 1 and level 5 throughout 2020, but most of the NPI's remained in place during the different levels. These NPIs may have indirectly affected the transmission of other pathogens that are mainly transmitted via the respiratory route, such as *Streptococcus pyogenes*. Studies conducted in Australia (non-invasive), China (non-invasive) and the US (invasive) reported a decrease in GAS disease following the COVID-19 pandemic (136–138). Health-seeking behavior may have also been affected causing individuals to be more reluctant to leave their homes for medical facilities in fear that they would contract SARS-CoV-2 and develop COVID-19. The restrictions may have also affected the availability of staff as some healthcare and laboratory institutions implemented shift work for social distancing and some staff members may have been isolating due to contracting SARS-CoV-2 or under quarantine following coming into contact with a case. Having less staff on

duty may have interrupted the normal work-flow thus negatively impacting the isolation of iGAS, the collection of iGAS isolates and the reporting of these cases. Most iGAS cases in the present study occurred in the colder months. We only have two years of data and the disruption caused by the COVID-19 pandemic in 2020 also makes it difficult to reach a final conclusion regarding the seasonality of iGAS disease in this setting. Previous studies have reported observing a seasonality of iGAS diseases, with most of the cases occurring in the winter months (98,100,101,104). It is not clear why iGAS disease is more prevalent in the colder months but it has been hypothesized that pathogens that are mainly transmitted via the respiratory route are more prevalent in colder months as a result of behavioral changes during that time (139). In the colder months, individuals are more likely to remain indoors resulting in indoor crowding with poor ventilation in an effort to preserve heat, which in turn promotes transmission (139). The inhalation of chilly air may also dry the mucosal membrane, thus resulting in epithelial damage which would make it easier for respiratory pathogens to colonize their host (140). However, some studies have reported not observing any seasonality at all for iGAS disease (116,141) while one study in Cape Town reported that iGAS disease was more common in the summer months and non-invasive disease was more common in the winter months (59).

The four most populated provinces in South Africa (Gauteng, Kwa-Zulu Natal, Western Cape and Eastern Cape) reported 94.0% (1398/1487) of the total cases. These provinces do not make up 94% of the total population in South Africa and it is therefore unlikely that the population size is the only reason why these provinces reported most of the iGAS cases. This may be attributed to the availability of resources in these settings such as better access to healthcare with good specimen collection practices and better equipped laboratories for testing.

iGAS disease cases are usually highest in the very young (<1 year) due to (but not limited to) an immature immune system and in the elderly (>65 years) as a result of (but not limited to) their waning immunity (99–101,105,108). Although most cases in the present study occurred in middle-aged individuals, the incidence of invasive GAS disease was highest in infants aged <1 years old and in the elderly aged ≥65 years old. This suggests that the greatest burden is seen in very young and elderly and these two groups are at a greater risk of developing iGAS disease and when

vaccines become available, these individuals should be considered as some of the first groups to be targeted by the intervention.

Other studies on iGAS conducted in Norway, Germany, Ireland, Brussels, USA, Spain and South Africa (59,102–104,108,113,142) reported that males accounted for most (53% - 63%) of the iGAS cases. This is similar to what was observed in the current study as a slightly higher proportion of cases were reported in males (52.8%; 785/1487) and it is not clear why this is so. Amongst the cases that occurred in males, 69.3% (544/785) of the cases occurred in adult males who were ≥25 years old [53% (360/679) for women] and thus a possible explanation for why cases were higher amongst males could be due to behavioral differences observed between males and females. Adult males may be disproportionally exposed to circumstances that are known to be risk factors for GAS disease such a homelessness (143) and the use of drugs or alcohol (144,145). In addition, it has been noted that males tend to delay seeking medical help when they are ill due to social beliefs such as 'masculinity' (146) and as a result may be less likely to deal with possible iGAS precursors (such as superficial skin infections and pharyngitis) and only present when illness has progressed and is more severe.

Data on race was poorly collected in this study. Of the cases where race was known, cases were more prevalent amongst black (73.3%; 77/105) and coloured (21.9%; 23/105) individuals. This is expected as black individuals accounted for 81% of the total South African population in 2019 (https://www.statssa.gov.za Accessed: 04 March 2024) and 2020 (https://www.statssa.gov.za Accessed: 04 March 2024) whilst coloured individuals accounted for 9% of the total South African population during the same time-period. This (together with how poorly race data was collected) makes it difficult to establish whether our results accurately indicate that these groups are disproportionally affected by iGAS as this may be a reflection of the dynamics of our population. In South Africa, black and coloured individuals come from a historically disadvantaged socio-economic background, which is linked to various GAS disease risk factors including (i) limited access to good health-care; (ii) poor living conditions (e.g. poor ventilation, limited resources for cleaning and washing); (iii) overcrowding in households; and (iv) environmental air pollutants (e.g. burning of fossil fuels for heat and cooking) (147–149). Studies in Canada and Australia have published data illustrating that disadvantaged individuals in their population report higher rates of

iGAS disease (111,141,150). This highlights the importance of accurately collecting race data.

Slightly over half (51.8%) of the iGAS cases in the present study were isolated from blood specimens. Research conducted in Brussels, Spain and India has indicated a comparable prevalence in blood specimens ranging from 44% to 55% whereas, other studies in Germany, Australia and the USA have reported an even higher (>90%) prevalence of GAS isolated from blood specimens (103,106,108,113,118,141,142,151). This suggests that bacteremia may be the most common clinical presentation among patients with iGAS disease. However, it is important to note that this may be a reflection of specimen collection practices as blood may be the easier specimen to collect.

Where in-hospital final outcome was known, mortality was recorded in 20.7% of the cases. This mortality rate is slightly higher than the mortality rates reported in similar studies conducted in Iceland, Finland, Norway, Brussels and Australia, which recorded mortality rates ranging between 5-19% (100–102,108,118,141). It is important to note that due to how poorly this variable was collected, the mortality rate may be underestimated in our population. Where these data were available, the information was collected from hospital records. Home deaths would have been missed as government death registries were not reviewed and would have been very important for patients that died of iGAS prior to presenting at hospital for investigation of their illness. Three of the studies (100,102,118) mentioned above describe that severe iGAS disease (sepsis/streptococcal toxic shock syndrome and necrotizing fasciitis) was associated with significantly higher mortality rates in their population. In the present study, 2.7% of the isolates analysed were collected from necrotic wound/tissue and although 51.8% of isolates were obtained from blood, data on whether this was associated with underlying necrotizing fasciitis or progressed to more severe disease such as sepsis or streptococcal toxic shock syndrome are not available. A Kenyan study reported a mortality rate (12.2%; 45/369) lower than our mortality rate (61). Other negative outcomes due to iGAS disease including surgical intervention (drainage of abscess, debridement, amputation etc.) have been described (100,108,118,152). These data were not collected in the present study.

5.2 Antimicrobial susceptibility

Similar to what has been reported globally (90,98,102,104,105,107,112,114,153), all the isolates that were tested in this study tested phenotypically susceptible to all three β-lactam antibiotics (penicillin, ceftaroline and ceftriaxone). Therefore, first-line treatment with these antibiotics remains appropriate in our setting. Mutations in the pbp2x region that may result in elevated MIC's against β-lactam antibiotics were first reported in 2020 during a GAS outbreak investigation in the US (46). In the course of the investigation, two near-identical invasive isolates collected from two patients shared an identical mutation in the pbp2x region, resulting in the pbp2x T553K substitution. This mutation resulted in elevated β- lactam MIC's. For both cases, patient interviews and medical record review revealed that these patients had a long and repeated history of antibiotic, including β-lactam, use which may have created a favorable environment for mutants to emerge. Since then, GAS isolates with reduced β-lactam susceptibility have been reported in other studies conducted in the US, Japan and Iceland (134,154–156). Although these mutations resulted in slightly elevated MIC's compared to the wild-type MIC's (i.e. decreased β-lactam susceptibility) the elevated MIC's remained below the susceptible clinical breakpoints. In the present study, 86.8% (501/577) of the isolates that were screened for mutations in the pbp2x region were the wild-type pbp2x –1 (pbp2x sequence type 1) i.e. no mutations associated with elevated MIC's were detected. Interestingly, 5 of these isolates displayed elevated MIC's against penicillin (0.12µg/ml; all 5 isolates) and ceftriaxone (0.5µg/ml; 1 isolate). This is contrary to what has been previously reported (134) as one does not expect to observe elevated MIC's in isolates that do not have a mutation in the pbp2x region. This could be the result of the inherent inaccuracy in phenotypic testing or an isolate mix-up and repeating the phenotypic and genotypic testing of these isolates may resolve the discordance. This discordance highlights that all isolates that display a higher MIC than what is commonly observed in our setting should be re-tested. Only one sequence type (pbp2x-10) previously associated with elevated MIC's in invasive isolates in the US (134) was detected in our study isolates. Sequence type pbp2x-10, with one amino acid substitution (M593T), was detected in 14 isolates, but only 1 of the 14 isolates displayed an elevated MIC against penicillin. Although the mutation was present in the other 13 isolates, this mutation does not always confer reduced

susceptibility on phenotypic testing (134). Globally, mutations in the pbp2x region resulting in reduced β-lactam susceptibility are rare (157) but ongoing surveillance is important for monitoring this emerging resistance.

The highest resistance rate (10.5%, 65/618) was seen against tetracycline. Previous studies in France, Norway, Ireland, Portugal, Brussels, New Caledonia and Spain have reported similar resistance rates in iGAS isolates ranging between 7–15% (98,102,104,105,108,116,142). A study in the US reported a slightly higher resistance rate of 17% in invasive isolates whereas studies conducted in Brazil and Lebanon reported even higher (>20%) resistance rates against tetracycline in non-invasive isolates (38,89,114). As a result of the high resistance rate against tetracycline, this antibiotic is not routinely used to treat infections caused by GAS and in addition, this antibiotic has also been reported to have negative effects on bone and teeth in children under 8 years of age, and in fetuses of mothers exposed to tetracycline antenatally (43). Although not routinely administered, the susceptibility and resistance patterns of GAS to tetracycline remain epidemiologically important to monitor any changes that may occur (41). The most common *tet* family gene detected in the present study was *tetM*, which was present in 48 isolates that were phenotypically resistant to tetracycline. Three studies conducted in France, Spain and in China have also reported a dominance of this gene among isolates resistant to tetracycline (41,98,158), suggesting that resistance to tetracycline may be predominantly driven by the *tetM* gene. Phenotypic and genotypic results were discordant for six isolates in the present study as the *tetM* gene was present in 2 isolates that were phenotypically susceptible and a *tet* family gene was not detected in 4 isolates that were phenotypically resistant to tetracycline. The phenotypic testing was repeated for these isolates and the results remained the same. Due to time constraints, the genotypic testing could not be repeated. It is thus not clear whether the discordant results are a lab error occurring in the genotyping workflow or the discordance may in fact represent a true difference in the sensitivity of the two methods. It may also be that although the *tet* gene is present, the genes may not be expressed in the isolates that are phenotypically susceptible, a phenomenon that has been described in other pathogens (159). Repeating the phenotypic and genotypic testing at the same time from the same culture may possibly resolve the discordance. A discordance between phenotypic and genotypic testing for tetracycline resistance has been previously reported in invasive and non-invasive

isolates by a group in China (158). This study reported that 7 isolates that were phenotypically susceptible carried the *tetM* gene and that 3 isolates that were phenotypically resistant to tetracycline did not carry a *tet* gene, resulting in a discordance rate for tetracycline (7.1%; 10/140) that was higher than the one obtained in the present study (1.0%; 6/577). The Chinese group did not provide a hypothesis for why the discordance between the phenotypic and genotypic results existed.

The resistance rate against erythromycin in the present study was 3.6% and this is comparable to previously published data (102,104,105,108) that reported resistance rates ranging 3-5% in invasive isolates. Studies from Finland (101), Spain (142) and the US (38) have reported even higher resistance rates in iGAS isolates which were 8.7%, 8.9% and 14.5%, respectively. Surprisingly, no resistance against clindamycin was detected in the present study, whereas clindamycin resistance rates between 3-4% in invasive isolates have been reported elsewhere (102,108,142). Studies that have reported even higher resistance rates against clindamycin include a Finland study (101) that reported a resistance rate of 9.2% for invasive isolates collected in 2013; a US study (38) that reported a resistance rate of 14.6% in invasive isolates collected 2011-2017 and finally a publication from Greece that reported that 18.8% of their invasive and non-invasive isolates were resistant to clindamycin (45). Clindamycin is recommended for the treatment of severe invasive disease (such as necrotizing fasciitis and streptococcal toxic shock syndrome) and considering that no isolates with clindamycin resistance were recorded in our study, this suggests that treatment of severe iGAS disease with this antibiotic remains ideal in our setting. It is important though to note that in our study, phenotypic inducible clindamycin resistance, which occurs in the presence of an inducing agent (160), was not tested for. The genomic data may assist in detecting isolates that are inducibly resistant to clindamycin as they should carry the *erm* gene even if they appear phenotypically susceptible to macrolides and clindamycin (38). Various studies (38,39,142,161) have tested for inducible clindamycin resistance, which highlights the importance of screening for this type of resistance. Inducible clindamycin resistance that is not detected could have public health implications as clindamycin resistance in patients may only develop during therapy, thus resulting in treatment failures (162). AMR predictions indicated that the *mef* gene seems to be driving macrolide resistance in our setting as this was the most commonly detected AMR determinant in isolates

resistant to erythromycin. This is contrary to other studies that have reported that the *erm* gene is most commonly detected AMR determinant in isolates that are resistant to macrolides (41,43,112,114,122,158). The presence of the *erm* gene may result in the cross-resistance to all macrolides, lincosamides and streptogramins B, while the presence of the *mef* gene may result in resistance against certain macrolides only (163) and thus this may explain why the prevalence of the *erm* gene did not dominate in our setting as no phenotypic resistance was seen against lincosamides (clindamycin). There was some discordance between the genotypic and phenotypic results (discordance rate of 0.5%; 3/577) even after repeating the phenotypic testing. The genotypic testing was not repeated due to time constraints. One isolate was phenotypically susceptible to both erythromycin and clindamycin while carrying the *erm* gene. This isolate may be inducibly resistant to clindamycin, hence the presence of the *erm* gene but further testing is needed to confirm this. A CDC publication reported a similar instance where one isolate was susceptible to erythromycin and clindamycin but carried the *ermT* gene (122). In the CDC study, inducible resistance to clindamycin was phenotypically screened for but not detected for the isolate. Another study conducted in China (158) reported that 4.3% (6/140) of their isolates carried the *ermB* gene but were phenotypically susceptible to erythromycin and clindamycin. These results suggest that the discordance may be due to differences in the sensitivity of the phenotypic and genotypic tests. In the present study, two isolates that were phenotypically resistant to erythromycin lacked both the *erm* or *mef* gene. The CDC publication mentioned above (122) also reported a similar discordant result where 2 of their isolates that were phenotypically resistant to erythromycin lacked erythromycin resistance determinants (122). A Chinese study reported a similar discordance and hypothesized that the observed discordance may be a result of an alternative mechanism of resistance or there may potentially be a mutation in the 23S rRNA (158).

Resistance to levofloxacin in our study was low (0.2%; 1/618) compared to the resistance rates reported in Germany (1.3%; invasive isolates), China (1.7%; invasive isolates), Greece (2.0%; invasive and non-invasive isolates) and Lebanon (5.0%; non-invasive isolates) (39,45,89,164). A study in China and Brazil did not detect any resistance to levofloxacin among their non-invasive isolates (37,114). Although the resistance rate against fluoroquinolones is low, this antibiotic is not commonly prescribed, especially to children, due to the negative effects it has on

bone and joint development (43). In South Africa, fluoroquinolones may be used to treat multi-drug resistant tuberculosis. In the present study, 7 isolates that were phenotypically susceptible to levofloxacin carried mutations in the *parC* region. Although these isolates were phenotypically susceptible to levofloxacin, the MICs for two isolates were slightly elevated and were at the cusp of susceptibility with an MIC of 2µg/ml. Not all mutations in this region lead to the phenotypic resistance of fluoroquinolones and mutations that occur in the *parC* may only lead to low-level resistance, whereas high-level resistance occurs when mutations occur in both the *parC* and *gyrA* region (165). One isolate was phenotypically resistant to levofloxacin but no mutations in the *parC*/*gyrA* region were detected. The phenotypic testing could not be repeated for this specimen as the laboratory did not have stock of the antibiotic and thus it is not clear whether this discrepancy is a laboratory error. Another possible explanation for the discordance is that resistance to levofloxacin in this isolate may be driven by a different resistance mechanism that was not screened for in this study. A study in China reported that they detected the *pmrA* gene, which is a drug efflux gene that confers resistance to fluoroquinolones through drug efflux.

5.3 *emm* type prevalence and the estimated vaccine coverage

Studies have shown that sporadic increases in the prevalence of GAS disease have been linked to the upsurge of specific *emm* types (166–169). This highlights the importance of monitoring the circulation of *emm* types. The most frequently detected *emm* types in the present study differ from those previously reported in high-income countries (**Table 2**). In studies conducted on the African continent (the Gambia, Kenya and two South African studies) between 2014 and 2023 using invasive and non-invasive isolates, the *emm* types detected in each of these studies differed to the *emm* types in present study, with the exception of one overlapping *emm* type per country i.e. *emm*80 was also detected in the Gambia, *emm*44 in Kenya and *emm*2 and *emm*92 in each of the South African studies (22,61,96,97). This difference may be because most of the isolates in the studies mentioned above were obtained from non-invasive GAS disease. Globally, *emm*1 is the most prevalent *emm* type isolated in high-income countries whereas in the present study, it was detected in only 3.7% of the isolates. The most prevalent *emm* type in the present study was *emm*76,

which was also the most commonly detected *emm* type amongst invasive isolates collected in a study conducted in Cape Town (59). This *emm* type is not included in the 30-valent M-protein-based vaccine, but is among the 24 *emm* types shown to be cross-opsonized (>40% bactericidal killing activity) in animal studies (57). Only 5 of the 10 most common *emm* types in the present study are included in the vaccine formulation. The 30-valent M-protein-based vaccine offers coverage to less than half (44.6%) of the isolates in the present study, consistent with results from studies conducted in other low-income countries and disadvantaged populations [Kenya (invasive isolates), Mali (non-invasive isolates), Australia (non-invasive isolates), South Africa (invasive and non-invasive isolates) and India (invasive isolates)] (59,61–63,96,151), which reported a potential vaccine coverage ranging 22 - 47% in their setting. Although the percent of isolates in the present study potentially covered by the 30-valent M-protein-based vaccine increases to 64.4% (352/547) when cross opsonization data was considered, this still was much lower that what has been reported (>90%) in studies from the US and Europe (107,112). The study in Cape Town reported a potential coverage of 58% and 60% among invasive and non-invasive isolates respectively when cross-protection was considered and also highlighted that the *emm* types causing invasive disease and those causing non-invasive disease were similar (59). It is not completely understood why cross-protection occurs, but the hypothesis that has been proposed is that it may be due to vaccine and non-vaccine *emm* types belonging to the same *emm* cluster, thus having high-sequence identity across the full length of the M-protein and therefore being similarly protected (119). A majority (27/40) of the M-proteins that were tested for cross-opsonization in animal studies cluster in *emm* clusters that have at least one *emm* type included in the 30-valent M-protein-based vaccine (57,119). Although cross-opsonic antibodies acting against non-vaccine *emm* types have been demonstrated in animal studies, clinical significance is not clear as it has not been determined whether the same will occur in human subjects (57). An Australian study demonstrated that potential coverage by the 30 valent M-protein-based vaccine was higher in cases of pharyngitis (66.7%) in comparison with cases of cutaneous (skin and soft tissue infections) infections (25.2%). This is not unexpected as *emm* types that are common causes of pharyngitis in the US and Europe were considered in the formulation of the 30-valent M-protein-based vaccine (57). Streptococcal pharyngitis is a precursor for ARF/RHD, which are responsible for most of the morbidity related

to GAS diseases. By reducing the incidence of strep throat, vaccination can significantly lower the risk of post-streptococcal complications. Therefore, data from low-income countries where streptococcal pharyngitis and ARF/RHD are common are essential to inform vaccine composition and deployment strategies for these vaccines. The Australian study mentioned above also highlights the importance of considering *emm* types that cause soft skin and tissue infections as these *emm* types seem to be an important cause of ARF in Australia, which has one of the highest burdens of ARF globally (63).

While iGAS can be severe, its overall burden may not justify a vaccine program solely aimed at reducing iGAS cases.

The majority of GAS isolates express the M related protein (Mrp) (58,65). Mrp was detected in most (91.5%; 528/577) of the isolates in the present study and therefore this provides support for the concept of formulating a vaccine that targets both the M protein and the M related protein.

The expansion of highly resistant *emm* types may play a role in driving resistance to specific antibiotics (38). A study in the US found that the majority (>90%) of *emm*92 isolates in their collection were non-susceptible to erythromycin and clindamycin (38). Another study in Spain also showed an association between *emm*22 isolates with tetracycline resistance, *emm*28 isolates with clindamycin resistance and both *emm*4 and *emm*12 isolates with erythromycin resistance (142). In the present study, resistance to tetracycline was most commonly seen in *emm*76 isolates and resistance to erythromycin was most commonly seen in *emm*95 isolates. These two *emm* types are not included in the 30-valent vaccine. Fortunately, resistance to erythromycin was not detected in any of the *emm*76 isolates which would have caused public health concern considering that *emm*76 is the most dominant *emm* type in this study and that macrolides are used for treatment in penicillin hypersensitive individuals. However, resistance to erythromycin has been reported in *emm*76 isolates as a study in Japan conducted on GAS isolates from streptococcal toxic shock syndrome cases reported that *emm*76 isolates accounted for 4.0% of erythromycin-resistant isolates and 19.9% of clindamycin-resistant isolates (109). This is concerning as *emm*76 GAS can be spread geographically. This is why it remains important to monitor antimicrobial resistance in relation to *emm* type and to assess any changes that may occur when vaccines are introduced.

In 2020 enhanced surveillance sites started collecting additional clinical data but this was unfortunately poorly collected and therefore, site of specimen collection was used *in lieu* of clinical manifestation in this study to investigate whether certain *emm* types aggregated by site of specimen collection. In the present study, there was no apparent pattern in how *emm* types aggregated based on specimen type. Our result is similar to a study that was conducted in a pediatric population in France, which also reported that a significant correlation between *emm* type and clinical manifestation was not found (170). However, other studies have demonstrated an association between clinical manifestation and *emm* type. A study in Spain reported that *emm*1 isolates were significantly associated with more severe clinical manifestations such as necrotizing fasciitis and streptococcal toxic shock syndrome (STSS), whereas *emm*3 and *emm*4 isolates were associated with the less severe scarlet fever (107). This is similar to what was reported in a study conducted in London where they observed that *emm*3 isolates were associated with scarlet fever (171). Another study in Norway reported that *emm*2 isolates were significantly associated with necrotizing fasciitis and *emm*12 isolates were associated with arthritis/osteomyelitis (102). Inferring associations between *emm* types and clinical manifestation/site of specimen collection is difficult to do and this is due to the high diversity of circulating *emm* types at any given time, especially in low-income countries which have been reported to have a higher diversity of *emm* types (95). In addition, isolate numbers in the present study were too low for some sites for us to see an association. In the present study, *emm* types were more diverse in blood specimens and 11 *emm* types were unique to blood specimens. This suggests that collecting only blood specimens for iGAS disease surveillance may be sufficient in our setting as it still provides a representative picture of strains that are circulating in the population.

We were able to assign 547 isolates to an *emm* cluster based on the *emm* cluster classification scheme (119). Due to the high diversity of *emm* types represented in this study, 15/48 described *emm* clusters were represented in this isolate collection. A Kenyan study, which also reported a high diversity of *emm* types (80 *emm* types), assigned 24 *emm* clusters to their study isolates (61).

5.4 Strain diversity by MLST

The study isolates were a diverse collection as 90 ST's were assigned to 560 isolates. The ST's in this study were largely different from ST's reported in other studies. A study conducted in Malaysia detected 27 ST's in their isolate collection and only 7 ST's (ST28, ST36, ST549, ST89, ST101, ST55 and ST5) matched ST's detected in the present study (172). A Romanian study reported detecting 23 different ST's, of which only ST378, ST28 and ST150 were also detected in the present study (91). Fifty-seven ST's were assigned to isolates collected in the Gambia and only 10 ST's (ST401, ST715, ST1219, ST229, ST268, ST721, ST120, ST258, ST129 and ST227) were also detected in the present study (97). A Chinese study reported 22 ST's in their collection, with 8 ST's (ST36, ST28, ST150, ST382, ST46, ST176, ST378 and ST63) also being detected in our study isolates. These differing results further emphasize that based on MLST, circulating GAS strains are diverse. In the present study, the 30 most prevalent ST's were largely represented by a single *emm* type which suggest that when the ST is known, the *emm* type can be predicted as previously reported (112). Isolates of different *emm* types also shared identical or nearly identical ST's further emphasizing the concordance between *emm* type and strain genetic similarities that have been reported in other studies done in the US, Romania and China (91,112,122,158).

It was very difficult to group ST's into clonal complexes (CC), defined by a founder ST and all other ST's that share four or more alleles with the founder ST (173), due to the diversity of ST's in our collection. This may explain why publications shy away from reporting CC's in their isolate collections.

5.5 Virulence related strain features

The 24 virulence related genes that were screened for amongst the study isolates were detected in varying ranges. One of the most important virulence related strain features, which are thought to play an important role in the pathogenesis of scarlet fever, are the streptococcal pyrogenic exotoxins (spe). In the present study, *spe*G and *sme*Z genes were the most common spe's, detected in 97.6% and 84.0% of the isolates, respectively. These genes are believed to be chromosomally encoded, hence their dominance in GAS isolates. Studies in Germany (invasive isolates),

Ireland (invasive isolates), the United States (invasive isolates), Australia (invasive and non-invasive isolates), India (invasive isolates), Denmark (invasive isolates) and Brazil (non-invasive isolates) have also reported a dominant prevalence of the *spe*G and *sme*Z genes amongst their isolates (103,104,112,151,174–176). Although believed to also be chromosomally encoded, the *spe*J gene was detected in only 42.6% of the isolates in the present study, consistent with other studies conducted in Australia (invasive and non-invasive isolates), the US (invasive isolates), Germany (invasive isolates) and data from 12 European countries (invasive isolates) which reported a *spe*J detection rate of 31% to 51% in their isolate collection (103,112,174,177). A Russian study reported an even lower prevalence (24%) of the *spe*J gene among their isolate collection (178). The rest of the spe genes (*spe*C, *spe*H, *spe*A, *ss*a, *spe*M, *spe*L, *spe*K and *spe*I) encoding the other streptococcal pyrogenic exotoxins were detected in varying ranges (10.7% to 42.5%) in the present study isolates. These spe genes are believed to be associated with bacteriophages and this is evident by their fluctuation in GAS isolates (174). None of the isolates in the present study were predicted to produce the exotoxins Spe_N, Spe_O and SpeP as the genes that code for these exotoxins were not detected in this study. At the time of data analysis of this study, we could not find any studies that had reported the detection of these genes (*spe*N, *spe*O and *spe*P). The presence of spe genes has been associated with specific underlying conditions. A German study reported that the presence of *spe*H, *spe*J and *spe*K was associated with chronic skin lesion whereas the presence of *spe*G was associated with diabetes (103). We unfortunately do not have data on underlying conditions in our study population and thus we were unable to assess this association. An association between the presence of specific spe genes and clinical manifestation has also been reported by an Irish group that published data highlighting that the frequency of *spe*H and *Spe*I was highest in cases of septic arthritis (104). On the other hand, a Chinese study reported that the presence of spe genes in their invasive and non-invasive isolates was not associated with clinical manifestation (43).

Different strains encode different combinations of spe genes commonly referred to as streptococcal superantigen (SAg) profiles (43). Streptococcal superantigen gene (SAg) profiles were diverse in the present study, with 70 different SAg profiles assigned to 577 isolates. Of the 9 *emm* types that were used to assess whether *emm* types exhibited a preference for specific antigen profiles, three *emm* types

(*emm*92, *emm*49 and *emm*53) showed a preference (≥80% isolates) for specific antigen profiles. In the present study, 45.5% (10/22) of *emm*1 isolates were SAg3 (A: G: J: Z), a profile reported in 93.1% (27/29) of *emm*1 isolates in an Australian study working on invasive and non-invasive isolates (174), 100% (43/43) of invasive *emm*1 isolates collected in the Netherlands (179) and 69.3% (311/449) of invasive *emm*1 isolates in a Spanish study (142). An Indian study (151) reported that 25% (1/4) of their invasive *emm*49 isolates had the same antigen profile (G: H: I: L:M) as 100% of the *emm*49 isolates in the present study. In addition to demonstrating that specific SAg profiles were associated with specific *emm* types, a Spanish study also demonstrated that isolates that were a specific *emm* type and exhibited a specific SAg profile, also showed an association with specific clinical manifestations(142). An example is their *emm* 1 isolates with SAg profile A:G:J:Z which were associated with the severe invasive diseases necrotizing fasciitis, pneumonia and streptococcal toxic shock syndrome. It is difficult to compare antigen profiles globally for two reasons: (i) there does not seem to be a standardized method or system on how to assign a SAg profiles. This is seen by how studies may randomly assign an arbitrary value (e.g. profile 1, profile 2, profile 3 etc.) to the combinations of streptococcal pyrogenic exotoxin genes detected in their studies (151,174) and (ii) *emm* type distribution differs geographically and thus so would SAg profiles.

One isolate in our collection was *emm* non-typable and no exotoxin genes were detected in this isolate. This isolate was positive for the group A carbohydrate which indicates that it is GAS. In addition, other GAS features (enn protein, fibronectin binding proteins and a hyaluronic acid capsule) were also detected, further providing support that this isolate is a GAS isolate.

5.6 Strain relatedness

Historically, identifying potential outbreaks relied on linking cases epidemiologically (time and place) as well as determining whether the isolates shared the same subtypes or sequence types through conventional molecular typing (121). Traditional bacterial typing methods may not be able to adequately discriminate between strains of the same lineage, thus making it difficult to accurately detect clusters and outbreaks (121). In contrast to conventional typing methods, whole genome sequencing exhibits superior discriminatory power when differentiating isolates

belonging to the same lineage and thus may offer valuable insight in outbreak investigations and uncovering hidden patterns of disease clusters in communities (180). In the present study, using whole genome sequence data to assess the phylogenetic relationship among circulating strains allowed us to identify the general genetic clustering of isolates by *emm* type. This suggests that isolates within the same *emm* type are genetically similar and share a common lineage. Our results are concordant with other studies which have also reported the clustering of isolates by *emm* type (121,180,181). In addition, our results suggest that there were no undetected iGAS disease outbreaks that occurred in this population during the surveyed time period as we did not observe any isolates within the same *emm* type aggregating by time and place.

To our knowledge, there is no published South African data on the usage of whole genome sequencing (WGS) to determine the genomic structure of iGAS and non-invasive GAS strains circulating in the population. The use of this method is more common in high-income countries which have shown how WGS data can be used to study relatedness among different strains, in order to detect outbreaks that may have gone undetected and to elucidate transmission dynamics. An Australian study published in 2016 used WGS in conjunction with epidemiology data to investigate whether transmission of GAS was predominantly occurring within households or across the community (182). By generating a phylogeny using WGS data, the investigators were able to discriminate between circulating strains of the same sequence types and they found that these strains were in fact genetically similar beyond just sharing an ST, thus suggesting that transmission was occurring at both the household and community level. A study in the UK identified a new emergent *emm*89 clade which was linked to the increase in *emm*89 iGAS isolates observed in the population (166). Using WGS data, phylogenetic analysis revealed two distinct clades of *emm*89 isolates and that isolates clustering in the new clade, although sharing an *emm* type, were genetically distinct from other *emm*89 isolates as a result of changes in the core genome. These studies highlight the importance and usefulness of WGS data in investigating differences beyond strain similarities based on *emm* type or ST to explain sudden surges in GAS disease.

5.7 Limitations

The whole-genome sequence based approach allowed strain characterization of surface structures and virulence related features which is lacking in our setting, but there are some limitations to our study.

We only examined iGAS disease in our setting and did not include non-invasive (cutaneous and pharyngitis) disease and so our results reflect the severe acute disease of GAS, but we have missed understanding the strains that may be associated with ARF/RHD, which are important outcomes of GAS disease especially in low-income settings like ours. Although incidence was highest in the extremes of age, not looking at pharyngitis, which is more common in 5-15 year olds, is a limitation as this group is the main transmitter of respiratory GAS disease. In addition, the genomic data of non-invasive infections is important to assess whether strain molecular characteristics are different or if there are features that are exclusive to invasive isolates.

We were unable to make a conclusion regarding the seasonality of iGAS disease in our setting due to the short time period assessed and the interruption caused by the COVID 19 pandemic. Continued surveillance is required.

iGAS disease is not notifiable in South Africa and therefore we may be missing some desirable data such as epidemiological links amongst cases, which is particularly important for detecting outbreaks/clusters. In the present study, cases are not actively followed up to investigate potential secondary cases. In addition, secondary cases may have presented with non-invasive disease and were thus not eligible for enrolment in the surveillance.

Participation of laboratories is voluntary and this means that laboratories that do not have the time or resources to keep track of iGAS isolates may not prioritize submitting these isolates for further phenotypic and genotypic characterization.

The surveillance of iGAS disease is new and thus was initially slow and incomplete, resulting in the lack of substantial patient clinical information such as underlying conditions. We were therefore unable to assess potential risk factors that are associated with an increased probability of developing iGAS disease. This is important in order to identify individuals at greater risk for developing iGAS disease so that they can be prioritized for vaccination when it becomes available. Data on clinical manifestation were also poorly collected which has limited our ability to

investigate the most common clinical manifestation amongst patients presenting with iGAS disease and in addition, if certain *emm* types are associated with more severe invasive infections. The knowledge of which *emm* types commonly result in more severe invasive infections may be important for vaccine development. Final in-hospital outcome of iGAS disease may be underestimated due to poor data collection of this variable in our study and other negative outcomes were also not captured. This means that the burden of morbidity and mortality of iGAS in our setting is still not completely understood.

Resistance to various antibiotics that are important were screened for in this study and the low resistance rate against these antibiotics is encouraging. However, inducible clindamycin resistance was not tested for and therefore phenotypic resistance against this antibiotic may be underestimated. It is important to note that for the present study, elevated MIC's against β- lactam antibiotics were defined as MIC values at the susceptibility breakpoint of these antibiotics as defined by CLSI. According to the CDC however, they have over the years collected baseline MIC data for wild-type isolates and used this data to define what constitutes an elevated MIC. iGAS disease surveillance is new in our setting and thus limited MIC data is available to determine our own MIC baseline. We cannot be certain whether all M related proteins (Mrp) detected in these isolates belongs to one of the three Mrp families that have been sequenced and targeted in the theoretical vaccine formulation. Thus, the potential vaccine coverage of a theoretical vaccine that targets both M and Mrp may be overestimated. We only performed genomic characterization of strain virulence and toxin production, phenotype testing for these features was not done to confirm whether these features are also phenotypically expressed. Finally, we did not receive isolates of all the reported cases and thus our genomic results are based on a subset of strains circulating in our setting.

5.8 Conclusion

Globally, iGAS disease has been reported as a major cause of morbidity and mortality. In our setting, the very young and elderly are at a greater risk of developing iGAS disease. The resistance rate against important antibiotics was low and no resistance was observed against β- lactam antibiotics. The dominant *emm* types in this study are different from *emm* types reported in high-income countries, and this is highlighted by the fact that *emm*1 is the most prevalent *emm* type in high income settings whereas this *emm* type was detected in less than 4% of our study isolates. The potential vaccine coverage (for iGAS) of the 30-valent M-protein based vaccine may not be sufficient in our setting and more data from African countries is needed so that African data can be considered and contribute to discussions regarding vaccine formulations. Data on the genomic surveillance of GAS is currently lacking, especially in low-middle- income settings. This work provides data and methods that could serve as the baseline for continued surveillance in an effort to control invasive disease caused by GAS.

6 REFERENCES

1. Mandell GL, Bennet JE, Dolin R. Mandell, Douglas and Bennetss's Principles and Practice of Infectious Diseases. Seventh. Mandell GL, editor. 2010. 2593–2610 p.
2. Facklam R. What Happened to the Streptococci: Overview of Taxonomic and Nomenclature Changes. Clin Microbiol Rev. 2002;15(4):613–30.
3. UK Standards for Microbiology Investigations. 2021. Available on: https://www.rcpath.org/static/ce35b1b6-9d79-4125-a0f64ba28e9a3584/UK-SMI-ID-4i4-Identification-of-Streptococcus-species-Enterococcus-species-and-morphologically-similar-organisms-September-2021.pdf. Accessed: 12-Oct-24
4. Walker MJ, Barnett TC, McArthur JD, Cole JN, Gillen CM, Henningham A, et al. Disease manifestations and pathogenic mechanisms of Group A Streptococcus. Clin Microbiol Rev. 2014;27(2):264–301.
5. Cunningham MW. Pathogenesis of group a streptococcal infections. Clin Microbiol Rev. 2000;13(3):470–511.
6. Bessen DE, Smeesters PR, Beall BW. Molecular Epidemiology, Ecology, and Evolution of Group A Streptococci. Microbiol Spectr. 2018;6(5):CPP3-0009-2018.
7. Hasty DL, Ofek I, Courtney HS, Doyle RJ. Multiple adhesins of streptococci. Infect Immun. 1992;60(6):2147–52.
8. Yamaguchi M, Terao Y, Kawabata S. Pleiotropic virulence factor - Streptococcus pyogenes fibronectin-binding proteins. Cell Microbiol. 2013;15(4):503–11.
9. Metzgar D, Zampolli A. The M protein of group A Streptococcus is a key virulence factor and a clinically relevant strain identification marker. 2011;2(5):402–12.
10. Shannon BA, McCormick JK, Schlievert PM. Toxins and Superantigens of Group A Streptococci. Microbiol Spectr. 2019;7(1):GPP3-0054-2018.
11. Lancefield RC. The antigenic complex of Streptococcus H~emolyticus. I. Demonstration of a type specific substance in extracts of Streptococcus H~molyticus. J Exp Med. 1928;47(1):91-103.
12. Påhlman LI, Malmström E, Mörgelin M, Herwald H. M protein from

Streptococcus pyogenes induces tissue factor expression and pro-coagulant activity in human monocytes. Microbiology. 2007;153(Pt 8):2458–64.
13. Sjöbring U, Ringdahl U, Ruggeri ZM. Induction of platelet thrombi by bacteria and antibodies. Blood. 2002;100(13):4470–7.
14. Shannon O, Hertzén E, Norrby-Teglund A, Mörgelin M, Sjöbring U, Björck L. Severe streptococcal infection is associated with M protein-induced platelet activation and thrombus formation. Mol Microbiol. 2007;65(5):1147–57.
15. Parks T, Smeesters PR, Curtis N, Steer AC. ASO titer or not? When to use streptococcal serology: a guide for clinicians. Eur J Clin Microbiol Infect Dis 2015;34(5):845–9.
16. Watson DW. Host-parasite factors in Group A streptococcal infections: Pyrogenic and other effects of immunologic distinct exotoxins related to scarlet fever toxins. J Exp Med. 1960;111(2):255
17. Smeesters PR, McMillan DJ, Sriprakash KS, Georgousakis MM. Differences among group A streptococcus epidemiological landscapes: consequences for M protein-based vaccines? Expert Rev Vaccines. 2009;8(12):1705–20.
18. Sanson MA, Macias OR, Shah BJ, Hanson B, Vega LA, Alamarat Z, et al. Unexpected relationships between frequency of antimicrobial resistance, disease phenotype and emm type in group A Streptococcus. Microb Genomics. 2019;5(11):e000316.
19. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. Lancet Infect Dis. 2005;5(11):685–94.
20. Gherardi G, Vitali LA, Creti R. Prevalent emm Types among Invasive GAS in Europe and North America since Year 2000. Front Public Heal. 2018;6:59
21. Procop G, Church D, Hall G, Janda W, Koneman E, Schreckenberger P, et al. Koneman's Color Atlas and Textobook of Diagnostic Microbiology. Lippincott Williams & Wilkins. 2017;70(46):1480.
22. Engel ME, Muhamed B, Whitelaw AC, Musvosvi M, Mayosi BM, Dale JB. Group a streptococcal emm type prevalence among symptomatic children in cape town and potential vaccine coverage. Pediatr Infect Dis J. 2014;33(2):208–10.
23. Carapetis JR, Beaton A, Cunningham MW, Guilherme L, Karthikeyan G, Mayosi BM, et al. Acute rheumatic fever and rheumatic heart disease. Nat Rev Dis Prim.2016;14;2:15084
24. Id PMG, Tong SYC, Holt DC, Ralph AP, Currie BJ. Concerns for efficacy of a

30-valent M-protein-based Streptococcus pyogenes vaccine in regions with high rates of rheumatic heart disease. PLoS Negl Trop Dis. 2019 Jul 3;13(7):e0007511.

25. Sliwa K, Carrington M, Mayosi BM, Zigiriadis E, Mvungi R, Stewart S. Incidence and characteristics of newly diagnosed rheumatic heart disease in Urban African adults: insights from the Heart of Soweto Study. Eur Heart J. 2010;31(6):719-27.

26. Gazzano V, Berger A, Benito Y, Freydiere A-M, Tristan A, Boisset S, et al. Reassessment of the Role of Rapid Antigen Detection Tests in Diagnosis of Invasive Group A Streptococcal Infections. J Clin Microbiol. 2016;54(4):994-9.

27. Thompson TZ, McMullen AR. Group A Streptococcus Testing in Pediatrics: the Move to Point-of-Care Molecular Testing. J Clin Microbiol. 2020;58(6).

28. Johnson DR, Kurlan R, Leckman J, Kaplan EL. The human immune response to streptococcal extracellular antigens: Clinical, diagnostic, and potential pathogenetic implications. Clin Infect Dis. 2010;50(4):481–90.

29. Ampofo K, Herbener A, Blaschke AJ, Heyrend C, Poritz M, Korgenski K, et al. Association of 2009 Pandemic Influenza A (H1N1) Infection and Increased Hospitalization With Parapneumonic Empyema in Children in Utah. Pediatr Infect Dis J . 2010;29(10):905–9.

30. Gewitz MH, Baltimore RS, Tani LY, Sable CA, Shulman ST, Carapetis J, et al. Revision of the Jones Criteria for the Diagnosis of Acute Rheumatic Fever in the Era of Doppler Echocardiography. Circulation. 2015 ;131(20):1806–18

31. Jones TD. THE DIAGNOSIS OF RHEUMATIC FEVER. J Am Med Assoc. 1944;126(8):481–4.

32. Taranta A, Stollerman GH. The relationship of Sydenham's chorea to infection with group A streptococci. Am J Med. 1956;20(2):170–5.

33. National guidelines on primary prevention and prophylaxis of Rheumatic Fever (RF) and Rheumatic Heart Disease (RHD) for health professionals at primary level.[cited2024Oct12].Available:https://www.westerncape.gov.za/text/2003/rheuma.pdf

34. Sánchez-Encinales V, Ludwig G, Tamayo E, García-Arenzana JM, Muñoz-Almagro C, Montes M. Molecular Characterization of Streptococcus pyogenes Causing Invasive Disease in Pediatric Population in Spain A 12-year Study. Pediatr Infect Dis J. 2020;38(12):1168–72.

35. Delpech G, Sparo M, Baldaccini B, Pourcel G, Lissarrague S, Allende LG. Throat carriage rate and antimicrobial resistance of streptococcus pyogenes in rural children in Argentina. J Prev Med Public Heal. 2017;50(2):127–32.
36. Anja A, Beyene G, S/Mariam Z, Daka D. Asymptomatic pharyngeal carriage rate of Streptococcus pyogenes, its associated factors and antibiotic susceptibility pattern among school children in Hawassa town, southern Ethiopia. BMC Res Notes. 2019;12(1):564.
37. Li H, Zhou L, Zhao Y, Ma L, Liu X, Hu J. Molecular epidemiology and antimicrobial resistance of group a streptococcus recovered from patients in Beijing, China. BMC Infect Dis. 2020;20(1).
38. Fay K, Onukwube J, Chochua S, Schaffner W, Cieslak P, Lynfield R, et al. Patterns of Antibiotic Nonsusceptibility Among Invasive Group A Streptococcus Infections—United States, 2006–2017. Clin Infect Dis. 2021; 73(11):1957-1964.
39. Imöhl M, Van Der Linden M. Antimicrobial susceptibility of invasive streptococcus pyogenes isolates in Germany during 2003-2013. PLoS One. 2015;10(9).
40. Mwaba C, Nakazwe R, Mpabalwani E, Lukwesa-Musyani C, Mwansa J, Chipeta J. Prevalence and antibiotic susceptibility of group A beta-haemolytic streptococcal isolates in children presenting with acute pharyngitis in Lusaka, Zambia. South African J Child Heal. 2020;14(2):99–103.
41. Villalón P, Bárcena M, Medina-Pascual MJ, Garrido N, Pino-Rosa S, Carrasco G, et al. National Surveillance of Tetracycline, Erythromycin, and Clindamycin Resistance in Invasive Streptococcus pyogenes: A Retrospective Study of the Situation in Spain, 2007–2020. Antibiot 2023;12(1):99.
42. Stevens DL, Bisno AL, Chambers HF, Dellinger EP, Goldstein EJC, Gorbach SL, et al. Practice guidelines for the diagnosis and management of skin and soft tissue infections: 2014 update by the Infectious Diseases Society of America. Clin Infect Dis. 2014;59(2):e10-52.
43. Zhang J, Gotoh K, Chen Y, Zheng Y, Yang Y, Yu D, et al. Molecular Characteristics of Streptococcus pyogenes Isolated From Chinese Children With Different Diseases. Front Microbiol 2021;12:722225.
44. Vitali LA, Zeng M, Shen Y, Cai J, Davies MR, Zhang C, et al. Identification and Characterization of Fluoroquinolone Non-susceptible Streptococcus pyogenes Clones Harboring Tetracycline and Macrolide Resistance in Shanghai, China.

Front Microbiol. 2018;1:542.
45. Meletis G, Lazaros A, Ketikidis S, Floropoulou N, Tychala A, Kagkalou G, et al. Antimicrobial resistance rates of Streptococcus pyogenes in a Greek tertiary care hospital: 6-year data and literature review. New Microbiol. 2023;46:37–42.
46. Vannice KS, Ricaldi J, Nanduri S, Fang FC, Lynch JB, Bryson-Cahn C, et al. Streptococcus pyogenes pbp2x Mutation Confers Reduced Susceptibility to β-Lactam Antibiotics. Clin Infect Dis . 2020;71(1):201–4.
47. Rampersadh K, Engel KC, Engel ME, Moodley C. A survey of antibiotic resistance patterns among Group A Streptococcus isolated from invasive and non-invasive infections in Cape Town, South Africa. Heliyon . 2024;10(13):e33694
48. Silva-Costa C, Friães A, Ramirez M, Melo-Cristino J. Macrolide-resistant Streptococcus pyogenes: prevalence and treatment strategies. Expert Rev Anti Infect Ther. 2015;13(5):615–28.
49. Weisblum B. Erythromycin resistance by ribosome modification. Antimicrob Agents Chemother. 1995 ;39(3):577.
50. Hraoui M, Boubaker IB Ben, Doloy A, Redjeb S Ben, Bouvet A. Molecular Mechanisms of Tetracycline and Macrolide Resistance and emm Characterization of Streptococcus pyogenes Isolates in Tunisia. Microb Drug Resist. 2011;17(3):377–82.
51. Yan SS, Fox ML, Holland SM, Stock F, Gill VJ, Fedorko DP. Resistance to Multiple Fluoroquinolones in a Clinical Isolate of Streptococcus pyogenes: Identification of gyrA and parC and Specification of Point Mutations Associated with Resistance. 2000;44(11):3196–8.
52. Vekemans J, Gouvea-Reis F, Kim JH, Excler J-L, Smeesters PR, O'brien KL, et al. Clinical Infectious Diseases The Path to Group A Streptococcus Vaccines: World Health Organization Research and Development Technology Roadmap and Preferred Product Characteristics. Clin Infect Dis. 2019;877(5):877–83.
53. Dale JB, Penfound TA, Tamboura B, Sow SO, Nataro JP, Tapia M, et al. Potential coverage of a multivalent M protein-based group A streptococcal vaccine. Vaccine. 2013;31(12):1576–81.
54. Pastural É, McNeil SA, MacKinnon-Cameron D, Ye L, Langley JM, Stewart R, et al. Safety and immunogenicity of a 30-valent M protein-based group a streptococcal vaccine in healthy adult volunteers: A randomized, controlled

phase I study. Vaccine. 2020;38(6):1384–92.
55. McNeil SA, Halperin SA, Langley JM, Smith B, Warren A, Sharratt GP, et al. Safety and immunogenicity of 26-valent group A streptococcus vaccine in healthy adult volunteers. Clin Infect Dis. 2005;41(8):1114–22.
56. Spellerberg B, Brandt C. Laboratory Diagnosis of Streptococcus pyogenes (group Astreptococci) [Internet]. Streptococcus pyogenes: Basic Biology to Clinical Manifestations. University of Oklahoma Health Sciences Center; 2016 [cited 2020 Sep 26]. Available from: http://www.ncbi.nlm.nih.gov/pubmed/
57. Dale JB, Penfound TA, Chiang EY, Walton WJ. New 30-valent M protein-based vaccine evokes cross-opsonic antibodies against non-vaccine serotypes of group A streptococci. Vaccine . 2011;29(46):8175–8.
58. Courtney HS, Niedermeyer SE, Penfound TA, Hohn CM, Greeley A, Dale JB. Trivalent M-related protein as a component of next generation group a streptococcal vaccines. Clin Exp Vaccine Res. 2017;6(1):45–9.
59. Barth DD, Naicker P, Engel K, Muhamed B, Basera W, Mayosi BM, et al. Molecular Epidemiology of Noninvasive and Invasive Group A Streptococcal Infections in Cape Town. mSphere. 2019;4(6):e00907-19.
60. Nelson GE, Pondo T, Toews KA, Farley MM, Lindegren M Lou, Lynfield R, et al. Epidemiology of invasive group a streptococcal infections in the United States, 2005-2012. Clin Infect Dis. 2016;63(4):478–86.
61. Seale AC, Davies MR, Anampiu K, Morpeth SC, Nyongesa S, Mwarumba S, et al. Invasive group a Streptococcus infection among children, rural Kenya. Emerg Infect Dis. 2016;22(2):224–32.
62. Tapia MD, Sow SO, Tamboura B, Keita MM, Berthe A, Samake M, et al. Streptococcal pharyngitis in schoolchildren in Bamako, Mali. Pediatr Infect Dis J. 2015;34(5):463–8.
63. Giffard PM, Tong SYC, Holt DC, Ralph AP, Currie BJ. Concerns for efficacy of a 30-valent M-protein-based Streptococcus pyogenes vaccine in regions with high rates of rheumatic heart disease. PLoS Negl Trop Dis. 2019 ;13(7):e0007511
64. Good MF, Pandey M, Batzloff MR, Tyrrell GJ. Strategic development of the conserved region of the M protein and other candidates as vaccines to prevent infection with group A streptococci. Expert Rev Vaccines. 2015;14(11):1459–70.

65. Dale JB, Niedermeyer SE, Agbaosi T, Hysmith ND, Penfound TA, Hohn CM, et al. Protective immunogenicity of group a streptococcal M-related proteins. Clin Vaccine Immunol. 2015;22(3):344–50.
66. Sekuloski S, Batzloff MR, Griffin P, Parsonage W, Elliott S, Hartas J, et al. Evaluation of safety and immunogenicity of a group A streptococcus vaccine candidate (MJ8VAX) in a randomized clinical trial. PLoS One. 2018;13(7):e0198658.
67. Batzloff MR, Hayman WA, Davies MR, Zeng M, Pruksakorn S, Brandt ER, et al. Protection against Group A Streptococcus by Immunization with J8-Diphtheria Toxoid: Contribution of J8-and Diphtheria Toxoid-Specific Antibodies to Protection. J Infect Dis. 2003;187:1598–608.
68. Barth DD, Moloi A, Mayosi BM, Engel ME. Prevalence of group A Streptococcal infection in Africa to inform GAS vaccines for rheumatic heart disease: A systematic review and meta-analysis. Int J Cardiol. 2020;307:200–8.
69. Barth D, M Mayosi B, Badri M, Whitelaw A, E Engel M. Invasive and non-invasive group A β-haemolytic streptococcal infections in patients attending public sector facilities in South Africa: 2003–2015. South African J Infect Dis . 2018;33(1):12–7.
70. Lamagni T, Guy R, Chand M, Henderson KL, Chalker V, Lewis J, et al. Resurgence of scarlet fever in England, 2014–16: a population-based surveillance study. Lancet Infect Dis . 2018;18(2):180–7.
71. Barth DD, Engel ME, Whitelaw A, Alemseged A, Sadoh WE, Ali SKM, et al. Rationale and design of the African Group a streptococcal infection registry: The AFROStrep study. BMJ Open. 2016;6(2):e010248.
72. Group A Streptococcal diseases: For health professionals - Canada.ca [Internet]. [cited 2020 Sep 28]. Available from: https://www.canada.ca/en/public-health/services/diseases/group-a-streptococcal-diseases/health-professionals.html#a5
73. NOTIFIABLE MEDICAL CONDITIONS | NICD [Internet]. [cited 2020 Sep 28]. Available from: https://www.nicd.ac.za/nmc-overview/
74. Nkgudi B, Robertson KA, Volmink J MB. Notification of rheumatic fever in South Africa -- evidence for underreporting by health care professionals and administrators. South African Med J. 2006;96(3):206–8.
75. GERMS-SA: ANNUAL SURVEILLANCE REVIEW. 2019.

76. Chen M, Yao W, Wang X, Li Y, Chan M, Wang G, et al. Outbreak of scarlet fever associated with emm12 type group A Streptococcus in 2011 in Shanghai, China. Pediatr Infect Dis J. 2012;31(9):e158-62.
77. Tse H, Bao JYJ, Davies MR, Maamary P, Tsoi H-W, Tong AHY, et al. Molecular Characterization of the 2011 Hong Kong Scarlet Fever Outbreak. J Infect Dis. 2012;206(3):341-51
78. Fernández-Prada M, Martínez-Diz S, Colina López A, Almagro Nievas D, Martínez Romero B, Huertas Martínez J. [Scarlet fever outbreak in a public school in Granada in 2012]. An Pediatr (Barc) [Internet]. 2014;80(4):249–53.
79. Lamden KH. An outbreak of scarlet fever in a primary school. Arch Dis Child. 2011;96(4):394–7.
80. Nanduri SA, Metcalf BJ, Arwady MA, Edens C, Lavin MA, Morgan J, et al. Prolonged and large outbreak of invasive group A Streptococcus disease within a nursing home: repeated intrafacility transmission of a single strain. Clin Microbiol Infect. 2019;25:248–9.
81. Nabarro LE, Brown CS, Balasegaram S, Decraene V, Elston J, Kapadia S, et al. Invasive Group A Streptococcus Outbreaks Associated with Home Healthcare, England, 2018–2019. Emerg Infect Dis. 2022;28(5):915.
82. Cornick JE, Kiran AM, Vivancos R, Aartsen J Van, Clarke J, Bevan E, et al. Epidemiological and Molecular Characterization of an Invasive Group A Streptococcus emm32.2 Outbreak. J Clin Microbiol. 2017;55(6):1837-1846.
83. Alcolea-Medina A, Snell LB, Alder C, Charalampous T, Williams TGS, Laboratory Group SM, et al. The ongoing Streptococcus pyogenes (group A streptococcus) outbreak in London, United Kingdom, in December 2022: a molecular epidemiology study. Clin Microbiol Infect. 2023;29(7):887-809
84. Disckson C, Pham MT, Nguyen V, Brubacher C, Silverman MS, Khaled K, et al. Community outbreak of invasive group A streptococcus infection in Ontario, Canada. Can Commun Dis Rep. 2018;44(7–8):182–8.
85. Worthing KA, Werno A, Pink R, McIntyre L, Carter GP, Williamson DA, et al. Biphasic Outbreak of Invasive Group A Streptococcus Disease in Eldercare Facility, New Zealand. Emerg Infect Dis. 2020;26(5):841–8.
86. Kemble SK, Westbrook A, Lynfield R, Bogard A, Koktavy N, Gall K, et al. Foodborne outbreak of group a streptococcus pharyngitis associated with a high school dance team banquet--Minnesota, 2012. Clin Infect Dis. 2013;57(5):648–

54.

87. Orataiwun P, Nilgate S, Leelasuprasri S, Cumchoo C, Pimtanothai N, Thitivichianlert S, et al. Outbreak of uncommon type of group A streptococcal pharyngitis among cadets at the Chulachomklao Royal Military Academy, Nakhon Nayok, Thailand. J Med Assoc Thail. 2002;85(1):378-82

88. 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(10):611–6.

89. Karaky NM, Araj GF, Tokajian ST. Molecular characterization of Streptococcus pyogenes group A isolates from a tertiary hospital in Lebanon. J Med Microbiol. 2014(Pt 9):1197-1204

90. Bahnan W, Hashwa F, Araj G, Tokajian S. emm typing, antibiotic resistance and PFGE analysis of Streptococcus pyogenes in Lebanon. J Med Microbiol. 2011 Jan;60(Pt 1):98-101.

91. Luca-Harari B, Straut M, Cretoiu S, Surdeanu M, Ungureanu V, Van Der Linden M, et al. Molecular characterization of invasive and non-invasive Streptococcus pyogenes isolates from Romania. J Med Microbiol. 2008 Nov;57(Pt 11):1354-1363

92. You YH, Song YY, Yan XM, Wang H Bin, Zhang MH, Tao XX, et al. Molecular epidemiological characteristics of Streptococcus pyogenes strains involved in an outbreak of scarlet fever in China, 2011. Biomed Environ Sci [Internet]. 2013;26(11):877–85.

93. You Y, Wang H, Bi Z, Walker M, Peng X, Hu B, et al. Molecular typing of Chinese Streptococcus pyogenes isolates. Mol Cell Probes. 2015 Jun 1;29(3):172–6.

94. Chen YY, Huang C Ter, Yao SM, Chang YC, Shen PW, Chou CY, et al. Molecular epidemiology of group A streptococcus causing scarlet fever in northern Taiwan, 2001–2002. Diagn Microbiol Infect Dis. 2007 Jul 1;58(3):289–95.

95. 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 Oct 1;9(10):611–6.

96. Khosa X V., Kgasha O, Mabhuza H, Moshe M, Engel K, Nchabeleng M. Molecular characterisation of group A streptococcus isolates recovered from the north-west of Pretoria, South Africa. S Afr Med J. 2021;111(5):487–90.

97. Bah SY, Keeley AJ, Armitage EP, Khalid H, Chaudhuri RR, Senghore E, et al. Genomic Characterization of Skin and Soft Tissue Streptococcus pyogenes Isolates from a Low-Income and a High-Income Setting. mSphere. 2023;8(1);e0046922.
98. Plainvert C, Doloy A, Loubinoux J, Lepoutre A, Collobert G, Touak G, et al. Invasive group A streptococcal infections in adults, France (2006-2010). Clin Microbiol Infect. 2012;18(7):702–10.
99. Darenberg J, Henriques-Normark B, Lepp T, Tegmark-Wisell K, Tegnell A, Widgren K. Increased incidence of invasive group A streptococcal infections in Sweden, January 2012-February 2013. Euro Surveill. 2013;18(14).
100. Olafsdottir LB, Erlendsdóttir H, Melo-Cristino J, Weinberger DM, Ramirez M, Kristinsson KG, et al. Invasive infections due to streptococcus Pyogenes: Seasonal variation of severity and clinical characteristics, Iceland, 1975 to 2012. Eurosurveillance. 2014;19(17):20784.
101. Smit PW, Lindholm L, Lyytikäinen O, Jalava J, Pätäri-Sampo A, Vuopio J. Epidemiology and emm types of invasive group A streptococcal infections in Finland, 2008–2013. Eur J Clin Microbiol Infect Dis 2015 3410 [Internet]. 2015;34(10):2131–6.
102. Naseer U, Steinbakk M, Blystad H, Caugant DA. Epidemiology of invasive group A streptococcal infections in Norway 2010-2014: A retrospective cohort study. Eur J Clin Microbiol Infect Dis. 2016;35(10):1639–48.
103. Imöhl M, Fitzner C, Perniciaro S, Van Der Linden M. Epidemiology and distribution of 10 superantigens among invasive Streptococcus pyogenes disease in Germany from 2009 to 2014. PLoS One. 2017;12(7):e0180757.
104. Meehan M, Murchan S, Gavin PJ, Drew RJ, Cunney R. Epidemiology of an upsurge of invasive group A streptococcal infections in Ireland, 2012–2015. J Infect. 2018;77(3):183–90.
105. Friães A, Melo-Cristino J, Ramirez M, Vaz T, Gião M, Ferreira R, et al. Changes in emm types and superantigen gene content of Streptococcus pyogenes causing invasive infections in Portugal. Sci Reports. 2019;9(1):1–10.
106. González-Abad MJ, Alonso Sanz M. Invasive Streptococcus pyogenes infections (2011–2018): EMM-type and clinical presentation. An Pediatría (English Ed. 2020;92(6):351–8.
107. Villalón P, Sáez-Nieto JA, Rubio-López V, Medina-Pascual MJ, Garrido N,

Carrasco G, et al. Invasive Streptococcus pyogenes disease in Spain: a microbiological and epidemiological study covering the period 2007–2019. Eur J Clin Microbiol Infect Dis. 2021;40(11):2295–303.

108. Zangarini L, Martiny D, Miendje Deyi VY, Hites M, Maillart E, Hainaut M, et al. Incidence and clinical and microbiological features of invasive and probable invasive streptococcal group A infections in children and adults in the Brussels-Capital Region, 2005–2020. Eur J Clin Microbiol Infect Dis [Internet]. 2023;42(5):555–67.

109. Ikebe T, Okuno R, Kanda Y, Sasaki M, Yamaguchi T, Otsuka H, et al. Molecular characterization and antimicrobial resistance of group A streptococcus isolates in streptococcal toxic shock syndrome cases in Japan from 2013 to 2018. Int J Med Microbiol. 2021 Apr 1;311(3):151496.

110. Rudolph K, Bruce MG, Bruden D, Zulz T, Reasonover A, Hurlburt D, et al. Epidemiology of Invasive Group A Streptococcal Disease in Alaska, 2001 to 2013. J Clin Microbiol [Internet]. 2016;54(1):134–41.

111. Tyrrell GJ, Fathima S, Kakulphimp J, Bell C. Increasing Rates of Invasive Group A Streptococcal Disease in Alberta, Canada; 2003-2017. Open forum Infect Dis [Internet]. 2018;5(8).

112. Li Y, Rivers J, Mathis S, Li Z, Velusamy S, Nanduri SA, et al. Genomic Surveillance of Streptococcus pyogenes Strains Causing Invasive Disease, United States, 2016–2017. Front Microbiol. 2020 Jul 24;11:1547.

113. Dunne EM, Hutton S, Peterson E, Blackstock AJ, Hahn CG, Turner K, et al. Increasing Incidence of Invasive Group A Streptococcus Disease, Idaho, USA, 2008–2019. Emerg Infect Dis. 2022;28(9):1785.

114. Arêas GP, Schuab RB, Neves FP, Barros RR. Antimicrobial susceptibility patterns, emm type distribution and genetic diversity of Streptococcus pyogenes recovered in Brazil. Mem Inst Oswaldo Cruz, Rio Janeiro. 2014;109(7):935–9.

115. Freschi de Barros S, De Amicis KM, Alencar R, Smeesters PR, Trunkel A, Postól E, et al. Streptococcus pyogenes strains in Sao Paulo, Brazil: molecular characterization as a basis for StreptInCor coverage capacity analysis. BMC Infect Dis. 2015;15(1).

116. Le Hello S, Doloy A, Baumann F, Roques N, Coudene P, Rouchon B, et al. Clinical and Microbial Characteristics of Invasive Streptococcus pyogenes Disease in New Caledonia, a Region in Oceania with a High Incidence of Acute

Rheumatic Fever. J Clin Microbiol [Internet]. 2010;48(2):526–30.
117. Williamson DA, Morgan J, Hope V, Fraser JD, Moreland NJ, Proft T, et al. Increasing incidence of invasive group A streptococcus disease in New Zealand, 2002-2012: A national population-based study. J Infect. 2015 Feb 1;70(2):127–34.
118. Boyd R, Patel M, Currie BJ, Holt DC, Harris T, Krause V. High burden of invasive group A streptococcal disease in the Northern Territory of Australia. Epidemiol Infect. 2016;144(5):1018-27.
119. Sanderson-Smith M, De Oliveira DMP, Guglielmini J, McMillan DJ, Vu T, Holien JK, et al. A systematic and functional classification of Streptococcus pyogenes that serves as a new tool for molecular typing and vaccine development. J Infect Dis. 2014 ;210(8):1325–38.
120. Kwong JC, Mccallum N, Sintchenko V, Howden BP. Whole genome sequencing in clinical and public health microbiology. Pathology. 2015;47(3):199–210.
121. Turner CE, Bedford L, Brown NM, Judge K, Török ME, Parkhill J, et al. Community outbreaks of group A Streptococcus revealed by genome sequencing. Sci Rep. 2017;7(1):8554.
122. Chochua S, Metcalf BJ, Li Z, Rivers J, Mathis S, Jackson D, et al. Population and whole genome sequence based characterization of invasive group a streptococci recovered in the United States during 2015. MBio. 2017;8(5).
123. Huebner RE, Klugman KP, Matai U, Eggers R, Hussey G, Brink A, et al. Laboratory surveillance for Haemophilus influenzae type B meningococcal, and pneumococcal disease [9]. South African Med J. 1999;89(9):924–5.
124. von Gottberg A, de Gouveia L, Tempia S, Quan V, Meiring S, von Mollendorf C, et al. Effects of Vaccination on Invasive Pneumococcal Disease in South Africa. N Engl J Med. 2014;371(20):1889–99.
125. Weinstein MP, Patel JB, Bobenchik AM, Campeau S, Cullen SK, Galas MF, et al. M100 Performance Standards for Antimicrobial Susceptibility Testing A CLSI supplement for global application. Performance Standards for Antimicrobial Susceptibility Testing Performance Standards for Antimicrobial Susceptibility Testing. 2020.
126. Metcalf BJ, Gertz RE, Gladstone RA, Walker H, Sherwood LK, Jackson D, et al. Strain features and distributions in pneumococci from children with invasive disease before and after 13-valent conjugate vaccine implementation in the

USA. Clin Microbiol Infect. 2016;22(1):60.e9-60.e29.
127. Inouye M, Dashnow H, Raven L-A, Schultz MB, Pope BJ, Tomita T, et al. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome Med. 2014;6(11):90.
128. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carriço JA. PHYLOViZ: Phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics [Internet]. 2012 May 8 [cited 2023 Dec 19];13(1):1–10.
129. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, et al. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother. 2012 Nov;67(11):2640-4.
130. Kumar Gupta S, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, et al. ARG-ANNOT, a New Bioinformatic Tool To Discover Antibiotic Resistance Genes in Bacterial Genomes. Antimicrob Agents Chemother. 2014;58(1):212-20.
131. Gardner SN, Slezak T, Hall BG. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics. 2015;31(17):2877–8.
132. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics [Internet]. 2014;30(9):1312–3.
133. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics. 2007;23(1):127-8
134. Chochua S, Metcalf B, Li Z, Mathis S, Tran T, Rivers J, et al. Invasive Group A Streptococcal Penicillin Binding Protein 2× Variants Associated with Reduced Susceptibility to b-Lactam Antibiotics in the United States, 2015-2021. Antimicrob Agents Chemother. 2022;66(9):e0080222.
135. Bulabula JKG, Ogunrombi AB, Molefe MB, Magumbeze V, Kumasamba KYG. The impact of the South African Covid-19 lockdown on incidence and severity of traumatic brain injury at Tshepong hospital: A multivariate retrospective cohort study. Heliyon. 2023;9(6):e16906.
136. Burrell R, Saravanos G, Britton PN.Unintended impacts of COVID-19 on the epidemiology and burden of paediatric respiratory infections. Paediatr Respir Rev. 2023:S1526-0542(23)00044-1.
137. Li H, Zhou L, Zhao Y, Ma L, Zhang H, Liu Y, et al. Epidemiological analysis of

Group A streptococcus infection diseases among children in Beijing, China under COVID-19 pandemic. BMC Pediatr. 2023;23(1):76.

138. McNeil JC, Flores AR, Kaplan SL, Hulten KG. The Indirect Impact of the SARS-CoV-2 Pandemic on Invasive Group a Streptococcus, Streptococcus Pneumoniae and Staphylococcus Aureus Infections in Houston Area Children. Pediatr Infect Dis J. 2021;40(8):E313–6.

139. Mäkinen TM, Juvonen R, Jokelainen J, Harju TH, Peitso A, Bloigu A, et al. Cold temperature and low humidity are associated with increased occurrence of respiratory tract infections. Respir Med. 2009;103(3):456-62

140. Cruz AA, Naclerio RM, Proud D, Togias A. Epithelial shedding is associated with nasal reactions to cold, dry air. J Allergy Clin Immunol. 2006 Jun 1;117(6):1351–8.

141. Wright CM, Moorin R, Pearson G, Dyer JR, Carapetis JR, Manning L. Increasing incidence of invasive group A streptococcal disease in Western Australia, particularly among Indigenous people. Med J Aust [Internet]. 2021;215(1):36–41.

142. Villalón P, Sáez-Nieto JA, Rubio-López V, Medina-Pascual MJ, Garrido N, Carrasco G, et al. Invasive Streptococcus pyogenes disease in Spain: a microbiological and epidemiological study covering the period 2007–2019. Eur J Clin Microbiol Infect Dis [Internet]. 2021;40(11):2295–303.

143. Watts V, Balasegaram S, Brown CS, Mathew S, Mearkle R, Ready D, et al. Increased Risk for Invasive Group A Streptococcus Disease for Household Contacts of Scarlet Fever Cases, England, 2011-2016. Emerg Infect Dis [Internet]. 2019;25(3):529–37.

144. Teatero S, McGeer A, Tyrrell GJ, Hoang L, Smadi H, Domingo MC, et al. Canada-Wide Epidemic of emm74 Group A Streptococcus Invasive Disease. Open forum Infect Dis [Internet]. 2018;5(5).

145. Factor SH, Levine OS, Schwartz B, Harrison LH, Farley MM, McGeer A, et al. Invasive group A streptococcal disease: risk factors for adults. Emerg Infect Dis [Internet]. 2003;9(8):970–7.

146. Galdas PM, Cheater F, Marshall P. Men and health help-seeking behaviour: literature review. J Adv Nurs [Internet]. 2005;49(6):616–23.

147. Francis JR, Gargan C, Remenyi B, Ralph AP, Draper A, Holt D, et al. A cluster of acute rheumatic fever cases among Aboriginal Australians in a remote

community with high baseline incidence. Aust N Z J Public Health. 2019 Jun 1;43(3):288–93.

148. Oliver JR, Pierse N, Stefanogiannis N, Jackson C, Baker MG. Acute rheumatic fever and exposure to poor housing conditions in New Zealand: A descriptive study. J Paediatr Child Health [Internet]. 2017;53(4):358–64.

149. Cannon JW, Abouzeid M, De Klerk N, Dibben C, Carapetis JR, Katzenellenbogen JM. Epidemiology and Infection Environmental and social determinants of acute rheumatic fever: a longitudinal cohort study. Epidemiol Infect. 2019;147:e79.

150. Tyrrell GJ, Bell C, Bill L, Fathima S. Increasing Incidence of Invasive Group A Streptococcus Disease in First Nations Population, Alberta, Canada, 2003-2017. Emerg Infect Dis [Internet]. 2021;27(2):443–51.

151. Haggar A, Nerlich A, Kumar R, Abraham VJ, Brahmadathan KN, Ray P, et al. Clinical and microbiologic characteristics of invasive Streptococcus pyogenes infections in north and south India. J Clin Microbiol [Internet]. 2012;50(5):1626–31.

152. Gazzaz N, Mailman T, Foster JR. Case Report: Recurrent invasive group A streptococcal infection with four-limb amputation in an immunocompetent child. BMJ Case Rep. 2018;2018:bcr2018225292.

153. Imöhl M, Imöhl I, Ren´ RR, Reinert R, Ocklenburg C, Van Der Linden M. Epidemiology of invasive Streptococcus pyogenes disease in Germany during 2003^2007. FEMS Immunol Med Microbiol. 2010;58(3):389-96

154. Ikeda T, Suzuki R, Jin W, Wachino JI, Arakawa Y, Kimura K. Isolation of group a streptococci with reduced in vitro β-lactam susceptibility harboring amino acid substitutions in penicillin-binding proteins in Japan. Antimicrob Agents Chemother. 2021;65(12):e0148221.

155. Musser JM, Beres SB, Zhu L, Olsen RJ, Vuopio J, Hyyryläinen HL, et al. Reduced in vitro susceptibility of streptococcus pyogenes to β-lactam antibiotics associated with mutations in the pbp2x gene is geographically widespread. J Clin Microbiol. 2020;58(4):1993–2012.

156. Southon SB, Beres SB, Kachroo P, Ojeda Saavedra M, Erlendsdóttir H, Haraldsson G, et al. Population Genomic Molecular Epidemiological Study of Macrolide-Resistant Streptococcus pyogenes in Iceland, 1995 to 2016: Identification of a Large Clonal Population with a pbp2x Mutation Conferring

Reduced In Vitro-Lactam Susceptibility. J Clin Microbiol. 2020;58(9):e00638-20
157. Hayes A, Lacey JA, Morris JM, Davies MR, Tong SYC. Restricted Sequence Variation in Streptococcus pyogenes Penicillin Binding Proteins. mSphere. 2020;5(2):e00090-20.
158. Lu B, Fang Y, Fan Y, Chen X, Wang J, Zeng J, et al. High prevalence of macrolide-resistance and molecular characterization of Streptococcus pyogenes isolates circulating in China from 2009 to 2016. Front Microbiol. 2017;8(6):1052.
159. Deekshit VK, Srikumar S. 'To be, or not to be'—The dilemma of 'silent' antimicrobial resistance genes in bacteria. J Appl Microbiol [Internet]. 2022;133(5):2902–14.
160. Schreckenberger PC, Ilendo E, Ristow KL. Incidence of Constitutive and Inducible Clindamycin Resistance in Staphylococcus aureus and Coagulase-Negative Staphylococci in a Community and a Tertiary Care Hospital. J Clin Microbiol. 2004;42(6):2777–9.
161. Powell LM, Choi SJ, Chipman CE, Grund ME, LaSala PR, Lukomski S. Emergence of Erythromycin-Resistant Invasive Group A Streptococcus, West Virginia, USA, 2020–2021. Emerg Infect Dis. 2023;29(5):898.
162. Lewis JS, Lepak AJ, Thompson GR, Craig WA, Andes DR, Sabol-Dzintars KE, et al. Failure of Clindamycin To Eradicate Infection with Beta-Hemolytic Streptococci Inducibly Resistant to Clindamycin in an Animal Model and in Human Infections. Antimicrob Agents Chemother. 2014;58(3):1327-31
163. Haenni M, Lupo A, Madec J-Y. Antimicrobial Resistance in Streptococcus spp . Microbiol Spectr. 2018;6(2).
164. Hua C-Z, Yu H, Xu H-M, Yang L-H, Lin A-W, Lyu Q, et al. A multi-center clinical investigation on invasive Streptococcus pyogenes infection in China. BMC Pediatr. 2019;19(1):181.
165. Malhotra-Kumar S, Van Heirstraeten L, Lammens C, Chapelle S, Goossens H. Emergence of high-level fluoroquinolone resistance in emm6 Streptococcus pyogenes and in vitro resistance selection with ciprofloxacin, levofloxacin and moxifloxacin. J Antimicrob Chemother. 2009;63(5):886-94
166. Turner CE, Abbott J, Lamagni T, G Holden MT, David S, Jones MD, et al. Emergence of a New Highly Successful Acapsular Group A Streptococcus Clade of Genotype emm89 in the United Kingdom. mBio. 2015 Jul

14;6(4):e00622
167. Proft T, Blagdon Snell L, Davies M, Beres SB, Saltini P, et al. Increase in invasive group A streptococcal infections in Milan, Italy: a genomic and clinical characterization. Front Microbiol. 2024;14:1287522.
168. Chiang-Ni C, Hsu C-Y, Yeh Y-H, Chi C-Y, Wang S, Tsai P-J, et al. Detection of toxigenic M1UK lineage group A Streptococcus clones in Taiwan. J Microbiol Immunol Infect. 2024;57(2):269-277.
169. Davies MR, Holden MT, Coupland P, Chen JHK, Venturini C, Barnett TC, et al. Emergence of scarlet fever Streptococcus pyogenes emm12 clones in Hong Kong is associated with toxin acquisition and multidrug resistance. Nat Genet 2014;47(1):84–7.
170. Henriet S, Kaguelidou F, Bidet P, Lorrot M, De Lauzanne A, Dauger S, et al. Invasive group A streptococcal infection in children: Clinical manifestations and molecular characterization in a French pediatric tertiary care center. Eur J Clin Microbiol Infect Dis [Internet]. 2010;29(3):341–6.
171. Turner CE, Pyzio M, Song B, Lamagni T, Meltzer M, Chow JY, et al. Scarlet Fever Upsurge in England and Molecular-Genetic Analysis in North-West London, 2014 - Volume 22, Number 6—June 2016 - Emerging Infectious Diseases journal - CDC. Emerg Infect Dis [Internet]. 2016;22(6):1075–8.
172. Nur Adila Hamzah S, Nasir Mohd Desa M, Syawani Jasni A, Mohd Taib N, Norbaya Masri S, Awang Hamat R. Distribution of virulence genes and the molecular epidemiology of Streptococcus pyogenes clinical isolates by emm and multilocus sequence typing methods. Med J Malaysia. 2021;76(2):164-170.
173. 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 [Internet]. 2004 Mar [cited 2022 Jul 20];186(5):1518.
174. Commons R, Rogers S, Gooding T, Danchin M, Carapetis J, Robins-Browne R, et al. Superantigen genes in group A streptococcal isolates and their relationship with emm types. J Med Microbiol. 2008;57(Pt 10):1238-1246.
175. Berman HF, Tartof SY, Reis JN, Reis MG, Riley LW. Distribution of superantigens in group A streptococcal isolates from Salvador, Brazil. BMC Infect Dis [Internet]. 2014 May 29 [cited 2022 Jul 25];14(1):294.
176. Luca-Harari B, Ekelund K, Van Der Linden M, Staum-Kaltoft M, Hammerum AM,

Jasir A. Clinical and epidemiological aspects of invasive Streptococcus pyogenes infections in Denmark during 2003 and 2004. J Clin Microbiol [Internet]. 2008;46(1):79–86.

177. Schmitz F-J, Beyer A, Charpentier E, Normark BH, Schade M, Fluit AC, et al. Toxin-Gene Profile Heterogeneity among Endemic Invasive European Group A Streptococcal Isolates. J Infect Dis [Internet]. 2003;188:1578–86.

178. Minko AG, Danilova TA, Danilina GA, Adzhieva AA, Tikhomirov EE, Zhukhovitsky VG. Molecular Genetic Characterization of Streptococcus pyogenes Strains Isolated from Patients with Various Manifestations of Streptococcal Infection. Bull Exp Biol Med 2023 [Internet]. 2023;1–5

179. Vlaminckx BJM, Mascini EM, Schellekens J, Schouls LM, Paauw A, Fluit AC, et al. Site-Specific Manifestations of Invasive Group A Streptococcal Disease: Type Distribution and Corresponding Patterns of Virulence Determinants. J Clin Microbiol. 2003;41(11):4941–9.

180. Tagini F, Aubert B, Troillet N, Pillonel T, Praz G, Crisinel PA, et al. Importance of whole genome sequencing for the assessment of outbreaks in diagnostic laboratories: analysis of a case series of invasive Streptococcus pyogenes infections. Eur J Clin Microbiol Infect Dis [Internet]. 2017 ;36(7):1173–80.

181. Chalker V, Jironkin A, Coelho J, Al-Shahib A, Platt S, Kapatai G, et al. Genome analysis following a national increase in Scarlet Fever in England 2014. BMC Genomics [Internet]. 2017;18(1).

182. Bowen AC, Harris T, Holt DC, Giffard PM, Carapetis JR, Campbell PT, et al. Whole genome sequencing reveals extensive community-level transmission of group A Streptococcus in remote communities. Epidemiol Infect. 2016;144(9):1991–8.

7 Appendices

7.1 Ethic clearance certificate (Appendix A)

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

R14/49 Ms. Sibusisiwe Zulu

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M210750

NAME: (**Principal Investigator**)	Ms. Sibusisiwe Zulu
DEPARTMENT:	School of Pathology Clinical Microbiology and Infectious Diseases National Institute for Communicable Diseases Centre for Respiratory Diseases and Meningitis
PROJECT TITLE:	Whole-genome sequence based characterization of Group A streptococci (GAS; Streptococcus pyogenes) isolates causing invasive disease in South Africa, 2019-2020
DATE CONSIDERED:	Ad hoc
DECISION:	Approved unconditionally
CONDITIONS:	Sub-study under M1809107
SUPERVISOR:	Prof A. von Gottberg and Dr S. Walaza
APPROVED BY:	Dr CB Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL:	03/08/2021

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary on the Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions.Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. I **agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in **July** and will therefore be due in the month of **July** each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).

Principal Investigator Signature Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

7.2 NICD GERMS-SA case definitions for isolate submission (Appendix B)

NICD GERMS-SA case definitions for isolate submission to Reference Laboratories for 2022

GERMS-SA laboratory surveillance: **ALL** laboratories, please submit the following bacterial or fungal pathogens to the National Institute for Communicable Diseases (NICD) on Dorset transport medium with a TrakCareLab/ private laboratory report. Alternatively, send specimen tube/blood culture bottle if uncertain of identification and/or no isolate cultured (please contact NICD reference lab before sending).

Pathogen	Specimen	Lab tests	NICD Centre
Streptococcus pneumoniae *Haemophilus* spp. *Neisseria meningitidis* Group A streptococcus (*Streptococcus pyogenes*) Group B streptococcus (*Streptococcus agalactiae*)	All normally-sterile site specimens, e.g. CSF, blood, pleural fluid, peritoneal fluid, pericardial fluid, joint fluid, tissue, etc. In addition to above for *Streptococcus pyogenes* include non-sterile body site (wound) specimen obtained from a patient with necrotizing fasciitis or streptococcal toxic shock syndrome	Culture positive **OR** Consistent Gram stain **OR** Latex positive For GAS and GBS culture and /PCR positive	CRDM 011 555 0315
Vibrio cholerae†† *Salmonella* Typhi *Salmonella enterica* serotype Paratyphi (A, B and C) Nontyphoidal *Salmonella* spp. *Shigella* spp. *Listeria* spp. *Campylobacter* spp. Diarrhoeagenic *E. coli* (**from rectal swab/stool only**)	Any specimen	Culture positive	CED 011 555 0333
Cryptococcus spp. (**no need to send isolates or specimens**)	Any specimen - enhanced surveillance laboratories need to inform the Surveillance Officers about cases (all year)	Culture positive **OR** CrAg test positive **OR** CSF India ink positive	CHARM – MRL 011 555 0384

††*Vibrio cholerae* isolates from human and non-human (environmental) specimens. CRDM = Centre for Respiratory Diseases and Meningitis, CED = Centre for Enteric Diseases, CHARM-Centre for Healthcare-Associated Infections, Antimicrobial Resistance and Mycoses, MRL = Mycology Reference Laboratory

To order a new batch of Dorset transport media, please call Siyanda Dlamini at telephone: 071 307 3262/ 062 605 5318

Level 1 GERMS-SA Surveillance **For other surveillance questions please call 011 386 6234/6012** **Version 19 for 2022**

7.3 Case report form (Appendix C)

LABORATORY CASE REPORT FORM

Lab Specimen No: ☐☐☐☐☐☐☐☐☐☐

Hospital Name: _____ Laboratory Name: _____

Lab Contact Person: _____ Lab Telephone No: _____

PATIENT

Surname: _____ First Name/s: _____

Sex: Male: ☐ Female: ☐ Date of Brth: d d m m y y y y

Ward: _____ Age: _____ Age Unit: Days: ☐ Months: ☐ Years: ☐

Hospital Number: ☐☐☐☐☐☐☐☐☐☐ Admission Date: d d m m y y y y

Province (Residence): _____ Town (Residence): _____

DIAGNOSIS AND OUTCOME

Clinical Diagnosis: Meningitis: ☐ LRTI: ☐ Bacteraemia or fungaemia: ☐
Dysentery: ☐ Diarrhoea: ☐ Other: ☐ Specify: _____

Outcome: Recovered: ☐ Died: ☐ Unknown: ☐

SPECIMEN

Collection Date: d d m m y y y y

Gram Stain: _____ India Ink: Pos: ☐ Neg: ☐ Not done: ☐

Latex Antigen / CrAg LFA: -*Cryptococcus*: Pos: ☐ Neg: ☐ Not done: ☐
-Bacterial: Pos: ☐ Specify: _____ Neg: ☐ Not done: ☐

Was the organism cultured? Yes: ☐ No: ☐

Haemophilus influenzae ☐	*Neisseria meningitidis* ☐	*Streptococcus pneumoniae* ☐
Salmonella Typhi ☐	*Shigella* species ☐	Non-typhoidal *Salmonella* species ☐
Campylobacter species ☐	*Vibrio cholerae* ☐	Diarrhoeagenic *E. coli* ☐
Enterohaemorrhagic *E. coli* ☐	*Cryptococcus* species ☐	*Candida* species (BC only) ☐
Candida auris (any site) ☐	*Streptococcus pyogenes* (GAS) ☐	*Streptococcus agalactiae* (GBS) ☐
Acinetobacter baumannii ☐		

Specimen type: Blood culture: ☐ CSF: ☐ Other site: ☐ Specify: _____

Was >1 organism type isolated from the specimen? Yes: ☐ No: ☐

If yes, specify organism/s: _____

PREVIOUS ISOLATES

Were any of the above organisms isolated from the patient previously? Yes: ☐ No: ☐

If yes, specify organism/s: _____

Date of Previous Isolate: d d m m y y y y Lab Specimen No: ☐☐☐☐☐☐☐☐☐☐

7.4 Turnitin report (Appendix D)

NATIONAL INSTITUTE FOR COMMUNICABLE DISEASES
Division of the National Health Laboratory Service

Centre for Respiratory Diseases and Meningitis

Modderfontein Road, Sandringham, 2131

Tel: +27 (0)11 386 6410 Fax: +27 (0)11 555 0437 Email: sibongilew@nicd.ac.za

University of the Witwatersrand
Faculty of Health Sciences, Postgraduate Office Parktown, 2193
Johannesburg, South Africa

To whom it may concern

RE: Turnitin report for Sibusisiwe Zulu (1304805) MScMed dissertation

Sibusisiwe Zulu is a part-time student doing her MScMed (Clinical Microbiology and Infectious Diseases) at University of the Witwatersrand. The overall similarity index score (19%) in the Turnitin plagiarism report was due to the following reasons:

- The review of literature included in the research may include some data and or quotes from existing research papers or content that is made available by experts in the field. Sources of the information have been integrated and cited accordingly.
- The report flagged methods/techniques from previous student manuscripts. These methods were conducted as per the manufacturer's instructions
- The report flagged data such as abbreviations and full names of commonly used terms in the field, names and abbreviations of features associated with *Streptococcus pyogenes*, gene and protein names, names of methods/techniques, institution and organisation names, as well as equipment names and reagents names, countries etc.

Kind regards,

Dr Sibongile Walaza

702185:S_Zulu_1304805_MSc_dissertation_FINAL.pdf

ORIGINALITY REPORT

19%	12%	14%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to North West University Student Paper	1%
2	hdl.handle.net Internet Source	1%
3	Submitted to University of Witwatersrand Student Paper	1%
4	www.ncbi.nlm.nih.gov Internet Source	1%
5	Sanderson-Smith, M., D. M. P. De Oliveira, J. Guglielmini, D. J. McMillan, T. Vu, J. K. Holien, A. Henningham, A. C. Steer, D. E. Bessen, J. B. Dale, N. Curtis, B. W. Beall, M. J. Walker, M. W. Parker, J. R. Carapetis, L. Van Melderen, K. S. Sriprakash, and P. R. Smeesters. "A systematic and functional classification of Streptococcus pyogenes that serves as a new tool for molecular typing and vaccine development", The Journal of Infectious Diseases, 2014. Publication	1%
6	academic.oup.com Internet Source	