The Molecular characterisation of Group A Streptococcus among Children with pharyngitis in the Vanguard Community (Bonteheuwel/Langa), Cape Town, South Africa

Babu Muhamed

A thesis presented in fulfillment of the requirement for the degree of Master of Science (Med) in the Department of Medicine, University of Cape Town

July 2012

DECLARATION

I, Babu Muhamed, hereby declare that the work on which this dissertation/theses is based is my original work (except where acknowledgements indicate otherwise), and that neither the whole work nor a part of it has been, is being, or it to be submitted for another degree in this or any other university.

I empower the University of Cape Town to reproduce for the purpose of research either the whole or any portion of the contents in any whatsoever.

_____

Signature

_____

Date

Thank you Lord for the most amazing ladies in my life, Johari Amisi Sefu, Asha Shangeshi Kambi and Esther Chiruza

ABSTRACT

Streptococcus pyogenes (Group A Sreptococcus or GAS) results in a variety of human disease including bacterial pharyngitis. Repeated GAS pharyngitis episodes, through the mechanism of molcular mimicry, may lead to acute rheumatic fever and a further sequella, rheumatic heart disease (RHD). In South Africa, 23.5 new cases of RHD/100,000 per annum have been estimated, being among the highest in the world; yet only few aspect of GAS diseases have been documented. This study invesitgated pharyngeal isolates from children in Cape Town, in an effort to contribute to the pathogenesis of GAS pahryngitis, were investigated.

The M-protein sequence typing (*emm* typing) techniques as well as restriction fragment length polymorphism were used to study the M-proteins (*emm* types) circulating among Cape Town children. Sequence analysis was also performed on the conserved region of the M protein for possible vaccine targets. A multiplex PCR assay was also optimised to study eight streptococcal pyrogenic exotoxins.

A total of 35 different *emm* types and subtypes from 143 pharyngeal GAS isolates were identified, *emm48.1* (12%) being the dominant *emm* type. RFLP revealed restriction profiles distinct to individual *emm* types. Sequence analyses of the conserved region of the M-protein reveal predominantly J14 and J14.1 vaccine targets while the multiplex PCR detected SAg profiles that were distributed across *emm* types.

This study present the first laboratory data conducted in a South African setting. Following *emm* typing, the experimental 26-valent GAS vaccine coverage was estimated at 51.7%, among the lowest compared to 85% coverage anticipated in developed countries such as the US and Canada. However, conserved region of the M-protein harboured predominately J14 and J14.1 sequences which are potential vaccine candidates conserved across GAS strains. Data on the extents of GAS diseases in South Africa are still limited hence more studies are essential in order to enhance understanding of these diseases and the vaccine design that will prove effective in preventing RHD.

ABBREVIATIONS

ºC	degrees celsius
ARD	acute respiratory disease
ARF	acute rheumatic fever
bp	base pair
BSA	bovine serum albumin
CDC	Centers for Disease Control and Prevention
GAS	group A streptococcus
MEME	multiple em motif elicitation
MHC	major histocompatibility complex
MLST	multilocus sequence typing
NEB	New England Biolabs
NF	necrotizing fasciitis
PCR	polymerase chain reaction
PFGE	pulsed-field gel electrophoresis
RFLP	restriction fragment length polymorphism
RHD	rheumatic heart disease
SAg,	superantigen(s)
SPE	streptococcal pyrogenic exotoxin
SARS	severe acute respiratory syndrome
SMEZ	streptococcal mitogenic exotoxin Z
SPA	streptococcal protective antigen
SSA	streptococcal superantigen A
SSTI	severe skin and soft tissue infections
STSS	streptococcal toxic shock syndrome
UAE	United Arab Emirates
USA	United States of America
WHO	World Health Organization

LIST OF FIGURES AND TABLES

Table of Contents

Literature review

1.1 *Streptococcus pyogenes*

1.1.1 Bacteriology of *Streptococcus pyogenes*

The genus *Streptococcus* comprises gram-positive, catalase-negative, oval or coccoid cells organised in pairs and chains hence the name "*Streptococcus"* [1]. *Streptococcus pyogenes* is a facultative anaerobe that grows optimally in the presence of media supplemented with blood. The haemoglobin-haptoglobin complex, hemoglobin, myoglobin, heme-albumin and catalase present in blood are sources of iron for optimum growth. This bacterial pathogen causes a wide array of clinically important infections such as pharyngitis, scarlet fever, impetigo, streptococcal toxic shock syndrome (STSS) as well as the post streptococcal diseases such as acute rheumatic fever (ARF) and rheumatic heart disease (RHD).

Streptococci were identified as early as 1874 from patients with cases of erysipelas and wound infection by Christian Albert Theodor Billroth [1], followed by demonstration of the organism in blood from patients with puerperal sepsis in 1879 by Pasteur [1]. In 1884, Rosenbach named the organism *Streptococcus pyogenes,* literally meaning pus forming chained round-shaped bacteria [1].

During the 1920s, *S. pyogenes* were classified using serological methods [2]. In 1924 Hitchcock identified a crude streptococcal extract that reacted to antibacterial sera [4]. In experiments following the discovery by Hitchcock, Lancefield discovered that antibacterial sera were homologous against streptococcal haemolytica extracts [7]. The extracts were then classified as protein (M protein) in nature following precipitation with dilute alcohol, acetic acid, picric acid and digestion by trypsin [6]. To date, the M protein remains the basis of GAS strain differentiation [8].

Investigation of scarlatinal streptococci revealed *S. pyogenes* isolates were not uniform in their serological characteristics [2], hence in 1932, Dr Rebecca Lancefield devised a serological method for characterising *S. pyogenes* using a precipitin test which classified streptococci into group A to E based on specific polysaccharides present on the cell wall [3]. The basis of this classification is on carbohydrate C, which was initially identified in isolates from animals [4] leading to investigation whether a similar carbohydrate existed on strains of human origin [5-7]. Anti-carbohydrate-C precipitin reactions showed similar titers with antiserum prepared against bacteria sharing similarities in their reaction [3]. Group A streptococci (GAS) are within the scope of this thesis hence will be the main focus of discussion.

1.1.2 Virulence of GAS

Streptococcus pyogenes has diverse virulence factors employing four main mechanisms; antiphagocytosis, adherence, colonisation and toxin production [8]. One of the primary human defence mechanisms involves phagocytosis, destruction and clearance of foreign invasion by polymorphonuclear leucocytes [8].

Group A streptococci avoid/bypass this defence mechanism through the antiphagocytosis mechanism. The M-protein is extracellular (Figure 1-1), a major virulence factor of GAS, a central focus in epidemiological studies and the main determinant of M-protein based vaccine coverage [8-11]. M-protein is encoded by the *emm* genes clustered at the *vir* locus of the chromosome [10]. The antiphagocytic effect of the M-protein, the major virulence property, is achieved through interference with opsonisation via the alternative complement pathway [12]. Binding of M-protein to complement control factors and other host proteins prevents activation of the alternate complement pathway thus protecting the bacteria from host phagocytosis by polymorphonuclear leukocytes [8, 12]. The M-proteins also bind to the Fc region of IgG to evade host phagocytosis [13, 14]. Group A streptococci treated with peptic digestion, known for removing the M-protein without distorting the integrity of the bacterial morphology, resulted in loss of serological response and function of the M-protein activities [12]. These observation were backed by resistance of M-protein positive strains exposed to fresh human blood, supporting the

hypothesis that the M-protein function in retarding the complement receptors, leading to resistance to opsonisation [12]. Another antiphagocytotic effect of GAS is achieved through a capsule comprising hyaluronic acid [15]. Encapsulated GAS resists phagocytosis while those lacking the capsule are susceptible to phagocytosis [15].

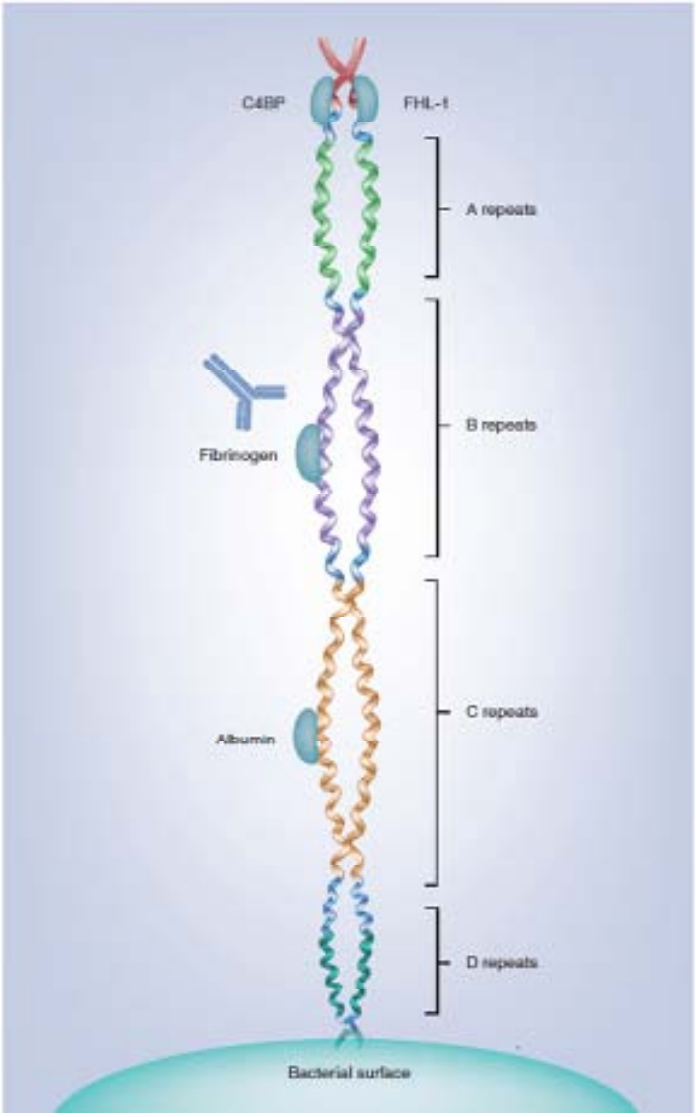


Figure 1-1. The M-protein structure. Adopted from [16].

Group A Streptococci adhere to host epithelial cells prior to colonisation and internalisation [17]. Adhesins present on GAS include; pili, lipoteichoic acid, M-protein and fibronectin binding protein [18]. Streptococcal pili have been shown to play a role as adhesins in GAS infections [19, 20]. Lipoteichoic acid adheres to fibronectin on human buccal epithelial cell [18]. Anti-lipoteichoic acid serum blocks colonization and infection of upper respiratory tract in mice [18]. The M-protein adheres to Hep-2 cells in tissue culture [21] and skin keratinocytes [22] while fibronectin binding protein such as protein F1 (streptococcal binding protein1) adheres to cutaneous Langerhans cells [23].

Colonisation is achieved by factors that evade immunoglobulins and the complement cascade [18]. Replication and production of cytolytic toxins may increase in the face of a good nutrient supply resulting in dominance over normal flora [18]. C5a peptidase produced by GAS inactivates C5a, a component of the complement pathway, minimising the influx of neutrophils thus increasing survival [18].

Among extracellular components, two haemolysins; streptolysin O and streptolysin S, play a crucial role in the pathogenesis of GAS [8]. Streptolysin O, a conserved pore-forming cytolysin, is ineffective in the presence of oxygen, lethal to erythrocytes and toxic to polymorphonuclear leukocytes, platelets, tissue-culture and lysosomes [8]. Similarly, streptolysin S is produced during growth in presence of serum albumin, alpha-lipoprotein and ribonucleic acid. Similarly to streptolysin O, streptolysin S (which is not affected by oxygen) has a destructive effect on polymorphonuclear leukocytes, platelets, tissue-culture cells and lysosomes but its activities can be affected by the presence of serum lipoproteins and phospholipids [8].

1.1.3 Streptococcal pyrogenic superantigens

The most studied superantigens (SAgs) are those seen in *S. pyogenes* and *Staphylococcus aureus* [24-27], although other microorganisms such as *Streptococcus dysgalactiae, Mycoplasma arthritidis* and *Yersinia pseudotuberculosis* also possess SAg activities. Superantigens are major virulent factors of GAS. They are extracellular protein toxins that are pyrogenic, increase host susceptibility to

endotoxic shock, suppress immunoglobulin production and have mitogenic activity for specific T-cell subsets [28]. Under normal circumstances, other antigens, when detected by the host immune system, are presented by means of antigen presenting cells on major histocompatibility complex (MHC) class II which results in induction of T-cell receptors. However, this is not the case with SAg which bypass antigen presenting cells and interact directly with MHC class II resulting in abnormal influx of T-cells [29], followed by over-secretion of pro-inflammatory cytokines (IFN-γ and lymphotoxin-α) and interleukins resulting in a high risk of developing STSS, multi-organ failure, acidosis and organ hypoperfusion [29].

There are eleven know streptococcal pyrogenic exotoxin (SPE); SPEA, SPEC, SPEG, SPEH, SPEI, SPEJ, SPEK, SPEL, SPEM, streptococcal mitogenic exotoxin Z (SMEZ) and streptococcal superantigen A (SSA) [30]. Some of these SAgs have been implicated in GAS diseases such as scarlet fever, STSS and ARF. The genes coding *speA* and *speC* have both been associated with scarlet fever and STSS [29, 31-33], while *speL* and *speM* have been identified from ARF patients, suggesting their participation in the host-pathogen interactions [34]. The streptococcal mitogenic exotoxin is a major immunological active agent and has also been associated with STSS [35]. Two toxins, SPEB and SPEF, initially classified as legitimate SAg, possess no SAg activities. *SpeB* codes for streptococcal extracellular cysteine proteinase while *speF* is identical to streptococcal DNase B [29]. Evidence of the involvement of SPEF in infected tissues (cellulitis, fasciitis and necrotising fasciitis) has also emerged [36].

Some SAg are chromosomally located while others are located on extrachromosomal elements such as phages or plasmids. [27, 30, 37, 38]. With the emergence of whole genome sequences of GAS, comes striking evidence of the presence and absence of genes coding SAg. This sequence based evidence, together with frequent detection of up to 100% of SAg genes suggests the presence of some of the SAg genes in the chromosomes [39-41]. Chromosomally implicated SAgs genes are *speG, speJ* and *smeZ*. The chromosomally coded SAg genes should be detectable in all the isolates

investigated, however, a 100% detection rate has not been achieved in all the studies [30, 37, 42], raising some doubt about the chromosomal location of these genes.

The surge in streptococcal diseases in 1980s mounted speculation of the involvement of SAg [42], with the increase largely noticed in invasive diseases such as scarlet fever and STSS [34]. However recent studies show no correlation between the presence of SAg and invasive versus non-invasive diseases [37, 41, 43], although one study characterised two pyrogenic toxins (SPEL and SPEM) associated with ARF [34]. The SAg, *speL* and *speM*, were isolated from thirty isolates associated with ARF outbreaks in the Salt Lake City, Utah, USA [44].

1.2 Molecular epidemiology of GAS

1.2.1 *emm* typing

The classical serological method devised by Lancefield initiated a platform for epidemiological studies of *S. pyogenes* [3]. More recently, with the advances in molecular biology, a sequence based method, *emm* typing, has been introduced and widely used in epidemiological studies of GAS [10]. The N-terminus of the M-protein is hypervariable (Figure 1-1) and differences in the DNA sequences of this region form the basis for the differentiation of the M-proteins, often referred as "*emm*" type. A particular *emm* type number is assigned to a sequence beginning with the three letters "*emm*" followed by a number e.g. *emm1*. Differences in the 5' region of a particular type determines the subtype e.g. *emm1.1*. The *emm* typing assay has been accepted as the most convenient and reliable method of strain typing and is now the gold standard in epidemiological studies of GAS [9, 10]. With more than 220 *emm* types documented, GAS isolates can easily be compared to known sequences that are catalogued by the Centre for Disease Control and Prevention (CDC) [45]. The *emm* typing technique will be further elaborated in Chapter Two.

1.2.2 Prevalence of *emm* types

The *emm* types circulating among the South African population are unknown. This is due to lack of molecular epidemiological studies of GAS. However, extensive review of studies and reports on global distribution of *emm* types shows similarities in the distribution of *emm* types in developed countries while on the other hand, developing countries depict differences in prevalent *emm* types [11]. Table 1-1 illustrates a summary of recent *emm* studies from different continents. The data from Table 1-1 as well as previous report on global distribution of *emm* types [11], show the geographical differences between developed and developing countries. Inter-continental similarities among the most prevalent *emm* types can be observed in developed countries while there are differences among the most prevalent *emm* types from developing countries (Table 1-1) [11]. Three studies conducted in the African continent (two from Ethiopia and one from Tunisia) revealed differences among five most prevalent types. Similarly, results from studies in Asia and South America have

shown variations in the "top five" *emm* types as shown in Table 1-1. However, North America and Europe share predominant *emm* types.

Table 1-1. Summary of recent *emm* typing studies from five continents

Continent	Country	Year	*n*	No. of *emm* types	Prevalent *emm* types (Descending order of prevalence)	Ref
Africa	Ethiopia	2006	82	43	39.2, 5.48, *st*62[a], 29.2[a], *st*463[a]	[46]
	Ethiopia	2005	217	78	12, 74, *st*62, 1, 25	[47]
	Tunisia	2011	103	38	118, 42, 1, st432, 28	[48]
Asia	China	2011	185	13	12, 1, 22, st1815, 6[a], 102[a]	[49]
	China	2008	145	5	1, 12, 4, 22, st5240	[32]
	China (Taiwan)	2009	830	21	1, 4, 12, 6, stIL103	[50]
	Fiji	2009	535	67	70, 33, 25, 11[a], 93, 69	[51]
	Fiji	2009	55	38	100, 76, 11, 33, 106	[52]
	India	2008	34	22	49, 82, 74, 71[a], 118[a], 121[a], 105[a]	[53]
	Japan	2002	316	19	12, 28, 1, 4, 13	[54]
	Lebanon	2011	103	33	1, 22, 28, 88, 4	[55]
	UAE	2010	38	25	st3211, 89, 75, st4695, st75	[56]
Europe	Austria	2006	140	25	1, 4, 12, 28, 77	[57]
	Denmark	2005	201	27	1, 28, 12, 6, 4	[41]
	Germany	2001	216	17	1, 12, 3, 28, 4	[58]
	Italy	2007	207	32	1, 12, 3, 4, 18	[59]
	Italy	2001	114	22	12, 89, 4[a], 1[a], 29, 6[a], 77[a]	[60]
	Norway	2008	101	19	28, 12, 87, 4, 89	[61]
	Romania	2008	147	19	1, 81, 76, 49, 78	[62]
	Spain	2010	673	25	6, 4, 28, 12[a], 11[a], 1	[63]
	Spain	2006	126	29	1, 3, 4, 12, 28	[37]
	Strept-EURO program	2009	4353	104	1, 28, 3, 89, 87	[64]
North America	Mexico	2003	282	27	1, 12, 75, 3, 2, 6	[65]
	Mexico	2002	40	5	2, 12, 1, 22, 89	[66]
	US	2009	7040	56	1, 12, 28, 4, 3,	[67]
	Canada	2009	1434	33	12, 1, 28, 4, 3,	[67]
	US (Hawaii)	2009	1482	93	1, 12, 28, 4,22	[68]
South America	Brazil	2006	128	48	53, 22,49, 58, 83, 8	[69]

a: similar number of isolate; *n, sample size*

1.2.3 Other molecular methods used to study the M-protein

Depending on the question investigators are trying to address in disease outbreaks or epidemiological studies, different molecular methods are used to compare bacterial strains with either local or global strains. Frequently used epityping or strain fingerprinting assays in GAS studies include pulse-field gel electrophoresis (PFGE), restriction fragment length polymorphism (RFLP) and multilocus sequence typing (MLST) [55, 70, 71].

In the context of local disease outbreaks, PFGE is considered as the gold standard [72]. This assay is second in a series of development of molecular typing techniques introduced in 1980s [73] and has since been widely used to study bacterial strains in hospital outbreaks. The principle of PFGE is based on DNA finger printing achieved by digesting the whole genome with restriction enzyme, separation of DNA fragments on agarose gel with alternating electric field and analysing the DNA fingerprint with dedicated software [72]. In GAS studies, PFGE has been used to investigate genetic relatedness and disease outbreaks [50, 55, 74-76] depicting discriminatory power to distinguish and relate GAS strains [55].

Multilocus sequence typing is a sequence based typing assay, highly effective and allows unambiguous characterisation of bacterial strain in global epidemiology [77]. This assay involves sequencing internal fragments of seven housekeeping genes; each sequence is assigned an allelic profile which defines a sequence type. Large numbers of alleles from the seven housekeeping genes provide the ability to distinguish different allelic profiles. A database has been developed for the allelic profiles of GAS representing worldwide diversity of the organism [77].

Discovered in 1970 [78, 79], restriction enzymes utilised in RFLP of GAS differentiated *emm* types based on restriction profiles following digestion of *emm* amplicon [60, 71, 80-82]. Restriction fragment length polymorphism has been recommended to developing countries due to promising results observed when

characterising the M-protein [71, 80, 83]. The effectiveness of RFLP in GAS studies will be further discussed in Chapter Three of the thesis.

1.2.4 Significance of *emm* typing

Molecular epidemiological studies utilise the *emm* typing technique to document prevalence of the *emm* types and in some cases estimate the coverage of M-protein based vaccines [10, 11, 84]. As will be elaborated in 1.6.1, the prevalent *emm* types in given area has a big impact on the coverage of M-protein based vaccines.

The *emm* typing technique has been useful in determining the GAS strain responsible for outbreak of certain diseases. For instance, in China, GAS strains responsible for scarlet fever among children were differentiated based on the *emm* types [32]; likewise isolates were differentiated from a scarlet fever outbreak in Taiwan in 2001 and 2002 [50]. Similarly GAS responsible for pharyngitis and ARF attacks have also been characterised using the *emm* typing [34, 85, 86]. Frequent isolation of certain GAS from disease sites triggers suspicion of the involment of some *emm* types in certain diseases manifestations [34]. For instance, a number of *emm* types have been labelled rheumatogenic due to their association with ARF (*emm1, emm3, emm5, emm6, emm18, emm14, emm19, emm24, emm27* and *emm29*) [34, 85]. However, the rheumatogenic *emm* types reported in a particular region may not be a true representation of global epidemiology. This is due to the fact that developed and developing countries do not share similar *emm* type distribution; hence the "rheumatogenic *emm* types" may be accounted for by circulating *emm* types Meanwhile, *emm2* has been associated with puerperal sepsis, *emm6* mostly associated with meningitis and *emm12* predominantly with GAS arthritis [35]. Two studies investigating GAS isolated from scarlet fever patients identified predominantly *emm1* and *emm12* [32, 50] while *emm1, emm3, emm6, emm12, emm18, emm19* and *emm24* have all been associated with pharyngitis [86].

Following *emm* typing, some superantigens have also been connected with certain *emm* types [30, 35, 87]. Some reports suggest the presence of conserved SAg profiles in *emm1, emm4 emm12, emm28*, and *emm75*, most notably *emm1* which has been associated with *speA, speG, speJ*, and *smeZ* but lack *speC*, *speH* and *ssa* genes [30, 35, 87]. With regards to *emm28*, one study has shown the present of *speC, speG* and *speJ,* the lack of *speA, speH* and *ssa* genes [30] although this was contradicted by another study which identified restriction profiles which depicted both presence and absence of these genes in *emm28* [42]. Interestingly, different studies have shown differences among SAg profiles from similar *emm* types, suggesting discrepancies in the present of certain SAg genes [30, 42].

1.3 Clinical manifestations of infection with GAS

1.3.1 Non-invasive diseases

1.3.1.1 Scarlet fever

Scarlet fever, mostly prevalent among children between 4 to 15 years of age [32], is associated with exotoxin SPEA [31] but other toxins such as SPEB and SPEC are produced by up to 90% of GAS [88]. Early stages of the disease are associated with white coated tongue (white strawberry tongue) and red swollen papillae after 2-3 days followed by "red strawberry tongue" after 4th and 5th day [89], disseminated rash to the head, neck and extremities, but sparing the palms and soles [88]. Post streptococcal toxin complications such as septic arthritis, rheumatic fever or post streptococcal glomerulonephritis may occur [89]. Epidemiological data on this disease remains limited in South Africa.

1.3.1.2 Impetigo

Impetigo results from superficial infection of GAS and *S. aureus* on skin [90]. Nonbullous impetigo occurs after the corneal layer of the epidermis is breached, granting GAS access to subcorneal keratinocytes [91]. The disease is common in hot, humid summer weather. Lack of epidemiological studies on impetigo may be the result of underestimation of disease in South Africa, as is the case with other streptococcal diseases. However, in Fiji, at least a quarter of school children (5-15 years of age) have active impetigo [90].

1.3.2 Invasive diseases

Severe skin and soft tissue infections

The common causative organisms in severe skin and soft tissue infections (SSTI) are *S. aureus* and GAS [87]. Common SSTI conditions are ecthyma, erysipelas, cellulitis, bursitis, necrotizing type II fasciitis and Fournier's gangrene. A breach in skin integrity renders the host susceptible to SSTI caused by GAS [88, 89]. Antimicrobial therapy with penicillin or clindamycin is often effective in treatment of these infections [89]. Epidemiological data on SSTI data are scarce in South Africa.

1.4 Streptococcal pharyngitis

1.4.1 Clinical Manifestations

Pharyngitis is defined as inflammation of the throat behind the soft palate (pharynx) [92]. Clinical manifestations such as marked odynophagia, exudative tonsillopharyngitis, anterior cervical adenitis, fever and leucocytosis are suggestive of infection by GAS [93]. Other symptoms include high temperature, headache, abdominal pain and greyish-yellow pharyngeal exudate [94]. Exudative pharyngitis is not common in children below three years of age [95]. Symptoms such as cough, hoarseness, diarrhoea and conjunctivitis are suggestive of viral infection, hence may distinguish bacterial from viral pharyngitis [93]. However, accurate differentiation of bacterial pharyngitis from viral is very difficult on clinical grounds alone.

Group A streptococcus is the most common bacterial cause of pharyngitis, primarily in children between 5 and 15 years of age [96, 97]. Other bacterial and viral etiological agents of pharyngitis exist (Table 1-2). The global burden of new GAS pharyngitis cases has been estimated around 686 million cases annually [97]. Chronic complications arising post streptococcal pharyngitis include ARF and RHD. Although post-streptococcal glomerulonephritis can occur post streptococcal pharyngitis, it is more commonly associated with impetigo. Acute rheumatic fever and RHD will be elaborated further in this chapter

Table 1-2. Microbial causes of acute pharyngitis

Pathogen	Syndrome / disease
Bacterial	
Streptococcus pyogenes (GAS)	Pharyngitis/tonsillitis, scarlet fever
Groups C and G *Streptococcus*	Pharyngitis/tonsillitis
Mixed aerobic / anaerobic infection	Gingivitis (Vincent's angina) Peritonsillitis/peritonsillar abscess (quinsy)
Neisseria gonorrhoeae	Pharyngitis
Corynebacterium diphtheria	Diphtheria
Corynebacterium ulcerans	Pharyngitis, diphtheria
Arcanobacterium haemolyticum (Corynebactirium haemolyticum)	Pharyngitis, scarletiniform rash
Yersinia enterocolitica	Pharyngitis, enterocolitis
Treponema pallidum	Secondary syphilis
Francisella tularensis	Oropharyngeal tularemia
Viral	
Rhinovirus	Common cold
Coronavirus	Common cold, SARS
Adenovirus	Pharyngoconjunctival fever, ARD
Herpes simplex	Pharyngitis, gingivostomatitis
Parainfluenza	Common cold, croup
Coxsackie A	Herpangina, hand-foot-mouth disease
Epstein-Barr virus	Infectious mononucleosis
Cytomegalovirus	Infectious mononucleosis
Human immunodeficiency virus	Primary HIV infection
Influenza A, B	Influenza
Chlamydial	
Chlamydophila pneumoniae	Pneumonia, bronchitis, pharyngitis
Chlamydophila psittaci	ARD, pneumonia
Mycoplasmal	
Mycoplasma pneumonia	Pneumonia, bronchitis, pharyngitis
Mycoplasma hominis (type 1)	Pharyngitis in volunteers

ARD, Acute respiratory disease; GAS, Group A streptococci; SARS, severe acute respiratory syndrome. Modified from [93, 98].

1.4.2 Epidemiology of GAS pharyngitis and asymptomatic carriage of GAS in South Africa and other African countries

Streptococcal pharyngitis is estimated to account for 5 to 15% of pharyngitis in adults and 20 to 30 % in children [99]. A plethora of epidemiological data on GAS pharyngitis comes from developed countries, while there is relatively little data from developing countries [100]. A South African study conducted in 1978 included children and teenagers between 2 and 19 years of age presenting with sore throat and were investigated for GAS. Positive cultures were seen in 48 (43%) of 112 children [100]. Another study conducted in Pretoria during the winter period (June-August) of 1979 and summer (October 1979 to January 1980) found beta-haemolytic

streptococci in 33.2% of 232 patients [101]. Data on scarlet fever in South Agrca is scant.

Asymptomatic carriage of GAS occurs in the throat of 15% to 20% healthy individual [102]. The GAS carriage rate observed in 1982 from a remote area of Venda (Limpopo province, South Africa) was 1.62% from 247 asymptomatic subjects between 5 and 25 years of age [103]. By contrast, a higher carriage was seen in 1983 from a study conducted in Johannesburg which also depicted seasonal variation in the prevalence of GAS from asymptomatic school children [104]. The children were from two racial groups in South Africa (Coloured and Indians) with mean age of 8.2 years for the Coloured and 7.9 years for the Indians children. Positive throat swab rates were 24.4% from 120 coloured children and 21.3% from 126 Indian children. The higher positive throat swabs within the urban setting may possibly be due to the overcrowding within Black communities at that time of South Africa's history, compared with the relatively open expanse of rural Venda

A few studies examining the epidemiology of pharyngitis have emerged from the African continent. A prevalence of 9.3% was reported in Morocco from 697 patients (494 children and 203 adults) [105]. The 697 patients were recruited prospectively in surveillance of patients presenting with sore throat in Rabat, Morocco from March 2006 to February 2007. A similar study looking at 504 children (2 to 10 years of age) in Tunisia presenting with sore throat was conducted in a primary care institute from June 2007 to May 2008. The study observed 166 (32.9%) children who were culture positive for GAS [106].

1.4.3 Treatment and control of pharyngitis

In South Africa, penicillin (oral Penicillin VK or intramuscular injection of benzathine penicillin) is the recommended first-line antibiotic for treatment of pharyngitis [107]. Intramuscular injection has shown to reduce the occurrence of ARF by 71-91% compared to oral administration [108]. Amoxycillin is recommended as an alternative to penicillin. Other antibiotics utilised for treatment of pharyngitis include amoxicillin-clavunate, azithromycin, clarithromycin,

cefpodoxime, cefprozil and cefuroxime [93, 107]. Alternatively, Alternatively, Cephalexin, Cephadroxil, Clindeamycin and Clarithromycin can be used to treat patients who are allergic to penicillin as recommended the Infectious Diseases Society of America (IDSA) the Infectious Diseases Society of America (IDSA)[109]. First generation cephalosporin are recommended for patients who do not have immediate hypersensitivity to beta-lactam antibiotics [110].

1.5 Acute rheumatic fever and rheumatic heart disease

1.5.1 Chronic but controllable disease: ARF and RHD

Acute rheumatic fever can follow GAS infection, and usually occurs following pharyngitis. Although the pathogenesis of this disease remains a mystery, molecular mimicry of the M-protein to the human cardiac myocin and other tissues is largely thought to result in development of this disease [111]. Consequently, this triggers the host antibody production against myocardium, cardiac connective tissue, valvular tissue and the smooth muscle cells of arteries [112]. Reports of rheumatogenic epitopes on S2 region of human cardiac myosin have also emerged [113]. The damage to cardiac tissue caused by an attack of ARF leads to RHD. At least 15.6 million people are affected by RHD worldwide [97], with Sub-Sahara Africa holding most cases of ARF/RHD [97].

In summary, following infection with GAS and mounting of immune response, GAS antigens are presented to CD4 T cells by antigen presenting cells. B-cells are then activated to elicit antibodies against GAS cell wall; however these antibodies also recognise host proteins in joints, myocardium, cardiac connective tissue, valvular tissue and the smooth muscle cells of arteries. This leads to stimulation of inflammatory cytokines followed by stimulation of complement cascade [12], neutrophils and macrophages which results in destruction of the host tissue. The aftermath of ARF, mostly as a result of ARF sequelae, present the most common heart and vulvular disease in developing countries, RHD [114].

1.5.2 Epidemiology of ARF and RHD in South Africa and Developed world

The Heart of Soweto study is the first study to present the incidence of new-onset RHD cases in patients above the age of 14 in Africa. Three hundred and forty four new cases of RHD were reported with the incidence of new cases calculated at 23.5 cases/100 000 per annum [115]. Developed countries, particularly the USA, have noticed a decline in number of ARF/RHD cases, and this has been attributed to the emergence of antibiotics and better living conditions [116]. As a result, an annual mean incidence rate of first attack of ARF in the USA and Western Europe is, 0.5-3

cases/100000 per annum [112, 117, 118]. The resurgence of ARF, diagnosed by echocardiography in Italy between January 2007 and December 2008 revealed an incidence of 23 and 27/100000 new cases in 2007 and 2008, respectively, in patients 5 to 15 years of age [116].

1.5.3 Treatment and control of ARF and RHD

In Africa, the Pan African Society of Cardiology in collaboration with the World Heart Federation and WHO adopted four strategies aimed at curbing the ARF/RHD burden. These strategies involve awareness, surveillance, advocacy and prevention (ASAP) [119]. The pillars proposed for the control of ARF and RHD include education, primary prevention through penicillin treatment of suspected streptococcal sore throat, secondary prevention through register based penicillin prophylaxis and surveillance through notification of ARF [120].

Education remains an essential tool in the fight against RHD. The Drakensberg Declaration on the control of ARF and RHD in Africa recognises awareness as means of educating the public and healthcare professionals [119]. Awareness of ARF/RHD is emphasised among child caregivers, teachers, and health professionals [121]. In a study which investigated eight guardians of children affected with ARF/RHD, six guardians admitted ignorance of the disease while seven parents lacked knowledge of ARF/RHD prior to the child's diagnosis [122]. The success of raising awareness through education can be witnessed by a study conducted in French Caribbean Islands (Martinique and Guadeloupe) which documented a decline in ARF/RHD incidences by 78% in Martinique and 74% in Guadeloupe, respectively, following 10 years of awareness programme on ARF [123].

Primary prevention of ARF is the first line prevention of the initial ARF attack. Penicillin is the main antibiotic proven to contain the burden of ARF/RHD [124]. Treatment of patients with known or suspected streptococcal sore throat with penicillin forms the primary prevention of ARF, hindering its notorious sequelae, RHD [124]. A meta-analysis of clinical trials looking at penicillin prescription

observed up to 70% reduction of ARF when penicillin was administered to patients with symptoms suggestive of streptococcal sore throat [124]. Intramuscular benzathine penicillin or oral phenoxymethyl penicillin are recommended for primary prevention of ARF [125].

Secondary prevention of ARF is the administration of antibiotic for prevention of recurrent ARF attack to individual with history of ARF. The antibiotics are administered to prevent new GAS infections which may result in new ARF attacks. Antibiotic treatment for secondary prevention of recurrence of ARF are similar to those prescribed for primary prevention of ARF except for single dose recommended for primary prevention whereas dosage is given two to four weekly in secondary prevention [125]. Intramuscular penicillin is much more effective in prevention of recurrence of ARF compared to oral penicillin [108]. When penicillin is administered intramuscularly, recurrence of ARF is decreased by 87-96% [108]. Frequent injections (2-weekly) reduce the recurrence of ARF by up to 50% compared to 4-weekly injections [108]. Patients with history of penicillin allergy can be treated with oral erythromycin [125]. The WHO recommendations for antibiotic duration for secondary prevention of ARF/RHD are presented in Table 1-3.

Table 1-3. Recommended duration of secondary prevention of ARF/RHD by WHO

Category of patients	Duration of secondary prevention
Patient without proven carditis	For 5 years after last attack, or until 18 years of age (whichever is longer)
Patients with mild carditis (mild mitral regurgitation or healed carditis	For 10 years after the last attack or at least until 25 years of age (whichever is longer)
More severe valvular disease	Lifelong
After surgery	Lifelong

Adapted from [125]

Notification of ARF/RHD is believed to be an important component in controlling the burden of this disease [126]. In the South African context, surveillance of ARF/RHD in the form of notification has been proven to have discrepancies resulting in under reporting of ARF/RHD cases [127]. Proper notification of initial diagnosis of ARF/RHD cases may provide more epidemiological data and enhance patient's enrolment in secondary prevention programmes [127]. Since it has been proven that patients enrolled in primary prevention are less likely to develop ARF/RHD [124], an efficient surveillance system may assist in monitoring the incidences of ARF/RHD [126].

1.6 GAS vaccine

1.6.1 M-protein vaccine

The M-protein based vaccine is the most recent GAS vaccine to reach human clinical trials [128-130]. The N-terminus domain or hypervariable region of the M-protein (10 to 35 amino acid residues) elicits an M-protein type-specific antibody, making it a vaccine target. M-protein based vaccines containing whole cell or highly purified M-protein developed in the 1930s and 1970s were abandoned due to cases of rheumatic fever following vaccination [131-135]. However, the development of the M-protein based vaccine was given a boost by advances in molecular biology towards the end of the twentieth century. Using recombinant techniques it was possible to prepare a vaccine which excluded the harmful tissue cross-reactive epitopes [136], thus minimising the possibility of ARF attacks in vaccine recipients. Successful immunogenicity of a recombinant tetravalent vaccine (M25, M5, M6 and M19) developed in 1993 [137] led to development of an octavalent vaccine (M24, M5, M6, M9, M1, M3, M18 and M2) in 1996 [138]. However, subsequent to the development of both the tetravalent and octavalent GAS vaccine, the carboxy-terminus components of the vaccine were found to have no immunogenicity or elicited low levels of antibodies that were not opsonic. Thus, in 1999 a vaccine containing six M-proteins (M24, M5, M6, M19, M1, and M3) was designed to optimize the immunogenicity [139]. Broader multivalent GAS vaccines containing 26 and 30 M-proteins were introduced in 2002 and 2011, respectively (Table 1-4) [140, 141]. The GAS vaccine containing 30 M-proteins is anticipated to have potential coverage that may extend beyond the M-proteins incorporated in the vaccine [140]. The global coverage of the experimental 26-valent vaccine has been estimated at 69.7% [11] while in North America, the figure is thought to be 85% [142]. At the time of writing this thesis, no published data on the efficacy and coverage of the 30-valent GAS vaccine were available.

Table 1-4. Progress in development of M-protein vaccine

Year	Vaccine	M-proteins included	Ref
1983	M5 vaccine	M5	[136]
1993	tetravalent	M25, M5, M6 and M19	[137]
1996	octavalent	M24, M5, M6, M9, M1, M3, M18 and M2	[138]
1999		M24, M5, M6, M19, M1, and M3	[139]
2002	26-valent	M24, M5, M6, M29, M14, M1, M12, M28, M3, M1.2, M18, M2, M43, M94, M22, M11, M59, M33, M89, M101, M77, M114, M75, M76, M92, spa	[141]
2011	30-valent	M1, M3.1, M6.4, M2, M18, M28, 12, spa, M4, M5.14, M11, M75, M19, M29.2, M14.3, M24, M77, M22, M73, M89, M58, M44, M78, M118, M83.1, M82, M81, M87, M49, M92, M114	[140]

spa, streptococcal protective antigen

The 26-valent vaccine remains the only GAS vaccine that has entered clinical trials after the year 2000 yet researchers are still concerned over geographical differences in the distribution of M-protein. With over 150 known M-proteins and only 30 of these protein included in the recent 30-valent vaccine [140], geographical distribution of *emm* types remains a concern although the developers of the 30-valent vccine have notice vaccine coverage that extends beyond the *emm* types included in the 30-valent vaccine [11]. Since global distribution of *emm* types has shown geographical differences among prevalent *emm* types in developed countries and developing countries (Table 1-1) [11] and only *emm* types of epidemiological importance in developed countries are included in the vaccine [140], the formulation of the 26-valent and the 30-valent vaccine has proved its potential effectiveness in developed countries but less effective in developing countries [11].

1.6.2 Conserved region of the M-protein: Answer to M-protein vaccine?

The C-repeats of M-protein are highly conserved across GAS strains [143], a property that brings hope of developing a GAS vaccine that does not confer type specificity, hence ideal for coverage of the multiple GAS strains [144]. Human serum albumin, CD46 and factor H bind to this proximal region of the M-protein [143]. Two peptide sequences on the C-repeats, J8 and J14 [145], when intranasally and subcutaneously delivered in mice models elicit immune response [146]. Following intraperitoneally challenging mice with GAS, immunized mice (10 days following immunization) showed a higher anti-J14 antibody titre, protecting 66-83%, contrasting with non-immunised mice with survival of between 0 to 16.6% [145].

Similarly, another study has shown 84% survival of mice immunised with J14 compare to 20% for non-immunized [147]. The C-repeat region of the M-protein will be discussed further in Chapter Four.

1.6.3 Non–M-protein vaccine

Mounting concerns over the M-protein's contribution to possible cross-reactivity with human tissues [131-135] and a low coverage reported in some developing countries [11] have contributed to the quest for alternative vaccine candidates. To date, a number of potential non-M-protein vaccine candidates have been proposed. These include streptococcal pili, streptococcal C5a peptidase and streptococcal hemoprotein receptor (shr).

Group A streptococcal pili are involved in bacterial-host interaction. This relationship is enhanced by the ability of pili to adhere onto host epithelial cells and form biofilm [20, 148]. The extracellular exposure of pili confers vaccine developer an opportunity to explore the potential use of this component of GAS as a vaccine target [149-152]. The pili based vaccine has been anticipated to confer protection against more than 90% of GAS strains [151].

C5a peptidase is a virulence factor of GAS which evades the complement system by inactivating serum chemotactic activity [153]. Group A streptococci have similar C5a peptidase hence this can serve as a vaccine target across GAS strains [154, 155]. When immunized with C5a peptidase, immunized mice cleared GAS from the nasopharynx and prevented nasopharyngreal colonisation compared to unimmunized mice [155]. Vaccination with streptococcal C5a peptidase derived from GAS induced sufficient immune response to clear streptococci from nasopharynx. Hence the C5a vaccine is anticipated to confer protection that will assist in clearance of skin infections [156]. Although there is promising progress in the development of C5a peptidase vaccine, human trials are yet to be conducted on this vaccine.

Streptococcus pyogenes requires hemoprotein receptor for iron acquisition although other mechanisms exist [157]. This haemoglobin binding protein is conserved among *S. pyogenes*, imbedded in the membrane and capable of inducing host immune response [157, 158]. Mice models challenged with purified streptococcal hemoprotein receptor (shr) expressed on *Lactococcus lactis* elicited shr-specific IgG and IgA after intraperitoneal and intranasal immunization, respectively [159]. Encouraging results were also observed from passive immunity with rabbit shr antiserum rendering this vaccine target essential in production of rabbit antiserum that may be used in the therapy of invasive diseases such as streptococcal toxic shock syndrome [159]. A vaccine targeting shr is also yet to proceed to clinical trials.

1.7 Conclusion

South Africa is among the developing countries which the prevalent of both ARF and RHD is still highly under reported [2, 5]. By the year 2010, 23.5 new cases of RHD/100000 per annum had been estimated [115], among the highest in the world yet the etiological agent, *S. pyogenes* (GAS) has not drawn attention from South African researchers. South Africa, together with other developing countries, is trailing in the efforts of containing the burden of ARF and RHD [127, 160]. This can be attributed to, in part, a lack of streptococcal data which, when intensively documented, can improved the understanding of GAS diseases epidemiology and change the approach towards this devastating disease. The extent of ARF and RHD in South Africa calls for a full investigation of the molecular epidemiology of GAS to best understand the circulating M-protein types, the most prominent virulence factor as well as a vaccine target. Early studies date back to the 1970s and 1980s [100, 101, 103, 104], and from the 1990s to the present, very few clinical or laboratory based studies have surfaced from South Africa, a handful being conducted by researchers in the aspects of ARF and RHD [119-122, 124, 125, 161].

SCOPE AND OBJECTIVES OF THE THESIS

Aim

- The primary aim of this study was to describe the molecular epidemiology of GAS isolates causing pharyngitis in children from Cape Town
- To estimate the coverage of the 26-valent GAS vaccine in the context of the molecular epidemiology of GAS from children in Cape Town

Objectives

- To perform *emm* typing of GAS isolates causing pharyngitis using PCR and DNA sequencing
- To perform RFLP analysis of the GAS isolates causing pharyngitis and to determine whether this approach is a useful alternative to sequence based analysis in South Africa
- To characterise the C-repeats of the M-protein using DNA sequence analysis
- To determine the SAg gene profile of pharyngeal strains of GAS
- To investigate the association between the presence of certain SAg and *emm* types

Characterisation of the *emm* gene using sequencing

2.1 Introduction

Although GAS has a number of potential vaccine targets, M-protein has become the most favourable and highly researched [46, 47, 129, 162]. M-protein is encoded by the *emm* gene [8] and consists of four structural repeat blocks that have been intensively explored in epidemiological studies of GAS [8]. The four repeat blocks, namely A-D (Figure 2-1), comprise amino acid sequences of different sizes [8]. Block A is hypervariable, block B is semi-variable, while block C is located proximal to the cell membrane and highly conserved among the GAS strains [102].

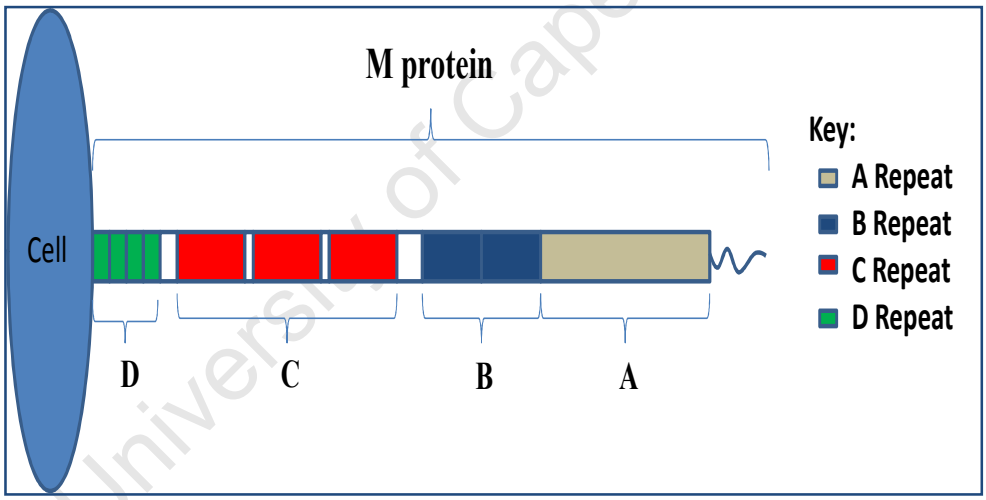


Figure 2-1. Schematic representation of the four repeats of M-protein based on literature [163].

The development of more broadly protective vaccines has resulted from an increased understanding of the epidemiology of GAS [142, 164]. The molecular technique of *emm* sequence typing of the N-terminal region of M-protein, was developed in 1996 and has subsequently become the method of choice for studying the molecular epidemiology of GAS [9, 10]. Briefly, the technique involves the amplification of a

large portion of the *emm* gene in a PCR assay and DNA sequencing of the N-terminal region is used to determine the *emm* type [10, 165].

A comprehensive review of articles and reports on the distribution of *emm* types revealed a plethora of data from developed countries compared with those from less developed countries [84]. Currently, Ethiopia and Tunisia are the only African countries which have documented the *emm* types circulating in their population [46, 48]. The Ethiopian study comprised 82 GAS isolates from throats of asymptomatic school children (6-14 years of age) while the Tunisian study included 103 isolates [46, 48]. In Ethiopia, 43 different *emm* types were identified in 82 GAS isolates with *emm3.19* being the most prevalent; whereas in Tunisia 38 *emm* types were identified in 103 GAS isolates with *emm118* the most prevalent in the GAS studied. These *emm* types were not common in studies on GAS reported from North America (USA and Canada) and Europe [142, 166]. The surveillance study from North America identified 56 *emm* types in 7040 GAS isolates from the USA among which *emm1* and *emm12* were the most prevalent [142]. In the same study involving 1434 GAS isolates from Canada, similar results were obtained [142]. As outlined earlier (1.2.2) the distribution of *emm* types in Europe is similar to those seen in North America. In Germany, among 216 isolates investigated, *emm1* and *emm12* were the most prevalent at 18.5% and 15.7% respectively. [58].

Epidemiological studies suggest a geographical difference in the distribution of GAS [11]. Because data from the developed world have been used to inform the selection of *emm* types in the 26-valent vaccine, the potential coverage of this vaccine is greater in these countries than in the developing countries [46, 64, 84, 166]. An Ethiopian study demonstrated that 46% of the 43 *emm* types were not included in the 26-valent GAS vaccine [46]. The low coverage in developing countries is primarily due to the lack of epidemiological studies that could be used to informing the selection of relevant *emm* types included in the vaccine [84].

Currently, no epidemiological data exists on the molecular epidemiology of GAS in South Africa; thus, the protective coverage of the 26-valent vaccine is unknown. This study aims at presenting the first examination of *emm* types in Cape Town, South Africa.

2.2 Experimental protocol

The *emm* typing technique involves a number of steps which are detailed below (Figure 2-2). As an overview, the procedure begins with the culture and identification of GAS from throat swab samples (A), after which DNA is extracted (B) and subjected to PCR amplification of the *emm gene* (C). The PCR products are separated and visualized using agarose gel electrophoresis (D), purified (E) and sequenced (F). The DNA sequences generated are submitted to the CDC for comparison with sequences in existing databases (G) thereafter the submitted sequence is assigned an *emm* type; e.g., *emm2* and when necessary a subtype, *emm2.1*.

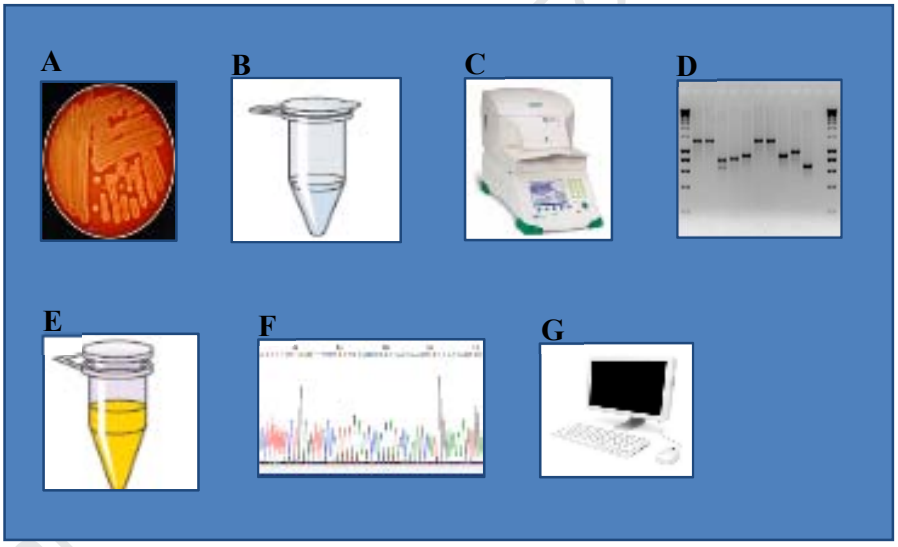


Figure 2-2. Sequence of steps in *emm* typing. **A** culture and identification of GAS. **B** genomic DNA extraction. **C** PCR amplification of the *emm* gene. **D** Agarose gel electrophoresis. **E** Purification of PCR amplicons. **F** sequencing of *emm* PCR products. **G** sequence submission to CDC database for assignation of *emm* types.

2.2.1 Selection of study population

Participants for this study were from Bonteheuwel and Langa, two periurban communites located approximately 10 kilometers from the central business distract of Cape Town [See Map]. The area is characteristed by lower socioeconmic conditions with high unemployment and a significant proportion of the residents of Black or mixed ancestery, living in informal housing. Pharyngeal isolates were collected from May 2008 until April 2012 from two groups of children: the first group, consisted of children between the ages of 3 and 15 years presenting with a sore throat to the Vanguard Community Health Centre. A research nurse obtained informed consent from parents or legal guardians of the children prior to their enrolment in the study, and then obtained pharyngeal swabs from the children. Individuals who received antibiotics within 30 days of enrolment were excluded. Where parents were unavailable or unwilling to allow participation in the study, the children were excluded with no predujice. The second group, consisting of asymptomatic school children across all school grades, had throat swabs taken after informed consent. Children older than 8 years had to provide assent in addition to the parents providing consent. Throat swabs were submitted to the National Health Laboratory Services, Microbiology Laboratory at Groote Schuur Hospital for culture and isolation of GAS.

Map: The Vanguard Community Demonstration Site

2.2.2 Culture and identification of β-haemolytic streptococci

Upon receipt at the Microbiology Laboratory, samples were inoculated on 2% blood (horse) agar and incubated in the presence of 5% CO_2 for 48 hours at 37°C (NuaAire Inc, USA). Culture on blood agar (horse) represents the gold standard for isolation of *S. pyogenes* [167]. *Streptococcus pyogenes* is requires supplementation of whole blood for observation of beta-haemlysis hence assist identification [167]. The colonies were further identified using a commercial streptococcal serogrouping kit by the Microbiology Laboratory. Isolates identified as belonging to Lancefield Group A, C or G were stored in storage media containing beads a -70°C freezer (New Brunswick Scientific, England). Only isolates conformed as belonging to Lancefield Group A were investigated further.

2.2.3 Genomic DNA extraction

Isolates were subcultured from frozen stocks onto blood agar and incubated in presence of 5% CO_2 for 24 hours. Single colonies were suspended in 1 ml of autoclaved water in Eppendorf tubes (Eppendorf, Germany). After emulsifying the colonies, the bacteria were pelleted using a microfuge for 1 minute at 16500 x *g* (Eppendorf, Germany). The supernatant was discarded and genomic DNA was extracted from the bacterial pellet using an InstaGene matrix DNA extraction kit (Bio-Rad, France) as recommended by the manufacturer. Using a 1000 µl pipette tip, as instructed, InstaGene matrix was added to the pellet and incubated for 30 minutes at 56°C (Kendro, German). After incubation, the tubes were vortexed at level 8 on a Vortex-2 Genie (Scientific Industries, USA) for 10 seconds and placed on a heating block (Techne LTD, USA) at 100°C for 8 minutes. The Eppendorf tubes were then vortexed again at level 8 for 10 seconds followed by centrifugation in a microfuge at 16500 x *g* for 3 minutes. The supernatant, containing the extracted DNA, was used as a template for PCR amplification.

2.2.4 Polymerase chain reaction

The *emm* gene was selectively amplified using PCR according to the CDC protocol [10, 168]. This molecular method selectively amplifies a target gene or a portion of a gene. The amplification is achieved by the use of sequence-specific oligonucleotides, often referred to as primers. Primers work in pairs by binding on

either strand of double-stranded DNA. The primers used in *emm* typing anneal to a conserved sequence internal to the *emm* gene [168]. The primers (Table 2-1) used in this assay were synthesized at the Department of Molecular and Cell Biology, University of Cape Town, South Africa. Polymerase chain reaction amplification of DNA is achieved by the work of an enzyme called DNA polymerase I (*Taq* Polymerase) [169] isolated from the bacterium *Thermus aquaticus* [169]. *Taq* Polymerase has the ability to withstand high temperatures during PCR cycles.

Table 2-1. Primers used for *emm* typing

Target gene	Primer	Primer sequence
***emm* gene**	Primer 1 (forward)	TAT T(C/G) GCT TAG AAA ATT AA
	Primer 2 (reverse)	GCA AGT TCT TCA GCT TGT TT
	Emmseq2	TAT TCG CTT AGA AAA TTA AAA ACA GG

To achieve a successful amplification of the *emm* gene, a combination of reagents were used at specific concentrations recommended by the CDC [168]. PCR reagents comprised 1X buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTP, 70 picomol/µl of the forward and reverse primers and 1.5U of SuperTherm *Taq* (JMR, Holdings, London UK) then brought to a final volume of 50 µl with DNAse-free water.

Cycling conditions for amplification of the *emm* gene involves the three basic principle amplification steps during PCR, namely; denaturation, annealing and elongation. Cycling conditions for *emm* typing, as recommended by the CDC [168], begins with initial denaturation at 94ºC for a minute followed by 30 cycles of denaturation at 94ºC for 15 seconds, annealing at 46.5ºC for 30 seconds and extension at 72ºC for 75 seconds. Additional 20 cycles follow with denaturation at 94ºC, annealing at 46.5ºC for 30 seconds, extension at 72ºC for 75 seconds with a 10 second increment for each of the subsequent 19 cycles. Final extension was carried out at 72ºC for 10 minutes.

2.2.5 Agarose gel electrophoresis

Agarose gel electrophoresis is a technique used to separate DNA fragments based on their sizes [170] when subjected to an electrical current. DNA has a negative charge due to the phosphate groups on its sugar-phosphate backbone and thus migrates to the positive end of the agarose gel when an electric current is passed through the gel. The percentage of agarose within a gel determines the pore size; hence different percentages of gels are prepared based on the size of DNA fragments to be separated. Larger DNA fragments migrate slower through an agarose gel than smaller fragments [170].

Agarose gel electrophoresis was used to separate the DNA fragments following PCR. As the expected sizes of *emm* amplicons range from 750 bp to 1400 bp, a 2% agarose gel was used to separate the PCR products. Agarose (SeaKem® LE Agarose, Lanza, USA) was dissolved in 1X TAE buffer (Appendix A) and ethidium bromide (10 ng/µl), which intercalates between base pairs of DNA and fluoresces under ultraviolet light. Eight microliters of the PCR product was loaded into each well, with 2 µl of loading dye. The fragments were separated by passing an electric current through the agarose for 3 hour of 60V to allow the negatively charged DNA fragments to migrate towards the positive electrode. The rate of migration is inversely proportional to the size of the fragments. A molecular marker (HyperLadder IV, Bioline, UK) was included to enable estimation of *emm* amplicon sizes.

2.2.6 DNA purification

The DNA sequencing assay is very sensitive and requires a purification step to minimise inhibition by PCR reagents within the sample. The MinElute PCR purification kit (Qiagen, Valencia, CA, USA) incorporates spin-column technology using silica-gel membranes with selective binding properties to purify DNA directly from PCR products.

Five volumes of PB buffer was added to one volume of PCR product and mixed as per manufacturer's instructions. To achieve a pH of 5.0 as recommended, 10 µl of 3

M sodium acetate was added and mixed. The mixture was then added to 2 ml MinElute columns and centrifuged in a microfuge (Eppendorf, German) for 1 minute at 17900 x *g*. The flow through products were discarded and 750 µl of Buffer PE was added to the column which was again centrifuged for a minute at 17900 x *g*. Again, the flow through product was discarded and the column was centrifuged for a minute at 17900 x *g*. MinElute columns were then transferred to sterile 1.5 ml Eppendorf tube and the DNA was eluted following the addition of 10 µl EB buffer to the centre of the column membranes and centrifuged for a minute at 17900 x *g*. Purified DNA products were stored at 4ºC.

2.2.7 Sequencing and assigning *emm* types

Sequencing of purified DNA was done using the ABI Prism® BigDye™ Terminator Cycle Sequencing kit (Applied Biosystems, USA) at Stellenbosch University, South Africa. The DNA concentration was adjusted to 10 ng/µl for amplicon less than 1000-bp and 20 ng/µl for amplicons more than 1000-bp as recommended by the sequencing facility at the Stellenbosch University, South Africa. Primer emmseq2 (Table 2-1), recommended by the CDC, was prepared to 1.1 ng/µl and used for sequencing reaction [168]. Sequences generated were analysed using BioEdit v7.0.9 (Ibis Biosciences, USA). The sequences were submitted electronically to the *S. pyogenes emm* sequence database centre at the CDC which assigned all the *emm* types and subtypes [45].

2.3 Results

2.3.1 *emm* types from GAS

A total of 143 GAS isolates obtained from children aged between 5 and 12 years were included for emm typing. Symptomatic children contributed 118 isolates while 25 came from asymptomatic children. Initial confirmation of successful amplification was observed on a 2% agarose gel. Following PCR and agarose gel electrophoresis, amplicons ranging in size from 720 bp to 1150 bp were generated from the isolates, as illustrated in Figure 2-3.

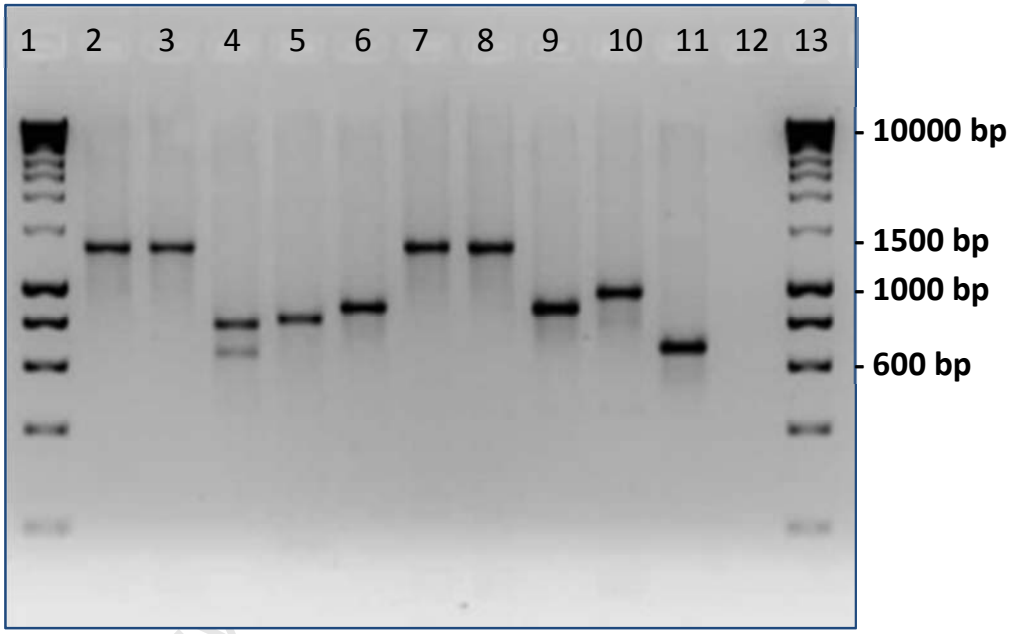


Figure 2-3. PCR amplicons of *emm* gene after electrophoresis on a 2% agarose gel. Lanes 1 and 13 are molecular markers. Lanes 2-11 are different clinical isolates of GAS. Lane 12 is a negative control of distilled water.

Sequences were obtained from all of the 143 isolates. Where necessary, electronic sequences were edited using information on the BioEdit chromatogram. Edited sequences were assigned a particular sequence type by the CDC BLAST-*emm* [45]. An *emm* type was assigned to each of the 143 isolates (Table 2-2). As shown in Table 2-2 and in the bar graph (Figure 2-4) a total of 35 different *emm* types and subtypes were identified with *emm48.1* being the dominant type. The eight most

prevalent *emm* types in descending order, accounted for 58% of the isolates, *emm48, emm12, emm4, emm89*, *emm94, emm75, emm1* and *emm22* (Table 2-2). Of the *emm* types identified in this study, 14 are included in the 26-valent GAS vaccine while 16 feature in the novel 30-valent vaccine (Table 2-2) [140, 141].

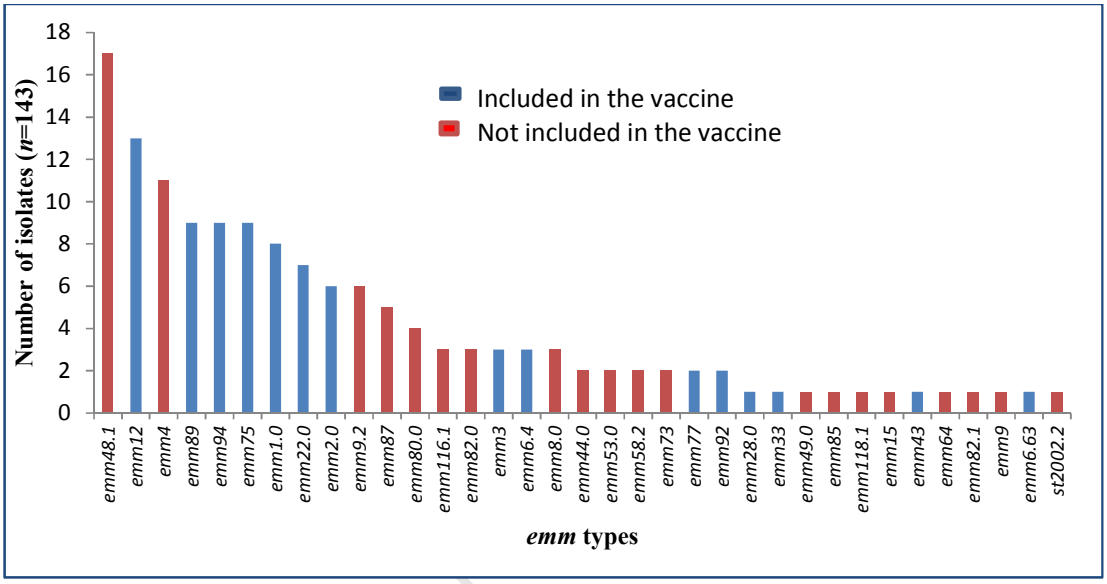


Figure 2-4. The diversity and distribution of *emm* types. The red bar indicates *emm* types not included in the 26-valent vaccine.

Table 2-2. The *emm* sequences (in descending order) recovered from symptomatic and asymptomatic children in Cape Town, South Africa

emm subtype/subtypes	*n (symptomatic/asymptomatic)*	%	Included in the 26-valent vaccine*	Included in the 30-valent vaccine*
emm48.1	17 (16/1)	12	No	No
emm12	13 (9/4)	9	Yes	Yes
emm4	11 (9/2)	8	No	Yes
emm89	9 (7/2)	6	Yes	Yes
emm94	9 (9/0)	6	Yes	No
emm75	9 (6/3)	6	Yes	Yes
emm1.0	8 (6/2)	5.6	Yes	Yes
emm22.0	7 (7/0)	5	Yes	Yes
emm2.0	6 (4/2)	4	Yes	Yes
emm9.2	6 (5/1)	4	No	No
emm87	5 (3/2)	3	No	Yes
emm80.0	4 (4/0)	2.7	No	No
emm116.1	3 (2/1)	2	No	No
emm82.0	3 (3/0)	2	No	Yes
emm3	3 (3/0)	2	Yes	Yes
emm6.4	3 (3/0)	2	Yes	Yes
emm8.0	3 (3/0)	2	No	No
emm44.0	2 (2/0)	1.3	No	Yes
emm53.0	2 (2/0)	1.3	No	No
emm58.2	2 (1/1)	1.3	No	Yes
emm73	2 (1/1)	1.3	No	Yes
emm77	2 (2/0)	1.3	Yes	Yes
emm92	2 (2/0)	1.3	Yes	Yes
emm28.0	1 (1/0)	0.7	Yes	Yes
emm33	1 (1/0)	0.7	Yes	No
emm49.0	1 (0/1)	0.7	No	Yes
emm85	1 (1/0)	0.7	No	No
emm118.1	1 (0/1)	0.7	No	Yes
emm15	1 (1/0)	0.7	No	No
emm43	1 (1/0)	0.7	Yes	No
emm64	1 (1/0)	0.7	No	No
emm82.1	1 (1/0)	0.7	No	No
emm9	1 (1/0)	0.7	No	No
emm6.63	1 (0/1)	0.7	No	No
st2002.2	1 (1/0)	0.7	No	No
Total	**143**	**100%**	**14/35 (51.7%)**	**16/35 (62.9%)**

* Refer to Table 1-4 for vaccine contents.

2.4 Discussion

This study reports on the *emm* types obtained from 143 pharyngeal isolates. A total of 35 *emm* types were identified with eight *emm* types; *emm48, emm12, emm4, emm89, emm22, emm94, emm1* and *emm75*, comprising 58% of the isolates. Of the *emm* types labelled as rheumatogenic; due to their association with ARF, only *emm1, emm3,* and *emm6* were identified in the isolates from Cape Town. Each of *emm3* and *emm6* were isolated from two patients while *emm1* was the seventh most prevalent *emm* type (Table 2-2). Interestingly, previously reported rheumatogenic *emm* types, *emm5, emm18, emm14, emm19, emm24, emm27* and *emm29*, not identified in this study are thought to be disappearing [85].

Previous studies from Africa have reported greater *emm* diversity than was observed in GAS from Cape Town. One study from Ethiopia identified 43 *emm* types in 82 GAS isolates [46] while 38 *emm* types were obtained from 103 GAS isolates from Tunisia [48] compared to 35 *emm* types from 143 Cape Town GAS. While the *emm* diversity in Cape Town may be lower than observed in the other two African countries, it is higher than the corresponding diversity observed in GAS from North America, Canada and Europe. Only 56 and 33 *emm* types were identified in 7040 US and 1434 Canadian GAS, respectively [67]. Studies from Europe show a similar diversity to that observed in North America. An Italian study that included 207 GAS strains obtained 32 *emm* types [166], while 25 *emm* types were reported from 140 Austrian GAS isolates [57]. Similarly, 19 *emm* types were identified from 101 and 147 GAS isolates from both Norway and Romania respectively [61, 62].

Among the isolates in this study, the *emm* types were not evenly distributed as some were common than others. The distribution of *emm* types was not even amongst Cape Town isolates. Eight *emm* types accounted for 58% of the isolates with *emm48* being the dominant type, comprising 12% of the GAS (Table 2-2). In Tunisia, eight *emm* types accounted for 66% of the isolates with *emm118* being dominant (9.7%). This *emm* type was not identified in the Cape Town GAS; conversely, the dominant local type (*emm48*) was identified in only one isolate from Tunisia [48]. The distribution of *emm* types in Ethiopia was broader than observed in both Tunisia and Cape Town. In Ethiopia, the eight most common *emm* types accounted for 38% of

the GAS with the dominant *emm3.19* contributing 15% of the isolates. This *emm* type was not identified in isolates from both Cape Town and Tunisia (Table 2-2).
A comparison of the five most frequently identified *emm* types, *emm48, emm12, emm4, emm89, emm94*, in Cape Town with the corresponding global information is presented in Table 2-3.

Table 2-3. Five prevalent *emm* types from Cape Town and other parts of the globe

Country	Sampled isolates	Number of *emm* types	Prevalent *emm* types (in descending order of prevalence)	Reference
Cape Town	143	35	48, 12, 4, 89, 94	This study
UAE	38	25	st3211, 89, st0721, 78.3, st4695	[56]
Lebanon	103	33	1, 22, 28, 88, 4	[55]
Germany	216	17	1, 12, 3, 28,4	[58]
Sweden	1 519	æ	89, 81, 28, , 12, 77	[171]
USA	7 040	56	1, 12, 28, 4, 3	[67]
Canada	1 434	33	12, 1, 28, 4, 3,	[67]
Brazil	128	48	53, 22, 49, 58, 83	[69]

UAE-*United Arab Emirates*
USA-*United States of America*
æ – not indicated

While there is some overlap between *emm* types identified in Cape Town and in other parts of the world as shown in Table 2.3, there are marked differences between the prevalence of individual types across the countries. For instance, *emm48* was the most prevalent in Cape Town GAS but was not identified in the USA, Canada and Germany [58, 67]. On the other hand, *emm1* which was prevalent in the USA, Canada and Germany [58, 67], comprised only 6% of the Cape Town GAS. Data collected through the Strep-EURO program revealed ten prevalent *emm* types in Europe; *emm1, emm28, emm3, emm89, emm87, emm12, emm4, emm83, emm81* and *emm5* [64]. Seven of these isolates; *emm1, emm28, emm3, emm89, emm87, emm12* and *emm4*, comprising 38.4% of the GAS, were identified in Cape Town (Table 2.2).

In a study from Lebanon involving 103 GAS, *emm22* and *emm4* accounted for 8.73% and 6.79% of the isolates, respectively [55]. These two *emm* types featured in Cape Town, accounting for a combined 13% of the GAS (Table 2-2). A study by Alfaresi in the UAE revealed differences between *emm* types from that country and those identified in Cape Town, as well as the USA and Europe (Table 2-3). Among the five most prevalent *emm* types, only *emm89* identified by Alfaresi was observed in

Cape Town (Table 2-3). A study of 34 GAS from India identified 22 *emm* types of which only one of the Cape Town isolates (*emm49*, 0.7%) was identified [53].

The efficacy of the experimental 26-valent GAS vaccine was considered in the context of the Cape Town isolates and other regions. Of the *emm* types identified in Cape Town, 14 of the 35 *emm* types are contained in the vaccine, accounting for an estimated coverage of 51.7% of the study population. In Ethiopia, 46% of the *emm* types were represented in the 26-valent vaccine [46]. On the other hand, coverage of the 26-valent vaccine has been estimated at 85% in the USA [142] and 69% in Europe [64]. An extensive review of *emm* type studies and reports from different parts of the world [11], estimated vaccine coverage of 43.7% from pharyngeal isolates in Africa and an overall coverage of 39% from all GAS diseases. Apart from China which has reported an overwhelming 93.5% vaccine coverage of their pharyngeal isolates [49], coverage of the 26-valent vaccine is lower in developing countries than in developed countries [46, 49, 56].

Recent development of a 30-valent vaccine is expected to have broader efficacy due to opsonisation with M-proteins which were not included in the previous 26-valent vaccine [140]. Eleven *emm* types, (*emm4, emm29, emm73, emm58, emm44, emm78, emm118, emm82, emm83, emm87* and *emm49*) not included in the former 26-valent vaccine are featured in the new 30-valent vaccine [140]. Five of these additions (*emm4, emm44, emm82, emm87* and *emm49*) were identified in Cape Town (Table 2-2). The inclusion of the *emm* types in the novel 30-valent vaccine was based on epidemiological data from North America, CDC and a StrepEuro study. Of the types present in the initial 26-valent vaccine but excluded from the novel 30-valent vaccine, *emm33 and emm43* were identified in this study. Although the included types do not alter the vaccine coverage in Cape Town significantly, 51.7% in the 26-valent vaccine compared to an estimated 63% in the 30-valent vaccine, the novel vaccine evokes cross-opsonic bactericidal antibodies against non-vaccine serotypes, [172] [140]. However, during writing of this thesis, the cross-opsonic bactericidal antibodies observed was still under investigation [172] The cross-opsonic antibodies evoked by the 30-valent vaccine against GAS will be welcomed in South Africa and the developing countries.

This data on the prevalent *emm* types from Cape Town provide a snapshot of the distribution of *emm* types circulating among Cape Town children. The data enhances our understanding of circulating *emm* types essential for consideration of a vaccine that will include *emm* types of epidemiological importance to the South African population.

Characterisation of the *emm* gene using restriction fragment length polymorphism

3.1 Introduction

Since its introduction in 1996, *emm* typing has been the method of choice to study the molecular epidemiology of GAS [9, 10]. However, this sequence based technique is impractical in some developing countries where sequencing facilities are scarce or costly. An alternative approach, restriction fragment length polymorphism (RFLP), has been used in the characterisation of the peptidyltranferase region of the 23S rRNA from *Helicobacter pylori*, the *flagellin A* gene from *Campylobacter jejuni*, the *omp1* gene from *Chlamydia trachomatis* and the *emm* gene from GAS [60, 71, 80, 81, 173-177]

PCR-RFLP is a technique in which nucleotide polymorphisms in a DNA sequence are used as markers to identify differences between specific genes from different organisms. Restriction endonucleases used in PCR-RFLP are enzymes which cleave DNA at a known sequence, usually 4-8 nucleotides long. Thus nucleotide polymorphisms at restriction sites will result in an enzyme being unable to cleave the DNA. Variations in restriction profiles are then used to detect the polymorphisms and identification is based on an individual profile in relation to those of others. Single enzyme digestion using the enzymes *Mbo*I, *Dde*I or *Sac*I, and double digestion using *Hae*III and *Hinc*II, have been recommended by the CDC for characterisation of *emm* types.

Although RFLP has proven useful in the characterisation of the M-protein, on its own, RFLP is not always able to elucidate the M-protein type and is frequently coupled with *emm* typing [66, 80]. However, better and well documented banding patterns associated with individual *emm* types could facilitate identification of *emm* types using RFLP. This chapter describes the use of RFLP to characterise local M-protein as an alternative to *emm* typing.

3.2 Experimental protocol

3.2.1 Sample selection

A total of 63 of the 143 isolates described in Chapter Two were included in this part of the study. These 63 isolates were chosen to provide at least two to three representatives of each of the 23 commonest *emm*-types (based on the results in Chapter 2). For reasons of practicality, at maximum of three isolates of a given *emm*-type were chosen. By analysing at least two representative isolates of each of these emm-types, the reproducibility of the RFLP assay could be evaluated. RFLP analysis of, emm-types that only occurred in one isolate was not performed.

3.2.2 Prediction of restriction product sizes

Using the New England Biolabs (NEBcutter V2.0) website [178] and DNA sequencing data from the GenBank and some DNA sequences provide by Pierre Smeesters from Université Libre de Bruxelles Belgium, restriction enzyme maps were generated for all the twenty three *emm*-types included in this section of the study. The primer binding sites were localised using BioEdit v7.0.9 (Ibis Biosciences, USA) and excess nucleotides from either flanking region of the five prime and three prime were trimmed to mimic the PCR product sizes that would be expected on agarose gel electrophoresis. As recommended by the CDC, *Dde*I, *Mbo*I and *Sac*I were chosen for single digests and *Hae*III and *Hinc*II for double digests.

3.2.3 Restriction enzyme digestion

Between 100-200ng of the each of the *emm* amplicons (2.2.3-2.2.6) estimated using PCR amplicon intensity (Appendix B) following agarose gel electrophoresis were digested with 5U of *Dde*I, *Mbo*I or *Sac*I, or *Hinc*II and *Hae*III in the recommended buffers (NEB, USA) and 1X bovine serum albumin (BSA) (NEB, USA). Reactions were made up to a final volume of 20 µl in molecular grade water (Promega, USA) and incubated in a waterbath at 37ºC for 60 minutes.

Restriction products were separated on 3% agarose gel (SeaKem® LE Agarose, Lanza, USA) containing EtBr (0.5µg/ml) for 6 hours at 60V (2.2.5) and photographed.

3.3 Results

3.3.1 Prediction of restriction maps by the NEBcutter

The NEBcutter V2.0 was able to excise DNA sequences where restriction sites were available. The program also provided possible sizes of restriction products as it would be expected on agarose gel electrophoresis (Figure 3-1). The nNEBcutter results were in conjuction with PCR products observed in agarose gel electrophoresis.

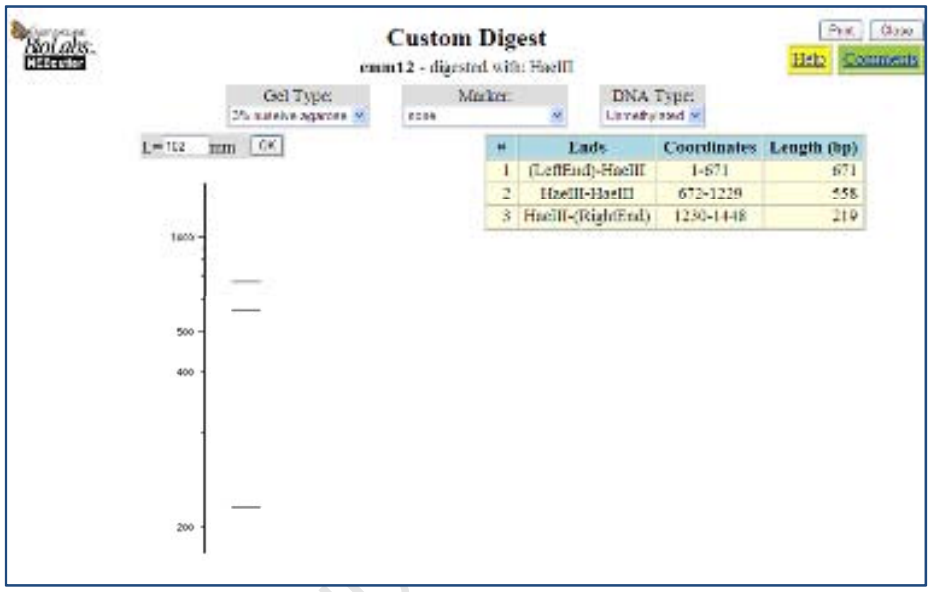


Figure 3-1. NEBcutter digest. Graphical simulation of anticipated restriction map of *emm12* sequence obtained following *Hae*III and *Hinc*II digest using NEBcutter V2.0.

3.3.2 Restriction digestion using *Dde*I

Restriction digestion with *Dde*I produced three to five restriction bands ranging between 50 and 500-bp (Table 3-1). At least two and where possible three isolates of the same *emm*-type were included to assess reproducibility. RFLP performed on different isolates of the same *emm*-type produced consistently reproducible profiles (Figure 3-2 lane 2 and lane 9). Restriction fragments less than 50 bp could not be visualised on the agarose gels and hence were excluded from Table 3-1. No isolates refractory to digestion were present among the twenty three *emm* types digested with *Dde*I (Refer to Appendix C for all the agarose gel electrophoresis results).

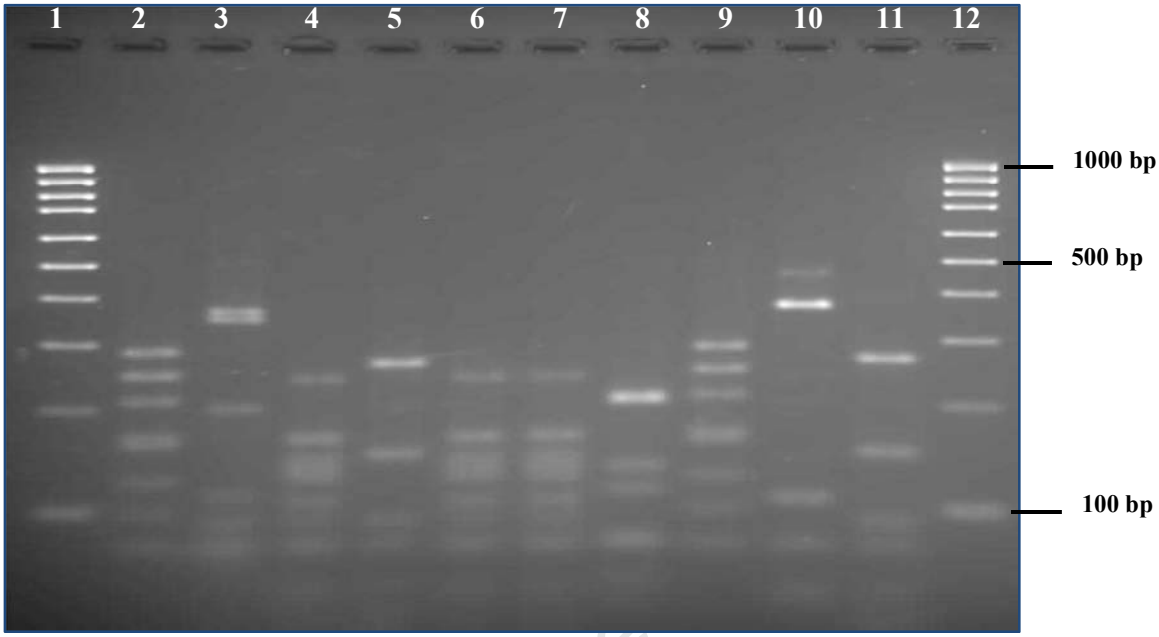


Figure 3-2. *Dde*I digests. A 3% agarose gel electrophoresis depicting restriction pattern of 10 isolates. Lane 2 *emm12*, Lane 3 *emm28*, Lane 4 *emm94*, Lane 5 *emm89*, Lane 6 *emm94*, Lane 7 *emm94*, Lane 8 *emm6.4*, Lane 9 *emm12*, Lane 10 *emm4* , Lane 11 *emm89*. Lane 1 and 12 are molecular markers (Appendix C)

Table 3-1. Restriction profiles obtained following digestion of twenty seven different *emm* types using *Dde*I

emm types	*n=63*	Restriction products (base pairs)						
emm1	3	270	200	180	150	70		
emm2	3	280	150	90	80			
emm3	3	280	200	150	100			
emm4	3	370	100	80	70			
emm6.4	3	200	190	130	100	80		
emm8	3	280	210	90	80			
emm9.2	3	260	170	90	60			
emm12	3	280	240	220	170	120	90	70
emm22	3	340	150	90	60			
emm44	2	160	150	120				
emm48.1	3	340	110 [D]	90	80			
emm53	2	250	140	120 [D]	80			
emm58.2	2	340	160	90	80			
emm73	2	300	150	90	80			
emm75	3	260	100	90 [D]	80			
emm77	2	240	150	90	80			
emm80	3	320	190	100	70			
emm82	3	250	150	100	80			
emm87	3	340	170	90	70			
emm89	3	270	150	90	80			
emm92	2	490	90	80	50			
emm94	3	250	170	150	100	90		
emm116	3	170 [D]	90 [D]					

[D] doublets

3.3.3 Restriction digestion using *Mbo*I

Two to four restriction bands of 90 to 760 bp were observed following digestion with *Mbo*I. Four *emm* types were refractory to *Mbo*I digestion (Table 3-2). For *emm3, emm4* and *emm8*, the respective bands less than 50 base pairs could not be determined following agarose gel electrophoresis (Table 3-2).

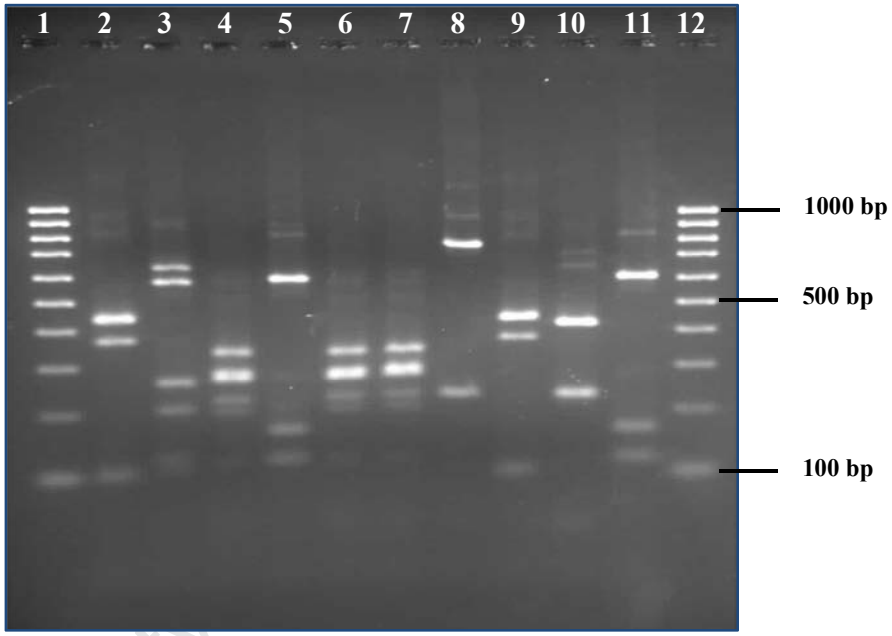


Figure 3-3. *Mbo*I digests. Agarose gel electrophoresis depicting 10 GAS isolates. Lane 2 *emm12*, Lane 3 *emm28*, Lane 4 *emm94*, Lane 5 *emm89*, Lane 6 *emm94*, Lane 7 *emm94*, Lane 8 *emm6.4*, Lane 9 *emm12*, Lane 10 *emm4* , Lane 11 *emm89*. Lane 1 and 12 are molecular markers.

Table 3-2. Restriction profiles obtained following restriction digest of twenty eight different *emm* types using *Mbo*I.

emm type	*n=63*	Restriction products (base pairs)			
emm1	3	520	480	120	
emm2	3	600	220	60	
emm3	3	700	200		
emm4	3	680	210		
emm6.4	3	760	220		
emm8	3	610	200	100	
emm9.2	3	480	340		
emm12	3	440	380	100	
emm22	3	500	340	100	
emm44	2	920*			
emm48.1	3	320	280	240	110
emm53	2	630	360		
emm58.2	2	990*			
emm73	2	480	270		
emm75	3	350	300	250	
emm77	2	480	180	60	
emm80	3	620	300	50	
emm82	3	960*			
emm87	3	950*			
emm89	3	610	160	120	
emm92	2	420	340	200	
emm94	3	350	290	270	
emm116	3	560	260	100	50

*No restriction site

3.3.4 Restriction digestion using *Sac*I

Of the twenty three *emm* types subjected to restriction digestion, only seven could be digested with *Sac*I

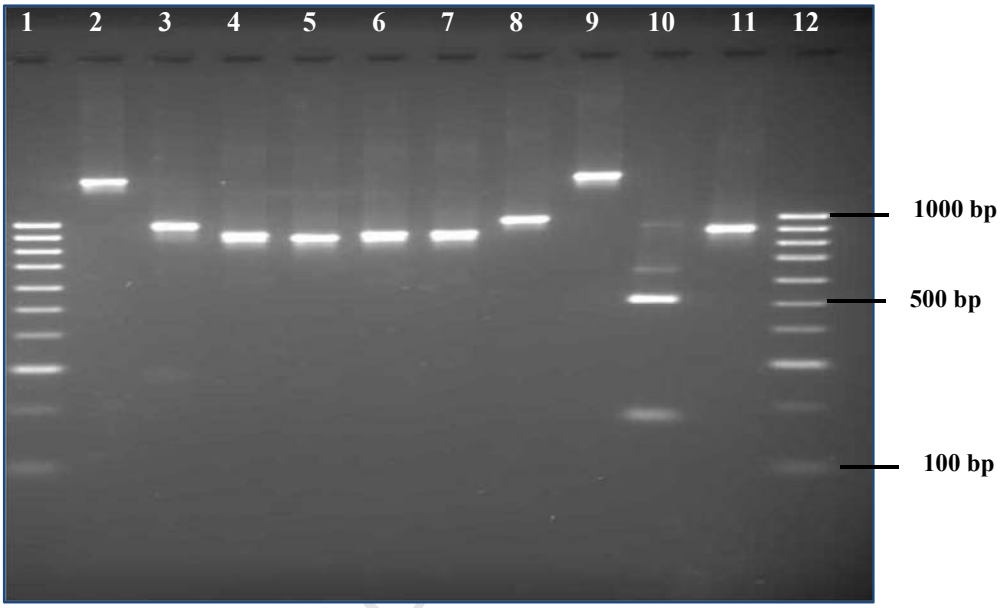


Figure 3-4. *Sac*I digests. Agarose gel electrophoresis depicting 10 GAS isolates. Lane 2 *emm12*, Lane 3 *emm28*, Lane 4 *emm94*, Lane 5 *emm89*, Lane 6 *emm94*, Lane 7 *emm94*, Lane 8 *emm6.4*, Lane 9 *emm12*, Lane 10 *emm4* , Lane 11 *emm89.* Lane 1 and 12 are molecular markers.

Table 3-3. Restriction profiles obtained following digestion of different *emm* types using *Sac*I

emm types	***n=63***	Restriction products (base pairs)	
emm1	3	1140*	
emm2	3	880*	
emm3	3	1240*	
emm4	3	800	150
emm6.4	3	1000*	
emm8	3	950	
emm9.2	3	900*	
emm12	3	1500*	
emm22	3	700	170
emm44	2	650	260
emm48.1	3	850	90
emm53	2	870	100
emm58.2	2	800	360
emm73	2	810*	
emm75	3	900*	
emm77	2	750*	
emm80	3	880	100
emm82	3	950*	
emm87	3	850*	
emm89	3	900*	
emm92	2	950*	
emm94	3	900*	
emm116	3	1000*	

* No restriction site

3.3.5 Restriction digestion using *Hae*III and *Hinc*II.

Double digestion with *Hae*III and *Hinc*II produced two to five restriction bands between 60 to 830 bp (Table 3-4). Two *emm* types were refractory to double digest (Table 3-4).

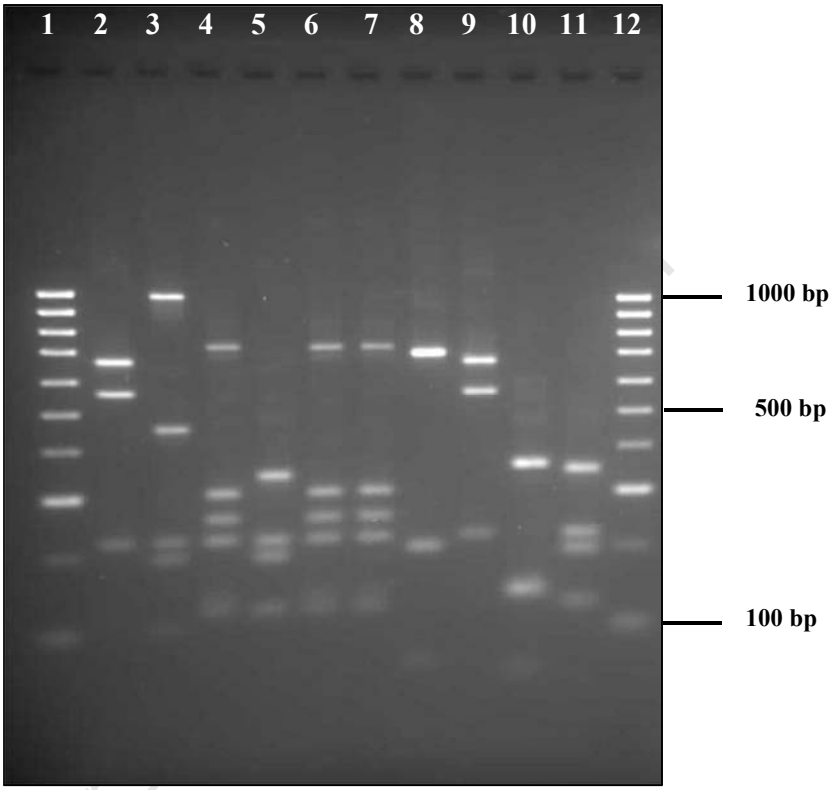


Figure 3-5. *Hae*III and *Hinc*II digests. Agarose gel electrophoresis of 10 GAS isolates. Lane 2 *emm12*, Lane 3 *emm28*, Lane 4 *emm94*, Lane 5 *emm89*, Lane 6 *emm94*, Lane 7 *emm94*, Lane 8 *emm6.4*, Lane 9 *emm12*, Lane 10 *emm4* , Lane 11 *emm89*. Lane 1 and 12 are molecular markers.

Table 3-4. Restriction profiles obtained following digestion of different *emm* types using *Hae*III and *Hinc*II

emm types	***n=63***	Restriction site	Restriction products (base pairs)				
emm1	3	*	830	220	130		-
emm2	3	**	520	350			-
emm3	3	*	340	300	180	90	-
emm4	3	***	750	120			-
emm6.4	3	**	720	220	60		-
emm8	3	***	470	320	220		-
emm9.2	3	§	900				-
emm12	3	*	680	560	220		-
emm22	3	***	460	220 [D]			-
emm44	2	***	340	240	230	130	-
emm48.1	3	§	800				-
emm53	2	*	740	220			
emm58.2	2	***	460	210			-
emm73	2		420	390			
emm75	3	*	700	220			-
emm77	3	*	550	240			-
emm80	3	*	550	240	220		
emm82	3	*	570	280	100		-
emm87	3	***	350	280	220	130	-
emm89	3	***	350	220	200	130	-
emm92	2	***	440	320	200		-
emm94	3	*	320	280	230	130	-
emm116	3	*	740	200			-

* Restriction site for *Hae*III
** Restriction site for *Hinc*II
*** Restriction site for both *Hae*III and *Hinc*II
§ No restriction sites for both *Hae*III and *Hinc*II
[D] doublets

3.4 Discussion

The twenty three *emm* types subjected to RFLP could be differentiated by restriction profiles (Table 3-1, 3-2 and 3.3; Figure 3-2, 3-3, 3-4 and 3-5). Three restriction enzymes, *Dde*I, *Mbo*I or *Sac*I, were used for single digestion and a combination of *Hae*III and *Hinc*II for double digests. None of the *emm* types in this study were refractory to digestion with *Dde*I, and all the profiles generated by *Dde*I could be distinguished from each other. These results correlate with those seen in another study which characterised GAS using *Dde*I [66]. During a scarlet fever outbreak in Mexico, RFLP using *Dde*I identified five *emm* types in 40 GAS isolates which correlated with sequencing results [66]. Sequencing confirmed the *emm* types as *emm2, emm12, emm1, emm22* and *emm89* [66].

Nineteen different *emm* types digested with *Mbo*I produced unique and readily distinguishable restriction profiles while four *emm* types could not digested (Table 3-2). Interestingly only seven *emm* types contained restriction sites recognised by *Sac*I (Table 3-3). However, for those *emm* types which were readily digested by *Sac*I, the profiles were unique enough to deduce the differences. To the best of the author's knowledge, there are no other studies reporting on either *Mbo*I or *Sac*I digests.

Following double digests with *Hae*III and *Hinc*II (Table 3-4), two *emm*-types were not digested by either *Hae*III and *Hinc*II (Table 3-4). However, in the remaining twenty one *emm*-types, the restriction profiles were discriminatory enough to identify the different *emm* types (Table 3-4), consistent with previous studies [66, 80, 83]. The study by Perea-Mejia during the scarlet fever outbreak in Mexico used *Hae*III and *Hinc*II to segregate forty isolates into five *emm* types [66] while a study by Beall [80] observed restriction profiles that were able to distinguish 32 *emm* types from 945 geographically diverse clinical isolates [80]. The 32 restriction profiles represented different *emm* types that had previously been identified using sequencing (*emm* typing) [80].

Restriction fragment length polymorphism has been useful in differentiating and grouping similar *emm* types [66, 80]. Representatives of the *emm* types in this study could be segregated based on their respective restriction profiles. These observations have also been seen in other studies which used RFLP to identify similar *emm* types. As previously mentioned, the study by Perea-Mejia was able to group 40 isolates into five groups based on restriction profile produced following digestion with *Dde*I [66]. Similarly, RFLP was useful in another study investigating 106 GAS isolates [71]. Although a different enzyme (*Alu*I) from those recommended by the CDC was utilised in the later study, nineteen different profiles distinct to different *emm* types were identified [71]. More impressively is the study by Beall which identified 32 *emm* types from 945 geographically diverse clinical isolates by RFLP [80], illustrating the ability of this relatively simple technique to distinguish the M-proteins. Hence RFLP can be used to differentiating *emm* types, particularly when a large number of isolates are involved. This minimizes the cost of further investigations such as sequencing.

The sequencing part of *emm* typing assay is the major limiting factor in developing countries. The *emm* typing assay requires generation of sequences in order to differentiate the N-terminal of the M-protein [168], most of the developing countries lack resources or sequencing facilities to carry out this most crucial part of the technique. Sequencing machines and associated reagents are costly thus posing a challenge to laboratories lacking resources. Other molecular typing techniques such as MLST and PFGE are laborious and require intense data analysis [72]. Similar to *emm* typing, MLST is a sequencing based technique. While PFGE lacks a sequencing step [72] other the limitations associated with PFGE includes standardisation, turn-around time of four to five days [179] and inter laboratory variation in results.

In this study, twenty three *emm* types subjected to RFLP were readily digested by *Mbo*I, *Dde*I, *Hae*III and *Hinc*II producing distinct restriction profiles. Although the *emm* typing techniques remains the gold standard in molecular epidemiology of GAS, RFLP can be an alternative assay in identification of *emm* types with similar

restriction patterns. It is evident that RFLP is ideal in segregating different *emm* types based on distinct profile. The restriction product sizes produced in this study can be useful in identifying local *emm* types, although studies using a greater range of local *emm* types may be needed to fully validate this approach.

C-repeats of the M-protein contain potential vaccine candidates conserved among GAS strains

4.1 Introduction

As elaborated in Chapter Two (2.1), the M-protein consists of regions A to D (Figure 2.1). The C region is conserved across GAS strains and contains potential vaccine candidates which may be an alternative to the M-protein specific vaccines [8]. The C-repeats of the M-protein, in which the vaccine candidates feature, are conserved sequences located on the C region.

Studies investigating the conserved region of the M-protein preferably utilise amino acid sequence analysis in place of nucleotides. This is due to the fact that amino acid sequences are more conserved than nucleotide sequences as a result of genetic code degeneracy. Of the sixty-one codons that make up the genetic code, there are only twenty amino acids (Appendix D). Therefore, two different codons may code for the same amino acid e.g., valine is coded by the codons GTT, GTC, GTA or GTG. Hence, protein sequences may remain conserved even though their respective nucleotide sequences may harbour variations. Different bioinformatics tools are used to align the protein sequences in order to identify the conserved regions [180, 181]. The M-protein contains at least three conserved sequences, C1, C2 and C3 [145, 180, 181]. These repeats are identified following multiple alignment of the protein sequence.

A region of one of the C-repeat sequences, J14, has been used in the construction of a vaccine target that is conserved among GAS strains. Three other J14 like sequences, namely; J14.1 J14.2 and J14.8, have also been identified as potential vaccine candidates [181]. The full amino acid sequence of J14 in the vaccine is KQAEDKVK**ASREAKKQVEKALE**QLEDKVQC [182]. The amino acids in bold are M-protein specific while the flanking sequences, underlined, are derived from yeast (GCN4) and assist in achieving the helical folding and conformational structure of

the M-protein [182]. Recently, antibodies against the J8 region have been raised in mice, and provides strong potential as a vaccine candidate given the cross-protective immunity against GAS isolates containing J14.0 and J14.[145]1.

Geographical differences in the distribution of the *emm* types and lower coverage in developed countries shown by M-protein vaccines targeting the hypervariable region (N-terminus, block A) have encouraged a focus on the conserved region as a potential vaccine target. The geographical differences among dominant *emm* types from different continents, as presented in Chapter One (Table 1-1), result in the type specific M-protein based vaccine having less coverage in developing countries as discussed in 2.4. The J14 vaccine is efficacious and does not cross-react with host tissues, as elaborated in Chapter One (1.6.2) [145, 181]. This part of the study aims to describe the sequences of the C-repeats of the M-protein as potential J14 vaccine targets that are shared across GAS strains.

4.2 Experimental protocol

4.2.1 Study population, identification of β-haemolytic streptococci and amplification of the *emm* gene

Sixty three isolates (used in Chapter Three) representing twenty three *emm* types were included for analysis of the conserved region of the M-protein. Where available, only *emm* types represented by more than a single isolate identified in Chapter Two (Table 2-2) were included for C repeat analysis. The DNA isolation and amplification were carried out as elaborated in Chapter Two (2.2.1-2.2.6).

4.2.2 Sequencing the C-terminus of the *emm* gene

Following purification of the amplified *emm* gene (2.2.6), DNA was quantified using the NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). In order to sequence the C-repeat regions of the M-protein, primer 2 [10], the reverse primer utilised in the amplification of the *emm* gene [10], is used. Primer 2 (Table 2-1) was prepared to 1.1 ng/µl and used in the sequencing reaction. The amplicons were analysed and edited using BioEdit v7.0.9 (Ibis Biosciences, USA).

4.2.3 Translation of M-protein sequences into amino acid

Since the reverse primer (primer 2) runs from the 3′ to 5′ direction and was utilized to generate the sequences of the C-repeat region, the reverse complement sequence was generated and then translated into the amino acid sequence using BioEdit v7.0.9 (Ibis Biosciences, USA).

4.2.4 Identification of the C-repeats within the M-protein sequences

The Multiple Em Motif Elicitation (MEME) online program aligns nucleotides or amino acid sequences in order to identify conserved regions within the sequences [183]. Depending on the number of conserved regions, specified by the investigator, the MEME program generates sequence alignments of each of the conserved region followed by generation of a sequence LOGO [183]. A sequence LOGO generated by MEME is the position-specific probability of a nucleotide or amino acid appearing in a predicted position within a specific conserved region.

The MEME online program was specified to identify three C-repeats in all the twenty three *emm* types, based on previous studies' reports [143, 145]. Sequences in FASTA format (4.2.3), were uploaded onto the MEME website and submitted for analysis, (http://meme.sdsc.ed/meme/cgi-bin/meme.cgi 16 February 2012).

4.3 Results

4.3.1 Multiple sequence alignment

Multiple sequence alignment of the conserved region was generated using MEME sequence alignment. Three different sequence alignments representing three C-repeats, C1, C2 and C3, were generated (Appendix D). Only sequences containing a respective C-repeat appeared in the alignment. The sequences were arranged based on their *p*-value. The *p*-value of the sequence alignment indicates the degree of motif match and not the degree of conservation of the sequences (Refer to Appendix D for sequence alignment results).

4.3.2 Identification of conserved sequences in C region of the M-protein

The MEME online program identified three C-repeat sequences in sixteen *emm* types, while seven *emm* types had only two C-repeats (Table 4-1). Amino acid alignments depict sequence variations within individual conserved regions (Appendix D). The first conserved region, C1, showed more sequence variation than the second and third conserved regions. The region C2 and C3, consisted mainly of the J14 and J14.1 variants, respectively. All the *emm* types included in this study contained either the J14 or J14-like sequence.

4.3.3 The MEME sequence LOGO

A MEME sequence LOGO was generated based on the sequences submitted to the MEME online program. Three different LOGO were generated, each for the three conserved regions, C1, C2 and C3 (Figure 4-1). Each LOGO contained sequences of the specific conserved region. The height of the letters in the LOGO depends on sequence variation within individual conserved region (Table 4-1).

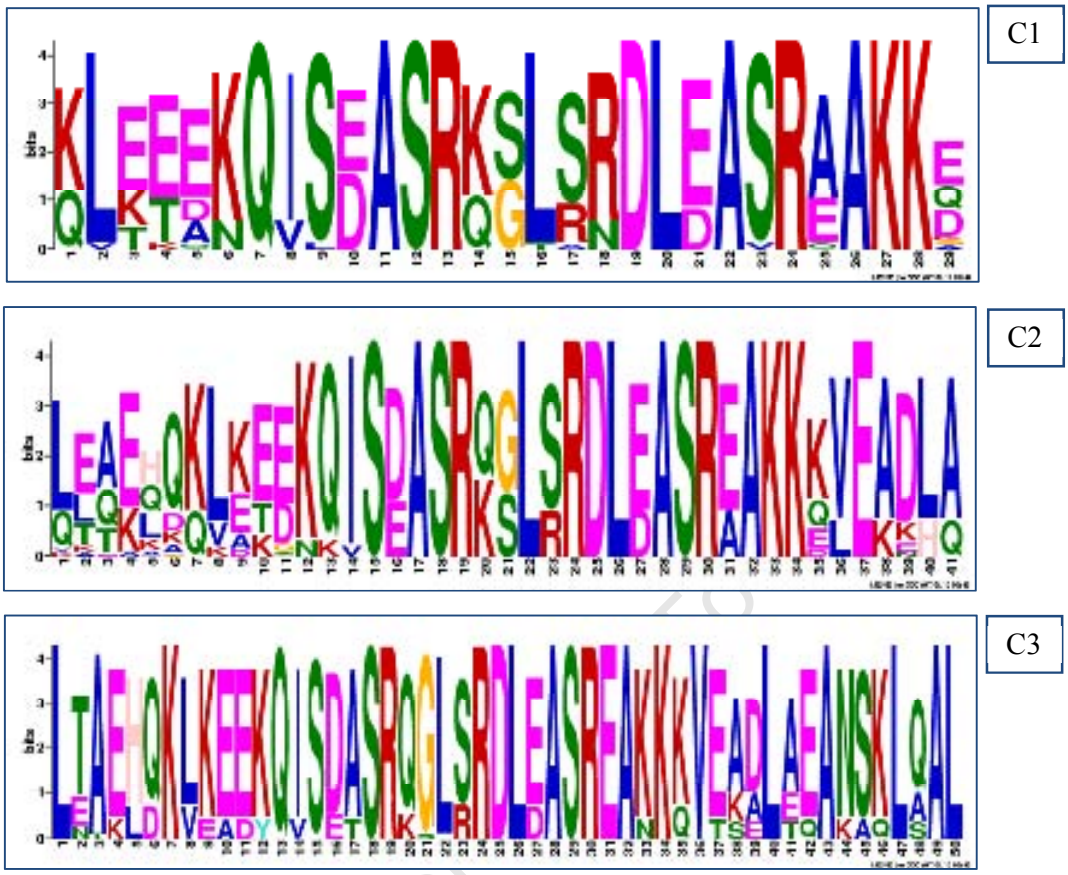


Figure 4-1. Sequence LOGO of three C-repeats. Each logo contains conserves sequences of the J14 or J14-like sequences. Each single letter amino acid height stands for a probability of that particular amino acid being located in the specified position.

Table 4-1. C-repeats sequences

emm type	*n=63*	C1-repeat	C2-repeat	C3-repeat
emm1	3	J14.2	J14.2	J14
emm2	3	J14.29	J14.12	J14.1
emm3	3		J14.2	J14
emm4	3	J14.36	J14.12	J14.1
emm6.4	3		J14.2	J14
emm8	3	J14.8	J14.1	J14.1
emm9.2	3		J14.1	J14.1
emm12	3	J14.8	J14.2	J14
emm22	3		J14.41	J14.1
emm44	2	J14.8	J14.1	J14.1
emm48.1	3	J14.36	J14.1	J14.1
emm53	2	J14.2	J14.2	J14
emm58	2	J14.29	J14.1	J14.1
emm73	2		J14.29	J14.1
emm75	3	J14.36	J14.1	J14.1
emm77	2		J14.8	J14.1
emm80	3	J14.2	J14.2	J14
emm82	3	J14.72	J14.1	J14.1
emm87	3		J14.1	J14.1
emm89	3	J14.29	J14.1	J14.1
emm92	2	J14.29	J14.1	J14.1
emm94	3	J14.12	J14.1	J14.1
emm116.	3	J14.2	J14.1	J14.1

J14 ASREAKKQVEKALE
J14.1 ASREAKKKVEADLA
J14.2, ASREAKKQVEKDLA
J14.8 ASRAAKKELEAEHQ
J14.12 ASRAAKKELEAKHQ
J14.29 ASRAAKKDLEAEHQ
J14.36 ASRAAKKELEANHQ
J14.41 ASRAAKKKVEADLA
J14.72 ASRAAKKGLEAEHQ

4.4 Discussion

The C-repeat region of the M-protein is highly conserved across GAS strains thus making it potentially an ideal vaccine target for multiple GAS strains. Sixty three GAS isolates, representing twenty three *emm* types were analysed to characterise the C-repeat regions. The C1-repeat of the M-protein contains more sequence variation compared to both C2 and C3. Among the twenty three *emm* types investigated in this study, there was sequence variation in the C1 region from different *emm* types (Table 4-1). The J14 sequence was not identified in region C1 of any of the sequences included in this study. These results observed from C1-repeats accords with the high level of sequence variation reported in the studies conducted by Vohra and Smeesters, respectively [145, 180].

Within the C2 region, thirteen different *emm* types contained J14.1. Other studies have also found a variety of sequences in the C2, however, J14.1 was the most common [145, 180]. Twenty-three out of thirty-one *emm* types investigated by Vohra corresponded to J14.1 while six had J14.4 sequences and one *emm105* isolate harboured J14.8 [145]. Similarly, J14.1 (30 out of 51 *emm* types) was dominant in C2 in a study by Smeesters, followed by J14.2 [180]. The C2 region reported by Smeesters showed higher sequence variation compare to this study and that of Vohra. However, it is important to note that the number of *emm* types investigated in this study (23) and those reported by Vohra (31) were less than those reported by Smeesters (51).

The C3 region, which is more proximal to the cell, harbours sequences corresponding to J14 and J14.1 **[145, 180].** Six *emm* types in this study harboured sequences corresponding to the J14 sequences while seventeen *emm* types contained the J14.1 sequence. The presence of either J14 or J14.1 is *emm* type specific. (Table 4-1). These results are similar to other studies that have investigated the conserved region of GAS [145, 180, 181]. Smeesters as well as Vohra have reported the presence of predominately J14 and J14.1 from the C3 region among the isolates investigated. In the study by Smeesters, eighteen *emm* types harboured sequences corresponding to J14, thirty *emm* types corresponded with the J14.1 sequence, two

emm types corresponded to J14-R11 and *emm78* contained a sequences different from J14 [180]. Similarly, the study by Vohra also found dominantly the presence of J14 or J14.1 in C3 being *emm* type specific [145]. Twenty-one out of thirty-one *emm* types investigated harboured sequences corresponding to J14.1 while the other ten corresponded to J14 [145]. Based on the results observed in this study (Table 4-1) and those reported by both Smeesters [180] and Vohra [145], respectively, C3 dominantly harbours sequences corresponding to J14.1 while a few *emm* types contain J14.

C3 is the most conserved region, relative to C2 and C1, with the degree of conservation increasing as one moves more proximally toward the cell. As it is evident from sequence alignment (Appendix D) and data presented on Table 4-1, the C3 region contains predominantly J14.1 whereas C2 has decreased sequence conservation, mainly J14.1 and other sequences (Table 4-1). Also, as previously elaborated, C1 displayed the most diverse sequences. This observation has also been seen by another study which identified C1 as the most diverse of the three C-repeats [181], similarly, all the other studies investigating the conserved region of the M-protein accords these similar observations [71, 145, 180]. Since C1 is less proximal to the cell compared to C2 and C3, one can suggest that the degree of sequence conservation of the M-protein decreases as one moves away from the cell. This is borne out by the fact that the N-terminus of the M-protein, which is known for its degree of hypervariety, is situated distally while the C-repeats are proximal to the cell,.

The four J14 sequences that are potential vaccine candidates are J14, J14.1, J14.2 and J14.8 [145, 180]. Sequences corresponding to these four vaccine candidates were identified in all the twenty three *emm* types investigated in this study (Table 4-1). The C3-region described previously harboured J14 and J14.1 while J14.2 was identified in the C2 region. The C3 region is highly conserved and harbours the J14 and J14.1 sequences [145], hence a vaccine designed to target the J14 or J14-like peptides can potentially induce protective antibodies that can protect against multiple GAS serotypes. This study and the studies conducted by Smeesters and Vohra,

respectively, have shown the presence of the J14 and J14.1 sequence among different *emm* types (Table 4-1) [145, 180]. The four vaccine candidates were identified among different *emm* types from different regions of the globe suggesting that protection by J14 or J14.1 based vaccines can be effective against multiple GAS strains from different regions of the globe.

The 26-valent type specific M-protein vaccines only offer protection against GAS strains included in the respective vaccine [46, 48], a property which makes these vaccines less favourable in some developing countries. With more than 220 *emm* types currently known and only 30 *emm* types included in the latest 30-valent vaccine [140], this raises concerns over the coverage of the type-specific vaccines in developing countries. However, a vaccine based on the conserved J14 is not M-protein specific due to the conserved sequences, observed in the C3-repeat (Table 4-1) [145, 180]. Hence this vaccine target can be an alternative to multiple GAS strains currently known, regardless of the prevalent *emm* type in a particular setting. Nevertheless, there are considerable concerns regarding their immunogenicity with consistent results not being observed between laboratories [184]

In summary, twenty three *emm* types analysed in this study harboured either the J14 or J14.1 sequences which are both potential vaccines candidates. Translating this into the vaccine perspective means that a vaccine designed to target the conserved region of the M-protein may be more ideal due to the level of conservation of this region across GAS strains. The J14 vaccines have shown promising results in protecting against GAS infections. Therefore, regarding J14 or J14.1, this study has shown a lack of M-protein specificity and geographical differences, making this vaccine candidate ideal in developing countries where high diversity of *emm* types exists.

Chapter Five: Streptococcal pyrogenic exotoxin

Streptococcal pyrogenic exotoxin gene profiles among Group A streptococci in Cape Town, South Africa

5.1 Introduction

Streptococcal superantigens (SAg) are prominent GAS virulent factors. As previously elaborated in Chapter One (1.1.3), eleven different streptococcal pyrogenic exotoxins (SPE); namely; SPEA, SPEC, SPEG, SPEH, SPEI, SPEJ, SPEK, SPEL, SPEM, streptococcal mitogenic exotoxin Z (SMEZ) and streptococcal superantigen A (SSA) have been described [27, 30, 34, 40, 185]. Considering the existence of the eleven SAg, detection of these genes can be laborious, time consuming and costly. Hence, different studies have opted for different molecular techniques for detection of these SAgs; conventional, multiplex or real-time PCR [27, 30, 42].

Conventional PCR has been widely used in the study of SAgs although this assay is laborious and time consuming especially when a different reaction is required for each gene [30, 32, 37, 54, 62]. However, most molecular SAg studies opt for detection of a single gene per reaction due to the number of parameters that would be needed to detect multiple genes [186, 187].

Multiplex PCR, devised in 1988 [188], detects more than two genes simultaneously, saving time, minimizing labour and is relatively cheaper than either conventional or real-time PCR [186, 187]. Multiplex PCR assays have been optimized to detect up to nine SAg genes (*speA, speB, speC, speF, speG, speH, speJ, smeZ* and *ssa*) simultaneously [42]. However, due to different conditions that may be required by individual primers, appropriate conditions that will accommodate all the primers need to be ascertained. A study by Lingtes investigating thirteen SAg genes opted for two sets of multiplex PCR for successful amplification [27]. The first set

contained *spea*1-3+5*, spec, speg, spej, spek* and *spel* while the second set consisted *spea*1-4, *speh, spei, spem, ssa* and *smez*. These two sets of genes were segregated due to the complexity of the multiplex PCR setup, in which primers pairs, reagent concentrations and PCR cycling conditions require common conditions to successfully amplify all the genes [186, 187]. Although multiplex PCR is more appropriate in detection of multiple genes simultaneously, the difficulty of this method lies in determining the optimal combination of reagents, primer concentrations and cycling conditions. However, successfully optimised multiplex PCR assays can save time and are relatively cheap.

Real-time PCR is the most sensitive method for detecting nucleic acid and hence may be ideal in SAg investigations. Real-time PCR, which amplifies and quantifies the target gene simultaneously, has also been applied in detection of SAg [27]. Interestingly, the multiplex real-time PCR was also able to identify a strain which did not harbour any SAg genes among the SAg investigated [27].

As is the case with other aspects of GAS in South Africa, there is lack of data on streptococcal superantigens. Hence the primary aim of this component of the study was to describe the distribution of SAg in association with *emm* types from patients with symptomatic pharyngitis as well as asymptomatic GAS carriers, using conventional multiplex PCR.

5.2 Experimental protocol

5.2.1 Study population and preparation of genomic DNA

A total of 98 GAS isolates representing 23 *emm* types were selected for investigation of SAg using conventional multiplex PCR. The same 23 emm-types that were analysed in Chapter 3 were chosen for this analysis. However, for this analysis we elected to analyse up to seven isolates (if possible) belonging to an individual emm-type in order to assess the degree of variation of the supernatigen profiles among different isolates belonging to the same emm-type. The selected samples included isolates representing more than a single isolate (Table 2-2). All the emm types identified represented by a single isolate were excluded for analysis (Table 2-2). Among the 98 isolates, 75 were obtained from symptomatic and 23 asymptomatic pharyngeal isolates (2.2.1). The genomic DNA investigated was obtained as described in 2.2.3. Among the 98 isolates, 75 were obtained from symptomatic and 23 asymptomatic pharyngeal isolates (2.2.1). The genomic DNA investigated was obtained as described in 2.2.3.

5.2.2 Single-plex PCR assay

PCR assays for detection of eight genes (*speA, speB, speC, speF, speG, speH, speJ* and *smeZ*) were performed as single reactions in order to ensure appropriate amplification of each gene. Each PCR reaction included a final concentration of 1X buffer, 2.5 mM $MgCl_2$, 0.4 mM dNTP, 0.2 µM of primer and 3.5U of SuperTherm *Taq* (JMR, Holdings, London UK) in a final volume of 50 µl as described by Schmitz [42], except for the primer concentrations. All primers were synthesized at the Department of Molecular and Cell Biology, University of Cape Town, South Africa. Respective primer sequences and expected product sizes are presented in Table 5-1.

The Applied Biosystems 2720 Thermocycler (Applied Biosystems, Carlbad, CA, USA) was used to amplify the SAg genes using PCR cycling conditions described by Schmitz [42]. Initial denaturation at 96ºC for a minute was followed by 30 cycles of denaturation at 96ºC for 50 seconds, annealing at 46ºC for 65 seconds and extension at 72ºC for 75 seconds. Final extension was carried out at 72ºC for 5 minutes.

Amplification products were analyzed using agarose gel electrophoresis in a 2% agarose gel (2.2.5) for 4 hours.

5.2.3 DNA sequence analysis of amplified products

Following PCR amplification of each of the eight genes, the amplicons were purified as previously described (2.2.6), quantified and submitted for DNA sequencing as elaborated previously (2.2.7), using the appropriately diluted forward primers (Table 5.1). The sequence data were analysed using BioEdit v7.0.9 (Ibis Biosciences, USA) and compared to existing sequences by BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi 28th November 2011).

Table 5-1. Primers sequences used for amplification of streptococcal superantigens

Target genes	Primers	Primer sequences [42]	Expected product sizes
speA	Forward	TAA GAA CCA AGA GAT GG	**248**
	Reverse	ATT CTT GAG CAG TTA CC	
speB	Forward	AAG AAG CAA AAG ATA GC	**955**
	Reverse	TGG TAG AAG TTA CGT CC	
speC	Forward	GAT TTC TAC TTA TTT CAC C	**584**
	Reverse	AAA TAT CTG ATC TAG TCC C	
speF	Forward	TAC TTG GAT CAA GAC G	**782**
	Reverse	GTA ATT AAT GGT GTA GCC	
speG	Forward	AGA AAC TTA TTT GCC C	**155**
	Reverse	TAG TAG CAA GGA AAA GG	
speH	Forward	AGA TTG GAT ATC ACA GG	**416**
	Reverse	CTA TTC TCT CGT TAT TGG	
speJ	Forward	ATC TTT CAT GGG TAC G	**535**
	Reverse	TTT CAT GTT TAT TGC C	
smeZ	Forward	TAA CTC CTG AAA AGA GGC T	**391**
	Reverse	CAT TGG TTC TTC TTG ATA AG	

Spe - Streptococcal pyrogenic exotoxin, SmeZ – Streptococcal mitogenic exotoxin Z

5.2.4 Multiplex PCR

During optimisation of the multiplex assays, a number of parameters were varied in order to determine the optimum conditions for a multiplex PCR assay.

5.2.4.1 $MgCl_2$

Three reagents utilised in PCR reaction, *Taq* polymerase, dNTPs and primers, require magnesium at an optimum concentration for optimal PCR fidelity [187]. Alteration in $MgCl_2$ concentration increases sensitivity and specificity of the reaction [189], hence, $MgCl_2$ was tested at a serial increments of 0.5 mM from 2 mM to 3.5 mM while maintaining the concentrations of other reagents (dNTP, *Taq* polymerase and primer concentrations).

5.2.4.2 dNTP

For optimum PCR reaction, a proper balance between $MgCl_2$ and dNTP is essential, as excessive dNTPs will inhibit the reaction [187]. The dNTPs work optimally at 0.2 mM when $MgCl_2$ concentration is between 1.5 and 2 mM [186, 187], thus, dNTP were tested at a serial increments of 0.1 mM from 0.2 to 0.5 mM. As the anticipated PCR product sizes ranged from 955 to 155 base pairs, a high dNTP concentration was required for complete elongation of double stranded DNA.

5.2.4.3 Primers

Primer concentrations and other PCR parameters such as melting and annealing temperatures play crucial roles during PCR [186, 187]. For optimum PCR amplification, depending on different loci, final primer concentrations of between 0.04 to 0.6 mM may be required for successful PCR amplification [186, 187]. In this study, all primers were tested at initial 0.2 mM and increased gradually by 0.1 mM up to a maximum of 0.7 mM if necessary.

5.2.5 Detection of streptococcal superantigens using optimised multiplex PCR assay

Molecular graded water (Promega, USA) was calculated to bring the final reaction volume to 50 µl following addition of 1X buffer, 2.5 mM $MgCl_2$, 0.4 mM dNTP, and 3.5U of SuperTherm *Taq* polymerase (JMR Holdings, London, UK). Primer concentrations utilized are presented in Table 5-2. PCR cycling conditions were identical to those utilized in detection of individual genes (5.2.2). Agarose gel electrophoresis for detection of amplification products was performed as described in 5.2.2.

Table 5-2. Multiplex PCR assay reagents

Multiplex	Buffer	$MgCl_2$	dNTPs	Primer speA, speC, speG, speH, smeZ	Primer speB, speJ	Primer speF
Initial	1X	0.5 mM	0.4 mM	0.2 µM	0.2 µM	0.2 µM
Optimised	1X	0.5 mM	0.4 mM	0.2 µM	0.3 µM	0.63 µM

5.3 Results

5.3 1 Amplification and sequencing of individual superantigen genes

Single-plex PCR assays were performed on 10 randomly selected isolates and each of the eight genes were amplified separately from at least one of the isolates. The DNA sequences of the amplification products showed sequence homology of at least 99% with published sequences of the respective genes (Appendix E). This was taken as evidence that the primers were amplifying the appropriate genes.

5.3.2 Positive control

No single isolate of the 10 randomly chosen isolates contained all eight genes. However, based on the results of single-plex assay, two isolates that between them contained all the eight genes investigated were identified. The first isolate contained seven genes, *SpeA, SpeB, SpeC, SpeF, SpeG, SpeJ* and *SmeZ* while the second isolate contained four genes, *SpeB, SpeF, SpeG* and *SpeH*. These two isolates were used as positive controls for all subsequent optimisation assays.

5.3.3 Initial multiplex PCR

In the initial multiplex PCR assay, only four genes, *SpeA, SpeC, SpeG and smeZ,* out of the seven genes expected from the first control were amplified while all the four anticipated genes from the second control, *SpeB, SpeF, SpeG* and *SpeH,* were amplified. However, *SpeB*, *SpeF*, *SpeG* and *SpeH* genes from the second control could were amplified but produced faint bands following agarose gel electrophoresis observation (Figure 5-1).

5.3.4 Optimum reagents and primer concentrations for multiplex PCR

Increasing the $MgCl_2$ concentration by 0.5 mM increased the PCR product yield with the best yield observed at 2.5 mM, hence this concentration was adopted. Serial increments of dNTP concentration by 0.1 mM did not result in any increase in amplification products, however, previous studies had observed improved product yield with $MgCl_2$ and dNTP concentrations of 2 mM and 0.4 mM, respectively, thus a final dNTP concentration of 0.4 mM was adopted [186, 187]. Optimisation of

individual primer concentrations had the most dramatic impact on the multiplex PCR. Increasing the concentrations of primers for *speF, speB* and *speJ,* which did not amplify in the initial attempt, (Figure 5-1), resulted in the amplification of *speF*, *speB* and *speJ* genes (Figure 5-2). Primers for *SpeA, SpeC, SpeG, speH* and *smeZ* genes were kept at the initial 0.2 µM while optimum yield of *speF* was observed with 0.63 µM (from the initial 0.2 µM) and both *speB* and *speJ* were amplified using 0.3 µM of primers (Table 5-2).

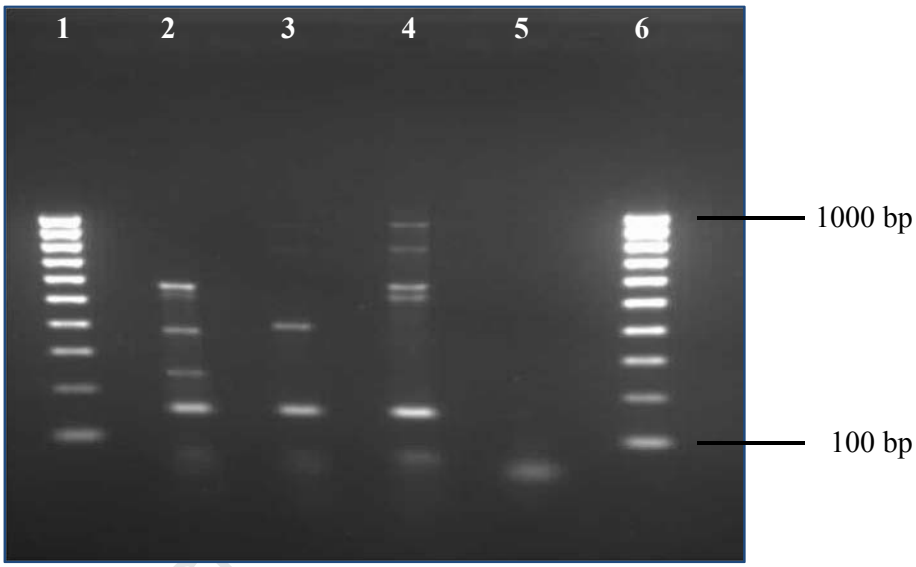


Figure 5-1. Initial multiplex PCR for the amplification of eight streptococcal pyrogenic exotoxins genes. Lane 1 and 6 are molecular markers (HyperLadder VI, Appendix B). Lanes 2 posiive control 1, lane 3 positive control 2 and lane 4 sample 1 and lane 5 negative control (H_2O).

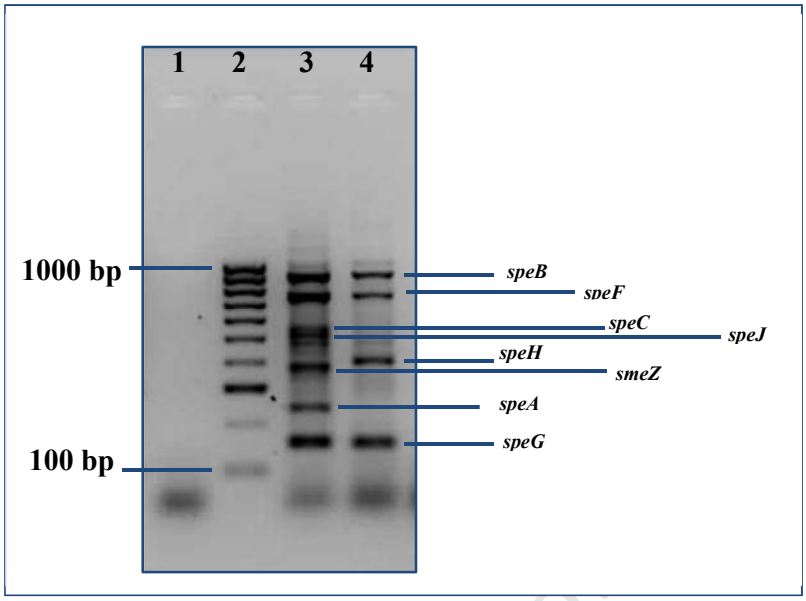


Figure 5-2. Agarose gel electrophoresis following multiplex PCR of eight streptococcal pyrogenic exotoxins. Lane 2 and 12, molecular marker, Lane 1 is a negative control while lane 3 and 4 are positive controls.

5.3.5 Detection of streptococcal pyrogenic superantigens using multiplex PCR

Out of the 98 isolates tested, 21 different superantigen profiles were obtained. These profiles were distributed across different *emm* types (Table 5-3). The most commonly detected gene was *speG* from 98 isolates (94%) and the least frequenly detected was *speA,* from 18 isolates (17%) (Table 5-3). None of the superantigen genes were found in six out of the 98 isolates.

There was a higher prevalence of some of the superantigen genes in isolates from symptomatic cases compared to asymptomatic cases (Figure 5-3), most notably *speA, speB, speC, speF, speH* and *speJ*. However, the difference was only statistically signifuicant for *speC*. . *SpeC* was detected in 71% of the isolates from symptomatic patients compared to 4% of the isolates from asymptomatic patients.

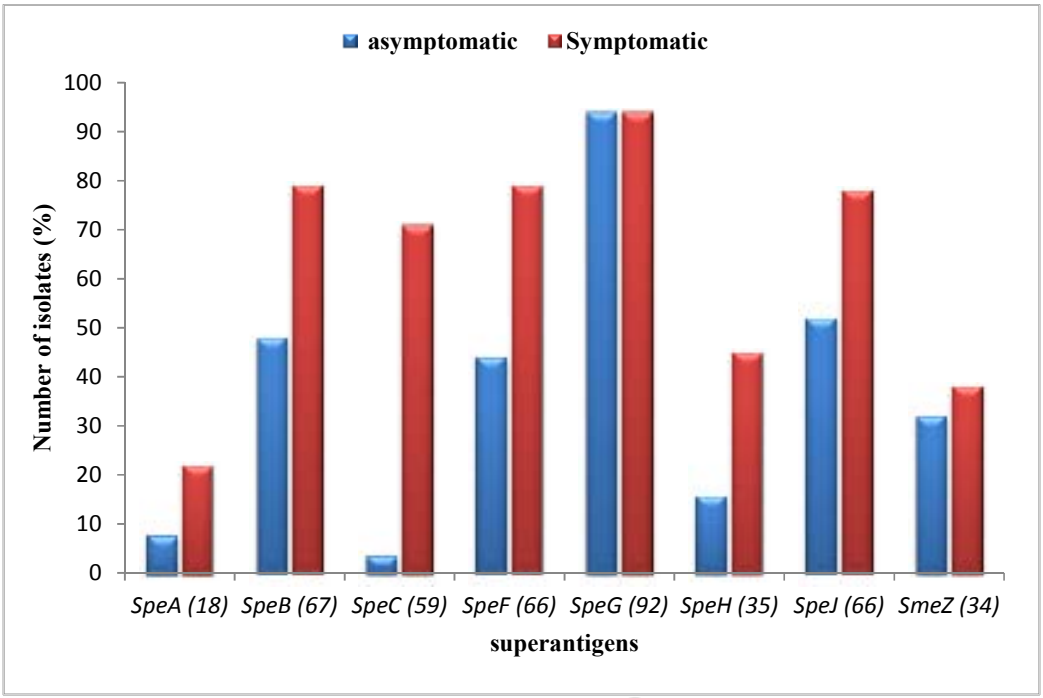


Figure 5-3. The distribution of amplified SAg gene among pharyngeal isolates. The total number of amplified genes are indicated in brackets.

Table 5-3. *emm* types and their respective superantigen profiles.

			Superantigen							
***emm* types**	Superantigen profile	*n*	*SpeA*	*SpeB*	*SpeC*	*SpeF*	*SpeG*	*SpeH*	*SpeJ*	*SmeZ*
emm48.1	A	2	-	-	-	-	+	-	-	-
	B	3	-	+	+	+	+	+	+	-
	D	5	-	+	+	+	+	-	+	+
	E	1	-	+	+	+	+	+	+	+
	F	1	+	+	+	+	+	-	+	+
emm12	A	3	-	-	-	-	+	-	-	-
	B	3	-	+	+	+	+	+	+	-
	G	3	+	+	+	+	+	+	+	-
	H	1	-	-	-	-	+	+	+	-
	P	1	-	-	-	-	+	+	-	-
emm4	B	1	-	+	+	+	+	+	+	-
	C	1	-	+	+	+	+	-	+	-
	D	5	-	+	+	+	+	-	+	+
	M	1	+	+	+	+	+	-	+	-
	N	1	-	+	+	+	+	-	-	+
	R	1	-	-	-	-	+	-	-	+
	S	1	-	+	-	+	+	-	+	+
emm89	A	1	-	-	-	-	+	-	-	-
	B	1	-	+	+	+	+	+	+	-
	C	1	-	+	+	+	+	-	+	-
	F	1	+	+	+	+	+	-	+	+
	H	1	-	-	-	-	+	+	+	-
	K	1	-	+	-	+	+	+	+	-
emm94	B	2	-	+	+	+	+	+	+	-
	C	1	-	+	+	+	+	-	+	-
	E	1	-	+	+	+	+	+	+	+
	F	1	+	+	+	+	+	-	+	+
	G	1	+	+	+	+	+	+	+	-
	P	1	-	-	-	-	+	+	-	-
	T	1	-	+	-	+	+	+	-	-
emm75	C	1	-	+	+	+	+	-	+	-
emm1.0	A	1	-	-	-	-	+	-	-	-
	E	1	-	+	+	+	+	+	+	+
	F	2	+	+	+	+	+	-	+	+
emm22.0	B	2	-	+	+	+	+	+	+	-
	C	3	-	+	+	+	+	-	+	-
	E	1	-	+	+	+	+	+	+	+
	G	1	+	+	+	+	+	+	+	-
emm2.0	A	3	-	-	-	-	+	-	-	-
	C	1	-	+	+	+	+	-	+	-
emm9.2	A	3	-	-	-	-	+	-	-	-
	C	3	-	+	+	+	+	-	+	-
	O	1	-	+	+	+	+	-	-	-
emm87	A	2	-	-	-	-	+	-	-	-
emm116	J	2	+	+	-	+	+	-	+	+
emm82.0	A	1	-	-	-	-	+	-	-	-
emm3	Q	1	+	+	+	+	+	-	-	+
emm6.4	A	1	-	-	-	-	+	-	-	-
	E	1	-	+	+	+	+	+	+	+
emm8.0	D	1	-	+	-	+	+	-	+	+
	F	1	+	+	+	+	+	-	+	+
	I	1	-	+	-	+	+	+	+	+
emm44.0	G	1	+	+	+	+	+	+	+	-
emm53.0	E	1	-	+	+	+	+	+	+	+

Continued from previous page

emm58.2	**A**	**1**	-	-	-	-	+	-	-	-
	L	1	+	+	+	-	+	-	+	+
emm73	B	1	-	+	+	+	+	+	+	-
emm92	E	1	-	+	+	+	+	+	+	+
	H	1	-	-	-	-	+	+	+	-
emm28.0	D	1	-	+	+	+	+	-	+	+
emm33	E	1	-	+	+	+	+	+	+	+
emm77	I	1	-	+	-	+	+	+	+	+
emm64	U	1	-	-	-	-	+	-	+	-
emm80	F	1	+	+	+	+	+	-	+	+
Emm49	A	1	-	-	-	-	+	-	-	-
emm118.1	A	1	-	-	-	-	+	-	-	-
emm75	A	1	-	-	-	-	+	-	-	-
st2002.2	B	1	-	+	+	+	+	+	+	-

- Undetected
+ detected

5.4 Discussion

Among the 98 GAS isolates investigated, the most common SAg was *speG*, present in 94% of the isolates while the least frequently detected SAg was *speA* in only 17% of the isolates (Table 5-3 and Figure 5-3). The *speG* gene is one of the three chromosomally coded genes [41] as elaborated in Chapter One (1.1.3) hence it is anticipated that it would be present in all the isolated investigated.

The other two chromosomally coded genes [41], *SpeJ* and *SmeZ* were present in 67% and 35% isolates, respectively. Up to 100% detection of *speG* has been reported by some studies [34, 41, 190, 191] suggesting chromosomal coding, while other studies have documented this gene in 90% [30] and 84.1% [37] of the isolates. The other chromosomally coded genes, *speJ* and *smeZ,* were detected in 33.3% and 91.7% of the isolates by Rivera [37], respectively, while Commons detected these genes in 51% and 95% of the isolates, respectively [30, 37]. These results accord with those reported by Schmitz which did not identify *speJ* and *smeZ* in all the isolates investigated [42]. The gene for *smeZ* is believed to have 23 alleles [37] hence the primer designed for a single allele could play a crucial role in the success of the multiplex PCR. Another factor that could interfere with the detection of these genes is the genomic DNA extraction method utilized. The study by Commons reported that the DNA extraction method may interfere with PCR amplification [30]. The authors observed that PCR amplification of DNA extracted by mechanical agitation was much more sensitive than PCR amplifying DNA extracted by the boiling method (heating at 100ºC for 10 minutes) [10, 30].

Six isolates in this study lacked all the eight SAg genes, *speA, speB, speC, speF, speG, speH, speJ* and *smeZ*, despite attempts to amplify *speG* and *smeZ* in single-plex reaction. The closed-tube multiplex PCR designed by Lingtes also identify a strain without any SAg gene [27]. The multiplex PCR was designed to detect two set of genes coding *spea1-3+5, spec, spej, spek, speal* and *speg* in the first set and the second set contained *spea1-4, speh, spei, spem, ssa* and *smez*, however genes coding *speB* and *speF* were not included [27]. Another study conducted by Rivera also identified an isolate with no SAg genes when nine genes, namely; *speA, speB, speC, speH, speI, speJ speM, ssa*, and *smeZ,* were investigated [37].

Twenty one different SAg profiles were identified among 98 GAS isolates (Table 5-3). Twenty one of the isolates (20%) contained only *speG* while the rest of the profiles were evenly distributed among the isolates. The distribution of the SAg profiles showed no association between any particular profile and isolates from symptomoatic versus asymptomatic patients.. Other studies have documented associations between SAg and particular diseases. For instance, *speA* and *speC* have both been associated with scarlet fever and STSS [29, 31-33]. A study conducted by Schmitz, observed no connection between GAS diseases and a particular SAg profile [42], with no significant (P>.01) differences between invasive, wound and pharyngeal isolates. Thus, according to Schmitz, possession of a particular SAg profile does not characterise GAS isolates [42]. In Japan, a study investigating 316 GAS isolates from invasive (bacteremia and toxic shock-like syndrome) and non-invasive infections (mucosa, pharynx, tonsils, respiratory track, ears, nose, vagina, stool and urethra) also found no correlation between SAgs and GAS isolation sites [54].

Some *emm* types are associated with distinct SAg profiles [37]. The twenty one SAg profiles observed in this study did not show any association with the predominant *emm* types (*emm48*, *emm12*, *emm4* and *emm89*) (Table 5-3). There is a random distribution of superantigens profiles identified among *emm* types, suggesting diversity rather than *emm* type specificity e.g. SAg profile A (*speG*) only was identified in thirteen different *emm* types; profile B (*speB,C,F,G,H,J*) was observed eight *emm* types (Table 5-3). The SAg profiles illustrated in Table 5-3 and those seen in another study [37] show that isolates belonging to the same *emm* type do not necessarily have similar SAg profiles. The study by Rivera identified at least a single toxin gene difference between similar *emm* types [37]. However, these finding contradicts previous studies which report restriction of certain SAg profiles to *emm* types [30, 42]. The study by Schmitz observed certain *emm* types were characterised by dominance of 1 or 2 toxin genes profiles e.g. 40 out of 72 *emm1* investigated contained *speA*, *speB*, *speF*, *speG*, *speJ* and *smeZ* [42]. However, other *emm1* with variable SAg profiles were present [42]. Interestingly 27 out of 29 *emm1* reported by Commons featured *speA*, *speG*, *speJ* and *smeZ* [30]. These SAg were also harboured in *emm4* [30].

Individual *emm* types such as; *emm1, emm4, emm12, emm28*, and *emm75,* are said to feature unique SAg genes [30, 35, 87]. In this study, *emm1* featured in three SAg profiles, A, E and F, which collectively featured all eight genes investigated, (Table 5-3). However the SAg gene identified in *emm1* are not in accordance with a study by Descheemaeker which documented presence of *speA, speB* and *speF* in *emm1* and also noticed the absence of *speC* and *ssa*, which was also observed in a study by Vlamininckx [35, 87]. The absence of *speC* from *emm1* contrasts with a Tunisian study which demonstrated presence of this gene in *emm1*, in line with two more studies [30, 37, 48]. Hence, with these results in mind, *emm1* in some instances does not exclusively lack *speC*, since results from this study and other studies [30, 37, 48] have identified this SAg gene from *emm1*.

The multiplex PCR in this study was successful in determining the SAg profiles from the 98 isolates. Regardless of the SAg profile identified, there was no close relationship with the *emm* types. The same applies to the most prevalent streptococcal exotoxin gene.

Conclusions

The results of the *emm*-typing show a greater diversity of *emm*-typpes among GAS isolates from Cape Town than has been reported in developed countries, and the proposed 26-valent vaccine would theoretically provide 51.7% coverage. This vaccine coverage supports calls for a reconsideration of the M-protein currently included in the 26-valent and the 30-valent GAS vaccine. The coverage observed is typical of developing countries where a diversity of *emm*-types exist, contrary to the high vaccine coverage observed in developed countries due to the dominance of only a few *emm*-types. Based on the data from this study, the dominant M48.1 does not featured in the experimental 26-valent GAS vaccines translating into a vaccine coverage of 51.7% contrasting up to 85% coverage anticipated from developed countries such as the US and Canada.

As an alternative to the sequencing assay, restriction fragment length polymorphism (RFLP) assay produced restriction patterns distinct enough to enable differentiation of the locally observed *emm*-types. The ability of restriction enzymes to produce distinguishable restriction patterns of M-proteins makes RFLP a useful assay for typing M-protein. Of the five enzymes recommended by the CDC, double digestion with *Hae*III and *Hinc*II produced distinct restriction patterns that were relatively easy to analyse. Although not commonly used in molecular epidemiology of the M-protein, RFLP results from this study suggests the effectiveness of this assay in determining differences among M-protein, particularly when a large number of isolates are involved. Hence, with regards to some developing countries where sequencing can be a challenge, RFLP can be useful in segregating similar M-protein, reducing the number of samples for sequencing, particularly when a large sample size is involved.

Sequence analyses of the conserved region of the M-protein confirm four potential vaccine targets conserved among GAS isolates. Contrary to the type specific 26-valent and 30-valent GAS vaccine in which a low vaccine coverage is evident in Cape Town, three out of the four vaccine targets are harboured in at least two conserved regions across all *emm*-types examined, regardless of the *emm*- type. The results from this study as well as reports from other studies across the globe have identified similar vaccine targets in GAS isolates. A vaccine targeting the four conserved regions has a potential to have universal relevance, unlike the M-protein based vaccine which has geographical variation in coverage.

The first multiplex PCR utilized to detect superantigen genes of GAS, streptococcal superantigens in South Africa, show no connection between SAg and M-protein types. Among the eight genes investigated, restriction profiles identified show no mutual relation with the M-protein types. Although superantigens are believed to largely contribute to the global surge of streptococcal diseases, no individual superantigen among those investigated in this study correlate with pharyngeal infection.

The thesis had a few limitations. The number of participants in each of the symptomatic and asymptomatic groups was small, thus not allowing for meaningful comparisons between the groups. The dearth of molecular data from South and sub-Saharan Africa also made it difficult to contextualise our results, especially against those data from developed countries where the majority of vaccine development is underway.

In summary, to the best of author's knowledge, this is the first study describing the molecular epidemiology of GAS in South Africa. The M-protein sequence typing as well RFLP proved useful in identifying GAS. Sequence typing remains the gold standard for molecular epidemiology studies of GAS. Both sequencing and RFLP show a diversity of M-protein circulating in Cape Town, calling for new vaccine formulation that will include prominent M-protein from Cape Town. Interestingly, sequence analysis of the conserved region of the M-protein identified potential vaccine targets conserved among GAS strains, injecting some hope of a much needed

GAS vaccine that is not M-protein specific and can be used regardless of the global distribution of the M-protein.

References

1. Mandell GL BJaDR. Streptococcus pyogenes. In: Dudlick THaM, ed. Principles and practice of Infectious diseases Vol. Volume 2, **2005**:2360-80.

2. Griffith F. The Serological Classification of Streptococcus pyogenes. J Hyg (Lond) **1934** Dec;34(4):542-84.

3. Lancefield RC. A Serological Differentiation of Human and Other Groups of Hemolytic Streptococci. J Exp Med **1933** Mar 31;57(4):571-95.

4. Hitchcock CH. Classification of the Hemolyttc Streptococci by the Precipitin Reaction. J Exp Med **1924** Sep 30;40(4):445-52.

5. Lancefield RC. The Antigenic Complex of Streptococcus Haemolyticus : Iii. Chemical and Immunological Properties of the Species-Specific Substance. J Exp Med **1928** Feb 29;47(3):481-91.

6. Lancefield RC. The Antigenic Complex of Streptococcus Haemolyticus : Ii. Chemical and Immunological Properties of the Protein Fractions. J Exp Med **1928** Feb 29;47(3):469-80.

7. Lancefield RC. The Antigenic Complex of Streptococcus Haemolyticus : I. Demonstration of a Type-Specific Substance in Extracts of Streptococcus Haemolyticus. J Exp Med **1928** Jan 1;47(1):91-103.

8. Bisno AL, Brito MO, Collins CM. Molecular basis of group A streptococcal virulence. *The lancet infectious diseases* **2003**;3:191-200.

9. Neal S BB, Ekelund K, Henriques-Normark B, Jasir A, Johnson D, Kaplan E, Lovgren M, Reinert RR and Efstratiou A Members of the Strep-EURO Study Group and International Streptococcus Reference Laboratories International Quality Assurance Study for Characterization of *Streptococcus pyogenes*. *Journal of clinical microbiology* **2007**;45:1175-9

10. Beall B, Facklam R, Thompson T. Sequencing emm-specific PCR products for routine and accurate typing of group A streptococci. J Clin Microbiol **1996** Apr;34(4):953-8.

11. 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;9(10):611-6.

12. Bisno AL. Alternate complement pathway activation by group A streptococci: role of M-protein. Infect Immun **1979** Dec;26(3):1172-6.

13. Kronvall G. A surface component in group A, C, and G streptococci with non-immune reactivity for immunoglobulin G. J Immunol **1973** Nov;111(5):1401-6.

14. Oehmcke S, Shannon O, Morgelin M, Herwald H. Streptococcal M proteins and their role as virulence determinants. Clin Chim Acta **2010** Sep 6;411(17-18):1172-80.

15. Wessels MR, Moses AE, Goldberg JB, DiCesare TJ. Hyaluronic acid capsule is a virulence factor for mucoid group A streptococci. Proc Natl Acad Sci U S A **1991** Oct 1;88(19):8317-21.

16. 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** Dec;8(12):1705-20.

17. Bisno AL, Brito MO, Collins CM. Molecular basis of group A streptococcal virulence. Lancet Infect Dis **2003** Apr;3(4):191-200.

18. Courtney HS, Hasty DL, Dale JB. Molecular mechanisms of adhesion, colonization, and invasion of group A streptococci. Ann Med **2002**;34(2):77-87.

19. Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G. Pili in gram-positive pathogens. Nat Rev Microbiol **2006** Jul;4(7):509-19.

20. Manetti AG, Zingaretti C, Falugi F, et al. Streptococcus pyogenes pili promote pharyngeal cell adhesion and biofilm formation. Mol Microbiol **2007** May;64(4):968-83.

21. Wang JR, Stinson MW. M protein mediates streptococcal adhesion to HEp-2 cells. Infect Immun **1994** Feb;62(2):442-8.

22. Okada N, Liszewski MK, Atkinson JP, Caparon M. Membrane cofactor protein (CD46) is a keratinocyte receptor for the M protein of the group A streptococcus. Proc Natl Acad Sci U S A **1995** Mar 28;92(7):2489-93.

23. Hanski E, Horwitz PA, Caparon MG. Expression of protein F, the fibronectin-binding protein of Streptococcus pyogenes JRS4, in heterologous streptococcal and enterococcal strains promotes their adherence to respiratory epithelial cells. Infect Immun **1992** Dec;60(12):5119-25.

24. Proft T, Sriskandan S, Yang L, Fraser JD. Superantigens and streptococcal toxic shock syndrome. Emerg Infect Dis **2003** Oct;9(10):1211-8.

25. Proft T, Fraser JD. Bacterial superantigens. Clin Exp Immunol **2003** Sep;133(3):299-306.

26. Langley R, Wines B, Willoughby N, Basu I, Proft T, Fraser JD. The staphylococcal superantigen-like protein 7 binds IgA and complement C5 and inhibits IgA-Fc alpha RI binding and serum killing of bacteria. J Immunol **2005** Mar 1;174(5):2926-33.

27. Lintges M, Arlt S, Uciechowski P, et al. A new closed-tube multiplex real-time PCR to detect eleven superantigens of Streptococcus pyogenes identifies a strain without superantigen activity. Int J Med Microbiol **2007** Oct;297(6):471-8.

28. Curtis N. Invesive group Astreptococcal infection. *Current oppinions on infectious Diseases* **1996**;9:191-202.

29. Sriskandan S, Faulkner L, Hopkins P. Streptococcus pyogenes: Insight into the function of the streptococcal superantigens. Int J Biochem Cell Biol **2007**;39(1):12-9.

30. Commons R RS, Gooding T, Danchin M, Carapetis J, Robins-Browne R and Curtis N. Superantigen genes in group A streptococcal isolates and their relationship with *emm* types. *Journal of Medical Microbiology* **2008**;57.

31. Stevens DL, Tanner MH, Winship J, et al. Severe group A streptococcal infections associated with a toxic shock-like syndrome and scarlet fever toxin A. N Engl J Med **1989** Jul 6;321(1):1-7.

32. Liang Y, Shen X, Huang G, Wang C, Shen Y, Yang Y. Characteristics of Streptococcus pyogenes strains isolated from Chinese children with scarlet fever. Acta Paediatr **2008** Dec;97(12):1681-5.

33. Musser JM, Hauser AR, Kim MH, Schlievert PM, Nelson K, Selander RK. Streptococcus pyogenes causing toxic-shock-like syndrome and other invasive diseases: clonal diversity and pyrogenic exotoxin expression. Proc Natl Acad Sci U S A **1991** Apr 1;88(7):2668-72.

34. Smoot LM MJ, Smoot JC, Hoe NP, Strickland I, Cole RL, Barbian KD, Earhart CA, Ohlendorf DH, Veasy LG, Hill HR, Leung DYM, Schlievert PM and Musser JM Characterization of two novel pyrogenic toxic superantigens made by an acute rheumatic fever clone of *Streptococcus pyogenes* associated with multiple disease outbreaks. *Journal of Infection and immunity* **2002**;70:7095-104.

35. Vlaminckx BJ, Mascini EM, Schellekens J, et al. Site-specific manifestations of invasive group a streptococcal disease: type distribution and corresponding patterns of virulence determinants. J Clin Microbiol **2003** Nov;41(11):4941-9.

36. Norrby-Teglund A, Thulin P, Gan BS, et al. Evidence for superantigen involvement in severe group a streptococcal tissue infections. J Infect Dis **2001** Oct 1;184(7):853-60.

37. Rivera A, Rebollo M, Miro E, et al. Superantigen gene profile, emm type and antibiotic resistance genes among group A streptococcal isolates from Barcelona, Spain. J Med Microbiol **2006** Aug;55(Pt 8):1115-23.

38. Weeks CR, Ferretti JJ. The gene for type A streptococcal exotoxin (erythrogenic toxin) is located in bacteriophage T12. Infect Immun **1984** Nov;46(2):531-6.

39. Proft T, Moffatt SL, Weller KD, Paterson A, Martin D, Fraser JD. The streptococcal superantigen SMEZ exhibits wide allelic variation, mosaic structure, and significant antigenic variation. J Exp Med **2000** May 15;191(10):1765-76.

40. Proft T, Webb PD, Handley V, Fraser JD. Two novel superantigens found in both group A and group C Streptococcus. Infect Immun **2003** Mar;71(3):1361-9.

41. Ekelund K, Darenberg J, Norrby-Teglund A, et al. Variations in emm type among group A streptococcal isolates causing invasive or noninvasive infections in a nationwide study. J Clin Microbiol **2005** Jul;43(7):3101-9.

42. Schmitz FJ, Beyer A, Charpentier E, et al. Toxin-gene profile heterogeneity among endemic invasive European group A streptococcal isolates. J Infect Dis **2003** Nov 15;188(10):1578-86.

43. Darenberg Jessica L-HB, Jasir A, Sandgeren A, Petterson H, Schalen C, Nrgen M, Romanus V, Norrby-Teglund and Normark BH Molecular and clinical characteristics of invasive Group A Streptococcal infection in Sweden
Journal of Clinical infectious disease **2007**;45.

44. Smoot JC, Korgenski EK, Daly JA, Veasy LG, Musser JM. Molecular analysis of group A Streptococcus type emm18 isolates temporally associated with acute rheumatic fever outbreaks in Salt Lake City, Utah. J Clin Microbiol **2002** May;40(5):1805-10.

45. CDC. Blast-emm. 2011 [cited; Available from: http://www.cdc.gov/ncidod/biotech/strep/strepblast.htm

46. Abdissa A AD, Kronvallb G, Shittu B, Achiko D, Zeidan M Yamuah LK and Aseffa High Diversity of Group A Streptococcal *emm* Types among Healthy Schoolchildren in Ethiopia. *Journal of Clinical Infectious Diseases* **2006**;49:1362-7.

47. Tewodros W, Kronvall G. M protein gene (emm type) analysis of group A beta-hemolytic streptococci from Ethiopia reveals unique patterns. J Clin Microbiol **2005** Sep;43(9):4369-76.

48. Hraoui M, Boutiba-Ben Boubaker I, Doloy A, Samir E, Ben Redjeb S, Bouvet A. Epidemiological markers of Streptococcus pyogenes strains in Tunisia. Clin Microbiol Infect **2011** Jan;17(1):63-8.

49. Chang H, Shen X, Huang G, et al. Molecular analysis of Streptococcus pyogenes strains isolated from Chinese children with pharyngitis. Diagn Microbiol Infect Dis **2011** Feb;69(2):117-22.

50. Chen YY, Huang CT, Yao SM, et al. Molecular epidemiology of group A streptococcus causing scarlet fever in northern Taiwan, 2001-2002. Diagn Microbiol Infect Dis **2007** Jul;58(3):289-95.

51. Steer AC, Magor G, Jenney AW, et al. emm and C-repeat region molecular typing of beta-hemolytic Streptococci in a tropical country: implications for vaccine development. J Clin Microbiol **2009** Aug;47(8):2502-9.

52. Steer AC, Jenney A, Kado J, et al. Prospective surveillance of invasive group a streptococcal disease, Fiji, 2005-2007. Emerg Infect Dis **2009** Feb;15(2):216-22.

53. Menon T, Lloyd C, Malathy B, Sakota V, Jackson D, Beall B. emm type diversity of beta-haemolytic streptococci recovered in Chennai, India. J Med Microbiol **2008** Apr;57(Pt 4):540-2.

54. Murakami J, Kawabata S, Terao Y, et al. Distribution of emm genotypes and superantigen genes of Streptococcus pyogenes isolated in Japan, 1994-9. Epidemiol Infect **2002** Jun;128(3):397-404.

55. Bahnan W HF, Araj G and Tokajian. emm typing , antibiotic resistance and PFGE analysis of *Streptococcus pyogenes* in Lebanon. Journal of Medical Microbiology **2011**;60.

56. Alfaresi MS. Group A streptococcal genotypes from throat and skin isolates in United Arab Emirates. BMC Research Notes **2010**;3.

57. Eisner A, Leitner E, Feierl G, Kessler HH, Marth E. Prevalence of emm types and antibiotic resistance of group A streptococci in Austria. Diagn Microbiol Infect Dis **2006** Aug;55(4):347-50.

58. Brandt CM SB, Honscha M, Troung ND, Hoenvener B and Lutticken R Typing of Streptococcus pyogenes Strains isolated from Throat infections in the region of Aachen Germany. *Journal of Infection and immunity* **2001**;2001:163-6.

59. Creti R IM, Baldassarri L, Pataracchia M, Recchia S, Alfarone G and Orefici G. emm Tyes, Virulence Factors, and Antibiotic Resistance of Invasive *Streptococcus pyogenes* Isolates from Italy: What Has Changed in 11 Years? Journal of Clinical Micribiology **2007**;45.

60. Dicuonzo G, Gherardi G, Lorino G, et al. Group A streptococcal genotypes from pediatric throat isolates in Rome, Italy. J Clin Microbiol **2001** May;39(5):1687-90.

61. Kittang BR LNaMH. Distribution of emm types and subtypes among noninvasive group A, C and G streptococcal isolates in western Norway. Acta Pathologica Microbiologica Et Immunologica Scandivanica **2008**;116.

62. Luca-Harari B, Straut M, Cretoiu S, et al. Molecular characterization of invasive and non-invasive Streptococcus pyogenes isolates from Romania. J Med Microbiol **2008** Nov;57(Pt 11):1354-63.

63. Montes M, Tamayo E, Orden B, Larruskain J, Perez-Trallero E. Prevalence and clonal characterization of Streptococcus pyogenes clinical isolates with reduced fluoroquinolone susceptibility in Spain. Antimicrob Agents Chemother **2010** Jan;54(1):93-7.

64. Luca-Harari B, Darenberg J, Neal S, et al. Clinical and microbiological characteristics of severe Streptococcus pyogenes disease in Europe. J Clin Microbiol **2009** Apr;47(4):1155-65.

65. Espinosa LE, Li Z, Gomez Barreto D, et al. M protein gene type distribution among group A streptococcal clinical isolates recovered in Mexico City, Mexico, from 1991 to 2000, and Durango, Mexico, from 1998 to 1999: overlap with type distribution within the United States. J Clin Microbiol **2003** Jan;41(1):373-8.

66. Perea-Mejia LM, Inzunza-Montiel AE, Cravioto A. Molecular characterization of group A Streptococcus strains isolated during a scarlet fever outbreak. J Clin Microbiol **2002** Jan;40(1):278-80.

67. Shulman ST TR, DaleJ B, Beall B, Kabat W, Cederlund KKE, Patel D, Rippe J , Li Z, Sakota V, and the North American Streptococcal Pharyngitis Surveillance Group Seven-Year Surveillance of North American Pediatric Group A Streptococcal Pharyngitis Isolates. *Journal of Clinical Infectious Diseases* **2009**;49:78-84.

68. Erdem G, Mizumoto C, Esaki D, Abe L, Reddy V, Effler PV. Streptococcal emm types in Hawaii: a region with high incidence of acute rheumatic fever. Pediatr Infect Dis J **2009** Jan;28(1):13-6.

69. Smeesters PR, Vergison A, Campos D, de Aguiar E, Miendje Deyi VY, Van Melderen L. Differences between Belgian and Brazilian group A Streptococcus epidemiologic landscape. PLoS One **2006**;1:e10.

70. Enright MC, Spratt BG, Kalia A, Cross JH, Bessen DE. Multilocus sequence typing of Streptococcus pyogenes and the relationships between emm type and clone. Infect Immun **2001** Apr;69(4):2416-27.

71. Yoonim N, Olive C, Pruksachatkunakorn C, Good MF, Pruksakorn S. M protein typing of Thai group A streptococcal isolates by PCR-Restriction fragment length polymorphism analysis. BMC Microbiol **2005**;5:63.

72. Olive DM, Bean P. Principles and applications of methods for DNA-based typing of microbial organisms. J Clin Microbiol **1999** Jun;37(6):1661-9.

73. MC AvBaE. Newer methods for bacterial strain typing. Clinican Microbiology newsletter **2008**;30:9.

74. Luca-Harari B SM, Cretoiu S, Surdeanu M, Ungureanu V, van de Linden M and Aftab J Molecular characterization of invasive and non invasive *streptococcus pyogenes* isolates from Romania. *Journal of Medical Microbiology* **2008**;57:1354-63.

75. Darenberg Jessica L-HB, Jasir A, Sandgeren A, Petterson H, Schalen C, Nrgen M, Romanus V, Norrby-Teglund and Normark BH Molecular and clinical characteristics of invasive Group A Streptococcal infection in Sweden. *Journal of Clinical infectious disease* **2007**;45.

76. Bianco S, Allice T, Zucca M, Savoia D. Survey of phenotypic and genetic features of streptococcus pyogenes strains isolated in Northwest Italy. Curr Microbiol **2006** Jan;52(1):33-9.

77. Enright MC, Spratt BG. Multilocus sequence typing. Trends Microbiol **1999** Dec;7(12):482-7.

78. Smith HO, Wilcox KW. A restriction enzyme from Hemophilus influenzae. I. Purification and general properties. J Mol Biol **1970** Jul 28;51(2):379-91.

79. Kelly TJ, Jr., Smith HO. A restriction enzyme from Hemophilus influenzae. II. J Mol Biol **1970** Jul 28;51(2):393-409.

80. Beall B, Facklam RR, Elliott JA, et al. Streptococcal emm types associated with T-agglutination types and the use of conserved emm gene restriction fragment patterns for subtyping group A streptococci. J Med Microbiol **1998** Oct;47(10):893-8.

81. Villasenor-Sierra A, Quinonez-Alvarado MG, Caballero-Hoyos JR. Interpersonal relationships and group A streptococcus spread in a Mexican day-care center. Salud Publica Mex **2007** Sep-Oct;49(5):323-9.

82. Mylvaganam H, Bjorvatn B, Hofstad T, Osland A. Emm gene polymorphism among temporally clustered group A streptococcal isolates in western Norway. APMIS **2000** Apr;108(4):303-12.

83. Balaraman M, Menon T. Restriction fragment patterns and emm types of group G streptococci. J Med Microbiol **2010** Aug;59(Pt 8):996-7.

84. Steer Andrew C LI, Matatolv Laisiana, Meall Bernard W and Carapetis Jonathan R Global *emm* type distribution of group A streptococci: systematic review and implications for vaccine development. *Lancet Infectious Disease* **2009**;9:611-6.

85. Shulman ST, Stollerman G, Beall B, Dale JB, Tanz RR. Temporal changes in streptococcal M protein types and the near-disappearance of acute rheumatic fever in the United States. Clin Infect Dis **2006** Feb 15;42(4):441-7.

86. Bessen DE, Carapetis JR, Beall B, et al. Contrasting molecular epidemiology of group A streptococci causing tropical and nontropical infections of the skin and throat. J Infect Dis **2000** Oct;182(4):1109-16.

87. Descheemaeker P, Van Loock F, Hauchecorne M, Vandamme P, Goossens H. Molecular characterisation of group A streptococci from invasive and non-invasive disease episodes in Belgium during 1993-1994. J Med Microbiol **2000** May;49(5):467-71.

88. Mandell GL BJaDR. Principles and practice of infectious diseases, **2005**.

89. Mahajan VK, Sharma NL. Scarlet fever. Indian Pediatr **2005** Aug;42(8):829-30.

90. Steer AC, Jenney AW, Kado J, et al. High burden of impetigo and scabies in a tropical country. PLoS Negl Trop Dis **2009**;3(6):e467.

91. Mandell GL BJaDR. Principles and practice of infectious diseases, **2005**.

92. Press OU. Concise Medical Dictionary. OXFORD. 8th ed Vol. 1. New York: Oxfrod University Inc., **2010**:561.

93. Alcaide ML, Bisno AL. Pharyngitis and epiglottitis. Infect Dis Clin North Am **2007** Jun;21(2):449-69, vii.

94. Mandell GL BJaDR. Principal and practice of infectious diseases. 6th ed, **2005**.

95. Bisno AL. Acute pharyngitis. N Engl J Med **2001** Jan 18;344(3):205-11.

96. Bisno AL. Group A Streptococcal infections and acute rheumatic fever. *New England Journal of Medicine* **1991**;325:783-93.

97. Carapetis JR, Steer AC, Mulholland E, K, M W. The global burden of group A streptococcal disease. *Lancet Infectious Disease* **2005**;5:685-94.

98. **Mandell GL DR, Bennett JE**. **Principles and practice of infectious diseases**. 6th ed: **New York: Churchill Livingstone**, **2006**.

99. Ebell MH, Smith MA, Barry HC, Ives K, Carey M. The rational clinical examination. Does this patient have strep throat? JAMA **2000** Dec 13;284(22):2912-8.

100. Olivier L, de Graad G. Streptococcal sore throat. S Afr Med J **1978** Dec 23;54(26):1082.

101. van Zyl ML, van Staden DA, Potgieter MD. [Beta-hemolytic streptococci as a cause of sore throat in the Pretoria area]. S Afr Med J **1981** May 23;59(22):783-4.

102. Mandell GL. *Principles and practice of Infectious diseases*, **2005**.

103. van Staden DA, Nel W, van Zyl ML. [Group A beta-hemolytic streptococci in a traditional black community]. S Afr Med J **1982** Oct 9;62(16):569-70.

104. Ransome OJ, Roode H, Spector I, Reinach SG. Pharyngeal carriage of group A beta-haemolytic streptococci in coloured and Indian schoolchildren. S Afr Med J **1983** Nov 5;64(20):779-81.

105. Benouda A, Sibile S, Ziane Y, Elouennass M, Dahani K, Hassani A. [Place of Streptococcus pyogenes in the throat infection in Morocco and overview of its susceptibility to antibiotics]. Pathol Biol (Paris) **2009** Feb;57(1):76-80.

106. Mezghani Maalej S, Rekik M, Boudaouara M, et al. [Childhood pharyngitis in Sfax (Tunisia): epidemiology and utility of a rapid streptococcal test]. Med Mal Infect Apr;40(4):226-31.

107. Brink AJ, Cotton MF, Feldman C, et al. Guideline for the management of upper respiratory tract infections. S Afr Med J **2004** Jun;94(6 Pt 2):475-83.

108. Manyemba J, Mayosi BM. Intramuscular penicillin is more effective than oral penicillin in secondary prevention of rheumatic fever--a systematic review. S Afr Med J **2003** Mar;93(3):212-8.

109. Shulman ST, Bisno AL, Clegg HW, et al. Clinical practice guideline for the diagnosis and management of group a streptococcal pharyngitis: 2012 update by the infectious diseases society of america. Clin Infect Dis **2012** Nov;55(10):e86-e102.

110. Bisno AL, Gerber MA, Gwaltney JM, Jr., Kaplan EL, Schwartz RH. Practice guidelines for the diagnosis and management of group A streptococcal pharyngitis. Infectious Diseases Society of America. Clin Infect Dis **2002** Jul 15;35(2):113-25.

111. Zabriskie JB. Mimetic relationships between group A streptococci and mammalian tissues. Adv Immunol **1967**;7:147-88.

112. Chang C. Cutting Edge Issues in Rheumatic Fever. Clin Rev Allergy Immunol **2011** May 20.

113. Ellis NM, Kurahara DK, Vohra H, et al. Priming the immune system for heart disease: a perspective on group A streptococci. J Infect Dis **2010** Oct 1;202(7):1059-67.

114. Chang C. Cutting edge issues in rheumatic fever. Clin Rev Allergy Immunol **2011** Apr;42(2):213-37.

115. 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** Mar;31(6):719-27.

116. Pastore S, De Cunto A, Benettoni A, Berton E, Taddio A, Lepore L. The resurgence of rheumatic fever in a developed country area: the role of echocardiography. Rheumatology (Oxford) **2011** Feb;50(2):396-400.

117. Tibazarwa KB, Volmink JA, Mayosi BM. Incidence of acute rheumatic fever in the world: a systematic review of population-based studies. Heart **2008** Dec;94(12):1534-40.

118. Markowitz M. The decline of rheumatic fever: role of medical intervention. Lewis W. Wannamaker Memorial Lecture. J Pediatr **1985** Apr;106(4):545-50.

119. Mayosi B, Robertson K, Volmink J, et al. The Drakensberg declaration on the control of rheumatic fever and rheumatic heart disease in Africa. S Afr Med J **2006** Mar;96(3 Pt 2):246.

120. Mayosi B. The four pillars of rheumatic heart disease control. S Afr Med J **2010** Aug;100(8):506.

121. Robertson KA, Volmink JA, Mayosi BM. Towards a uniform plan for the control of rheumatic fever and rheumatic heart disease in Africa--the Awareness Surveillance Advocacy Prevention (A.S.A.P.) Programme. S Afr Med J **2006** Mar;96(3 Pt 2):241.

122. Robertson KA, Volmink JA, Mayosi BM. Lack of adherence to the national guidelines on the prevention of rheumatic fever. S Afr Med J **2005** Jan;95(1):52-6.

123. Bach JF, Chalons S, Forier E, et al. 10-year educational programme aimed at rheumatic fever in two French Caribbean islands. Lancet **1996** Mar 9;347(9002):644-8.

124. Robertson KA, Volmink JA, Mayosi BM. Antibiotics for the primary prevention of acute rheumatic fever: a meta-analysis. BMC Cardiovasc Disord **2005**;5(1):11.

125. Mayosi BM. Protocols for antibiotic use in primary and secondary prevention of rheumatic fever. S Afr Med J **2006** Mar;96(3 Pt 2):240.

126. McDonald M, Brown A, Noonan S, Carapetis JR. Preventing recurrent rheumatic fever: the role of register based programmes. Heart **2005** Sep;91(9):1131-3.

127. Nkgudi B, Robertson KA, Volmink J, Mayosi BM. Notification of rheumatic fever in South Africa -- evidence for underreporting by health care professionals and administrators. S Afr Med J **2006** Mar;96(3):206-8.

128. McNeil SA HS, Langley JM, Smith B, Warren A, Sharratt G P, Baxendale DM., Reddish MA., Hu MC, Stroop SD, Linden J, Fries LF, Vink PE and Dale JB

Safety and Immunogenicity of 26-Valent Group A *Streptococcus* Vaccine in Healthy Adult Volunteers *Clinical Infectious Diseases* **2005**;41:1114-22.

129. Hu MC WM, Stroop SD , Mark AR, Beall B and Dale JB Immunogenicity of a 26-valent Group A Streptococcal Vaccine. *Journal of Infection and Immunity* **2002**;70:2171-7.

130. Fischetti VAKF Jk, Manjula BN and Scott JR Streptococcal M6 protein expressed in Escherichia coli. Localization, purification and comparison with streptococcal-derived M protein. *Journal of Experimental medicine* **1984**;159:1083-95.

131. Massell BF, Honikman LH, Amezcua J. Rheumatic fever following streptococcal vaccination. Report of three cases. JAMA **1969** Feb 10;207(6):1115-9.

132. Massell BF, Michael JG, Amezcua J, Siner M. Secondary and apparent primary antibody responses after group A streptococcal vaccination of 21 children. Appl Microbiol **1968** Mar;16(3):509-18.

133. Dale JB, Beachey EH. Epitopes of streptococcal M proteins shared with cardiac myosin. J Exp Med **1985** Aug 1;162(2):583-91.

134. Bronze MS, Dale JB. Epitopes of streptococcal M proteins that evoke antibodies that cross-react with human brain. J Immunol **1993** Sep 1;151(5):2820-8.

135. Stollerman GH. Prospects for a vaccine against group A streptococci: the problem of the immunology of M proteins. Arthritis Rheum **1967** Jun;10(3):245-55.

136. Dale JB, Seyer JM, Beachey EH. Type-specific immunogenicity of a chemically synthesized peptide fragment of type 5 streptococcal M protein. J Exp Med **1983** Nov 1;158(5):1727-32.

137. Dale JB, Chiang EY, Lederer JW. Recombinant tetravalent group A streptococcal M protein vaccine. J Immunol **1993** Aug 15;151(4):2188-94.

138. Dale JB, Simmons M, Chiang EC, Chiang EY. Recombinant, octavalent group A streptococcal M protein vaccine. Vaccine **1996** Jul;14(10):944-8.

139. Dale JB. Multivalent group A streptococcal vaccine designed to optimize the immunogenicity of six tandem M protein fragments. Vaccine **1999** Jan;17(2):193-200.

140. 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** Sep 12.

141. Hu MC, Walls MA, Stroop SD, Reddish MA, Beall B, Dale JB. Immunogenicity of a 26-valent group A streptococcal vaccine. Infect Immun **2002** Apr;70(4):2171-7.

142. Shulman ST, Tanz RR, Dale JB, et al. Seven-year surveillance of north american pediatric group a streptococcal pharyngitis isolates. Clin Infect Dis **2009** Jul 1;49(1):78-84.

143. Smeesters PR, McMillan DJ, Sriprakash KS. The streptococcal M protein: a highly versatile molecule. Trends Microbiol **2010** Jun;18(6):275-82.

144. Batzloff MR, Hayman WA, Davies MR, 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** May 15;187(10):1598-608.

145. Vohra H, Dey N, Gupta S, et al. M protein conserved region antibodies opsonise multiple strains of Streptococcus pyogenes with sequence variations in C-repeats. Res Microbiol **2005** May;156(4):575-82.

146. Batzloff MR, Hartas J, Zeng W, Jackson DC, Good MF. Intranasal vaccination with a lipopeptide containing a conformationally constrained conserved minimal peptide, a universal T cell epitope, and a self-adjuvanting lipid protects mice from group A streptococcus challenge and reduces throat colonization. J Infect Dis **2006** Aug 1;194(3):325-30.

147. Brandt ER, Sriprakash KS, Hobb RI, et al. New multi-determinant strategy for a group A streptococcal vaccine designed for the Australian Aboriginal population. Nat Med **2000** Apr;6(4):455-9.

148. Abbot EL, Smith WD, Siou GP, et al. Pili mediate specific adhesion of Streptococcus pyogenes to human tonsil and skin. Cell Microbiol **2007** Jul;9(7):1822-33.

149. Margarit I, Rinaudo CD, Galeotti CL, et al. Preventing bacterial infections with pilus-based vaccines: the group B streptococcus paradigm. J Infect Dis **2009** Jan 1;199(1):108-15.

150. Mora M, Bensi G, Capo S, et al. Group A Streptococcus produce pilus-like structures containing protective antigens and Lancefield T antigens. Proc Natl Acad Sci U S A **2005** Oct 25;102(43):15641-6.

151. Falugi F, Zingaretti C, Pinto V, et al. Sequence variation in group A Streptococcus pili and association of pilus backbone types with lancefield T serotypes. J Infect Dis **2008** Dec 15;198(12):1834-41.

152. Buccato S, Maione D, Rinaudo CD, et al. Use of Lactococcus lactis expressing pili from group B Streptococcus as a broad-coverage vaccine against streptococcal disease. J Infect Dis **2006** Aug 1;194(3):331-40.

153. Wexler DE, Chenoweth DE, Cleary PP. Mechanism of action of the group A streptococcal C5a inactivator. Proc Natl Acad Sci U S A **1985** Dec;82(23):8144-8.

154. Chmouryguina I, Suvorov A, Ferrieri P, Cleary PP. Conservation of the C5a peptidase genes in group A and B streptococci. Infect Immun **1996** Jul;64(7):2387-90.

155. Ji Y, Carlson B, Kondagunta A, Cleary PP. Intranasal immunization with C5a peptidase prevents nasopharyngeal colonization of mice by the group A Streptococcus. Infect Immun **1997** Jun;65(6):2080-7.

156. Cleary PP, Matsuka YV, Huynh T, Lam H, Olmsted SB. Immunization with C5a peptidase from either group A or B streptococci enhances clearance of group A streptococci from intranasally infected mice. Vaccine **2004** Oct 22;22(31-32):4332-41.

157. Bates CS, Montanez GE, Woods CR, Vincent RM, Eichenbaum Z. Identification and characterization of a Streptococcus pyogenes operon involved in binding of hemoproteins and acquisition of iron. Infect Immun **2003** Mar;71(3):1042-55.

158. Fisher M, Huang YS, Li X, McIver KS, Toukoki C, Eichenbaum Z. Shr is a broad-spectrum surface receptor that contributes to adherence and virulence in group A streptococcus. Infect Immun **2008** Nov;76(11):5006-15.

159. Huang YS, Fisher M, Nasrawi Z, Eichenbaum Z. Defense from the Group A Streptococcus by active and passive vaccination with the streptococcal hemoprotein receptor. J Infect Dis **2011** Jun 1;203(11):1595-601.

160. Garapetis JR, Steer AC, Mulholland E, K, M W. The global burden of group A streptococcal disease. *Lancet Infectious Disease* **2005**;5:685-94.

161. Robertson KA, Mayosi BM. Rheumatic heart disease: social and economic dimensions. S Afr Med J **2008** Oct;98(10):780-1.

162. Dale JB aBE. Localization of protective epitopes of the amino terminus of type 5 streptococcal M protein. *Journal of Experimental Medicine* **1986**;163:1191-202.

163. Bauer MJ, Georgousakis MM, Vu T, et al. Evaluation of novel Streptococcus pyogenes vaccine candidates incorporating multiple conserved sequences from the C-repeat region of the M-protein. Vaccine **2012** Mar 9;30(12):2197-205.

164. Dale JB, Penfound T, Chiang EY, Long V, Shulman ST, Beall B. Multivalent group A streptococcal vaccine elicits bactericidal antibodies against variant M subtypes. Clin Diagn Lab Immunol **2005** Jul;12(7):833-6.

165. CDC. emm Types as Proportions of Total Disease Isolates in Six Global Regions. 2011 [cited 2011 16 May 2011]; Available from: http://www.cdc.gov/ncidod/biotech/strep/emmtype_proportions.htm

166. Brandt CM, Spellerberg B, Honscha M, Truong ND, Hoevener B, Lutticken R. Typing of Streptococcus pyogenes strains isolated from throat infections in the region of Aachen, Germany. Infection **2001** May-Jun;29(3):163-5.

167. Gieseker KE, Mackenzie T, Roe MH, Todd JK. Comparison of two rapid Streptococcus pyogenes diagnostic tests with a rigorous culture standard. Pediatr Infect Dis J **2002** Oct;21(10):922-7.

168. CDC. Protocol for emm typing. 2011 [cited 2011 16 May]; CDCwebsite]. Available from: http://www.cdc.gov/ncidod/biotech/strep/protocol_emm-type.htm

169. Saiki RK, Gelfand DH, Stoffel S, et al. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **1988** Jan 29;239(4839):487-91.

170. Stellwagen NC. Electrophoresis of DNA in agarose gels, polyacrylamide gels and in free solution. Electrophoresis **2009** Jun;30 Suppl 1:S188-95.

171. Darenberg J, Luca-Harari B, Jasir A, et al. Molecular and clinical characteristics of invasive group A streptococcal infection in Sweden. Clin Infect Dis **2007** Aug 15;45(4):450-8.

172. 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 Oct 26;29(46):8175-8.

173. Gita S, Suneeta M, Anjana S, Niranjan N, Sujata M, Pandey RM. C. trachomatis in female reproductive tract infections and RFLP-based genotyping: a 16-year study from a tertiary care hospital. Infect Dis Obstet Gynecol;2011:548219.

174. Mylvaganam H, Bruun T, Vindenes HA, Langeland N, Skrede S. Molecular epidemiological investigation of an outbreak of invasive beta-haemolytic streptococcal infection in western Norway. Clin Microbiol Infect **2009** Mar;15(3):245-52.

175. Desai M, Tanna A, Efstratiou A, George R, Clewley J, Stanley J. Extensive genetic diversity among clinical isolates of Streptococcus pyogenes serotype M5. Microbiology **1998** Mar;144 (Pt 3):629-37.

176. Tsai HJ, Huang HC, Tsai HL, Chang CC. PCR-based restriction fragment length polymorphism (RFLP) analysis of Campylobacter jejuni isolates from humans, chickens and dogs in northern Taiwan. J Vet Med Sci **2006** Aug;68(8):815-9.

177. Agudo S, Perez-Perez G, Alarcon T, Lopez-Brea M. Rapid detection of clarithromycin resistant Helicobacter pylori strains in Spanish patients by polymerase chain reaction-restriction fragment length polymorphism. Rev Esp Quimioter **2011** Mar;24(1):32-6.

178. BioLabs NE. NEBcutter V2.0. 2011 [cited 2011 5th July]; Available from: http://tools.neb.com/NEBcutter2/index.php

179. Bannerman TL, Hancock GA, Tenover FC, Miller JM. Pulsed-field gel electrophoresis as a replacement for bacteriophage typing of Staphylococcus aureus. J Clin Microbiol **1995** Mar;33(3):551-5.

180. Smeesters PR, Mardulyn P, Vergison A, Leplae R, Van Melderen L. Genetic diversity of Group A Streptococcus M protein: implications for typing and vaccine development. Vaccine **2008** Oct 29;26(46):5835-42.

181. Yoonim N, Olive C, Pruksachatkunakorn C, Pruksakorn S. Bactericidal activity of M protein conserved region antibodies against group A streptococcal isolates from the Northern Thai population. BMC Microbiol **2006**;6:71.

182. Olive C, Clair T, Yarwood P, Good MF. Protection of mice from group A streptococcal infection by intranasal immunisation with a peptide vaccine that contains a conserved M protein B cell epitope and lacks a T cell autoepitope. Vaccine **2002** Jun 21;20(21-22):2816-25.

183. Bailey L Thimothy CE. "Fitting a mixture model by expectation maximization to discover motifs in biopolymers": AAAI Press, Menlo Park, California, **1994**.

184. Penfound TA, Chiang EY, Ahmed EA, Dale JB. Protective efficacy of group A streptococcal vaccines containing type-specific and conserved M protein epitopes. Vaccine **2010** Jul 12;28(31):5017-22.

185. Proft T, Herrmann R. Identification and characterization of hitherto unknown Mycoplasma pneumoniae proteins. Mol Microbiol **1994** Jul;13(2):337-48.

186. Henegariu O, Heerema NA, Dlouhy SR, Vance GH, Vogt PH. Multiplex PCR: critical parameters and step-by-step protocol. Biotechniques **1997** Sep;23(3):504-11.

187. Markoulatos P, Siafakas N, Moncany M. Multiplex polymerase chain reaction: a practical approach. J Clin Lab Anal **2002**;16(1):47-51.

188. Chamberlain JS, Gibbs RA, Ranier JE, Nguyen PN, Caskey CT. Deletion screening of the Duchenne muscular dystrophy locus via multiplex DNA amplification. Nucleic Acids Res **1988** Dec 9;16(23):11141-56.

189. Markoulatos P, Samara V, Siafakas N, Plakokefalos E, Spyrou N, Moncany ML. Development of a quadriplex polymerase chain reaction for human cytomegalovirus detection. J Clin Lab Anal **1999**;13(3):99-105.

190. Bisno AL. The resurgence of acute rheumatic fever in the United States. Annu Rev Med **1990**;41:319-29.

191. Proft T, Fraser JD. Streptococcal superantigens. Chem Immunol Allergy **2007**;93:1-23.

Appendix A

50X stock TAE buffer

242 g Tris base

27.1 ml glacial acetic acid

37.2 g $Na_2EDTA\ 2H_2O$

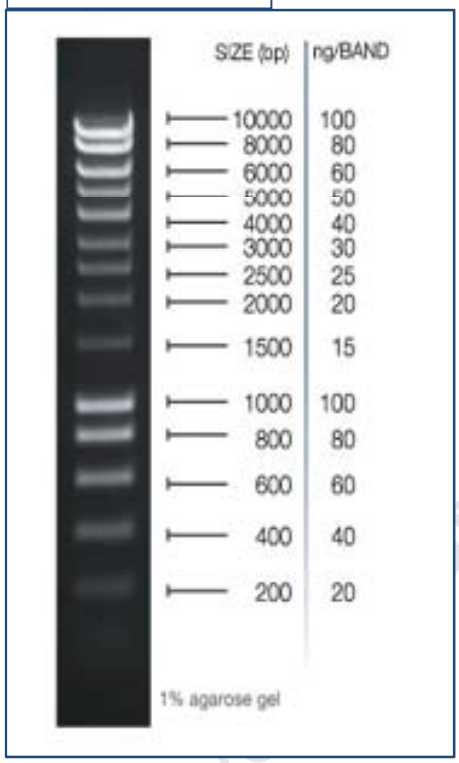
pH 8.5

Prepare to 1 litre with distilled water

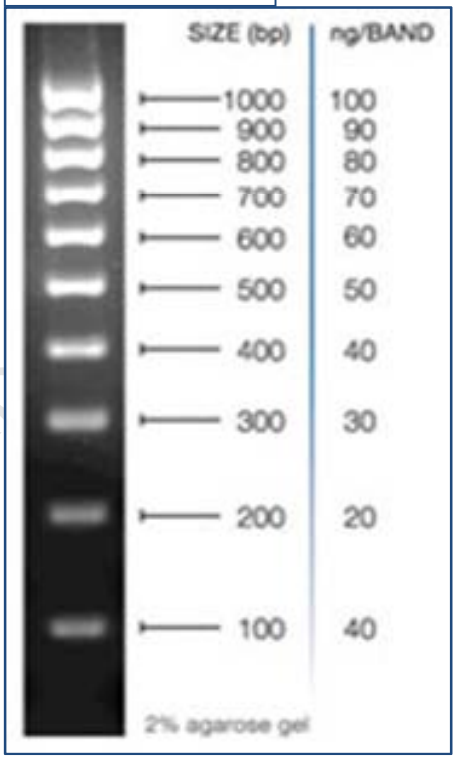
Appendix B

Molecular markers used in agarose gel electrophoresis

HyperLadder I



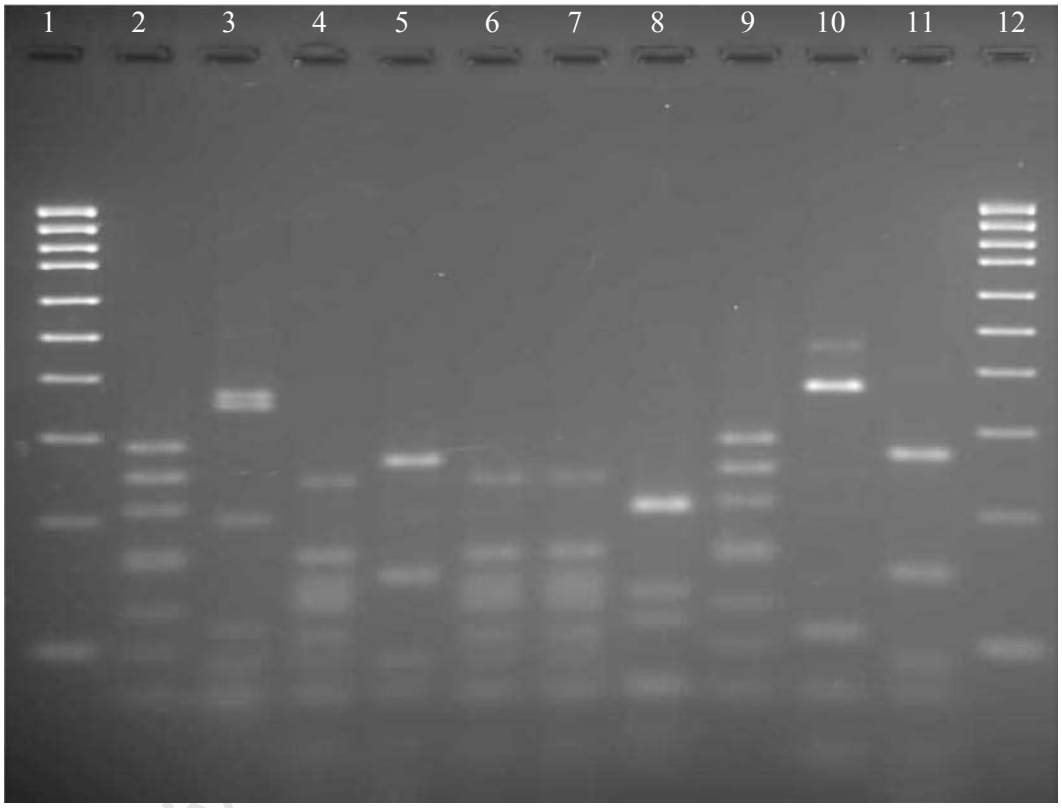
HyperLadder IV



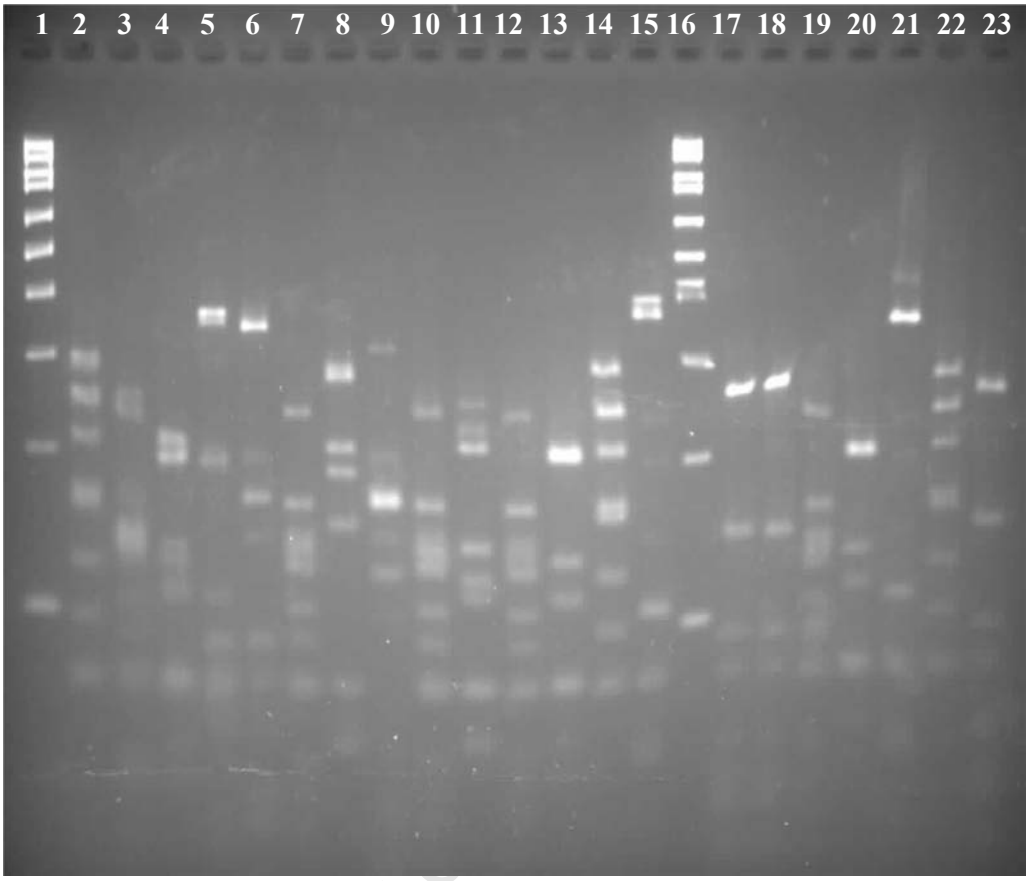
Appendix C

*Dde*I digests

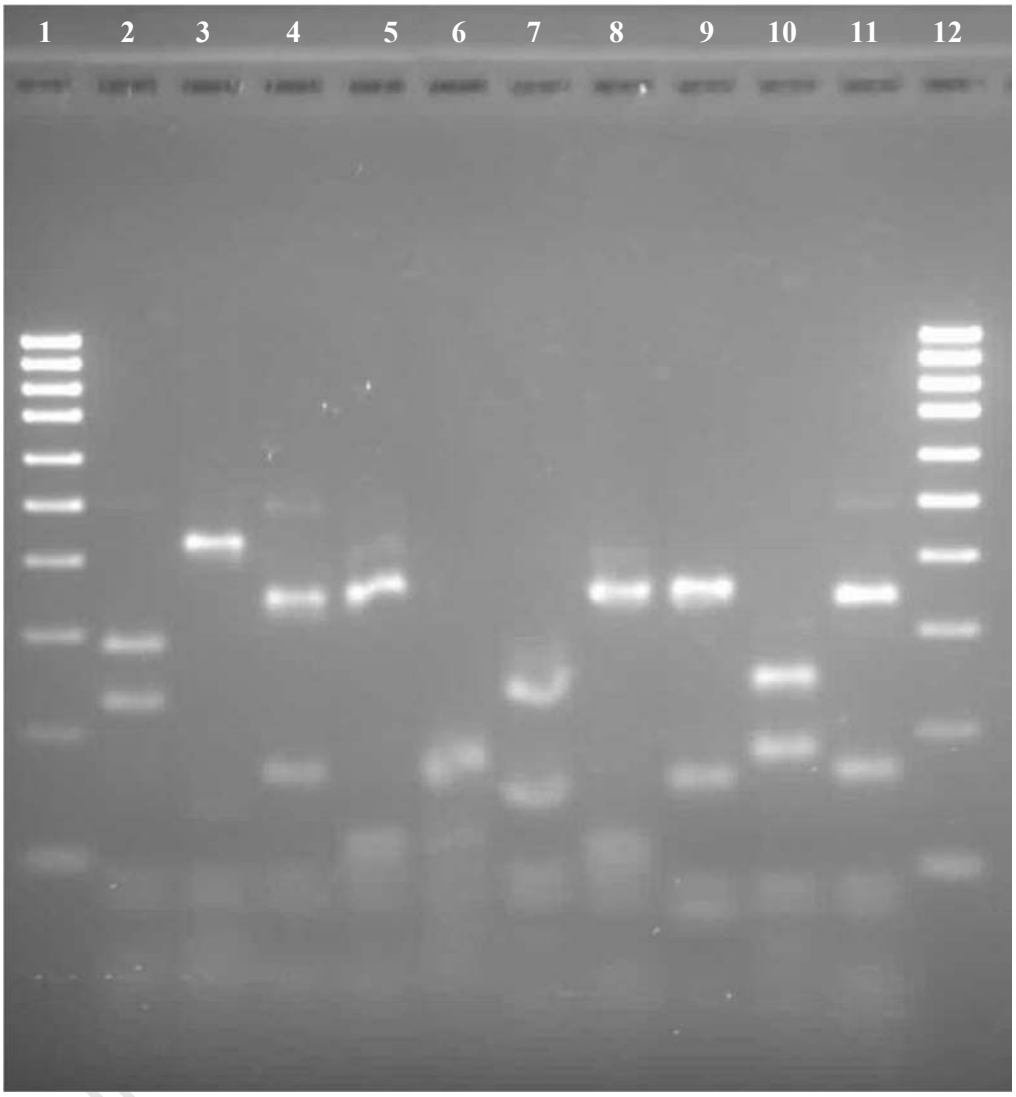
Restriction pattern of representative *emm* types following digestion with *Dde*I, separated on a 3% agarose gel electrophoresis for 6 hours



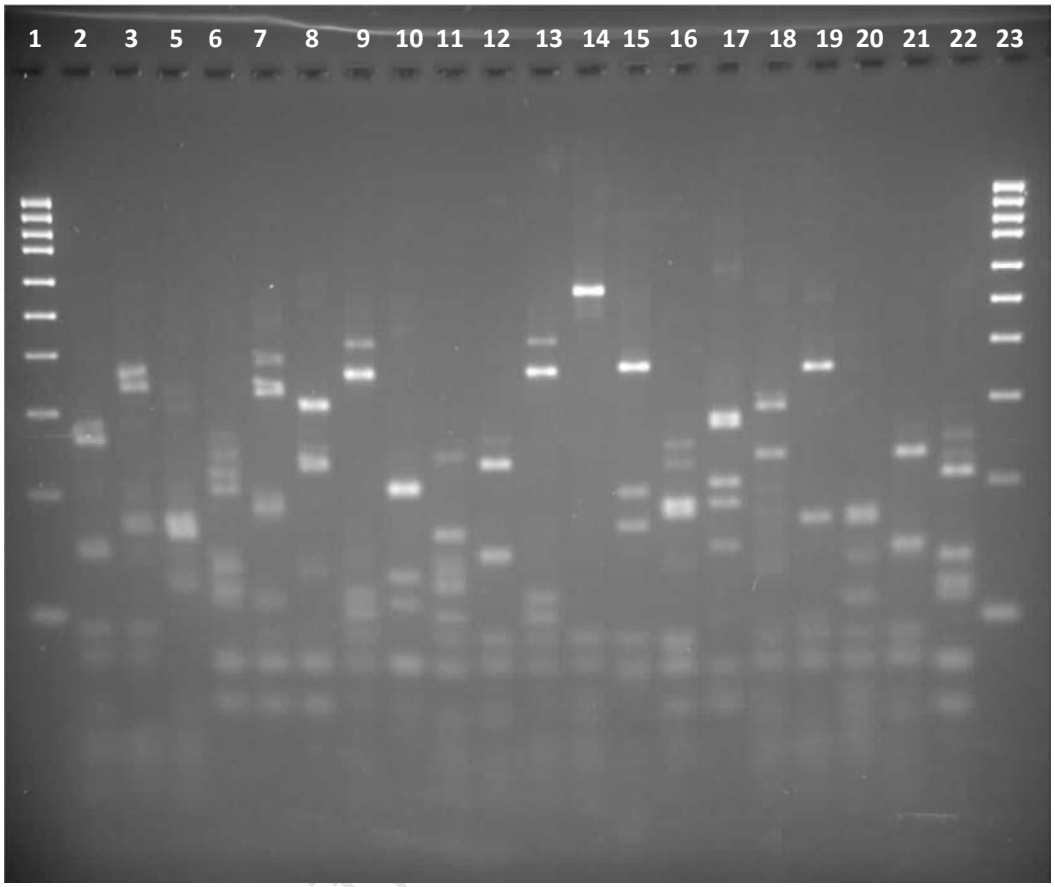
Lane 1 and 12 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm12*, Lane 3 *emm28*, Lane 4 *emm94*, Lane 5 *emm89*, Lane 6 *emm94*, Lane 7 *emm94*, Lane 8 *emm6.4*, Lane 9 *emm12*, Lane 10 *emm4* , Lane 11 *emm89*



Lane 1 and 16 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm12*, Lane 3 *st2002.2*, Lane 4 *emm116.1*, Lane 5 *emm28*, Lane 6 *emm87*, Lane 7 *emm94*, Lane 8 *emm1*, Lane 9 *emm89*, Lane 10 *emm94*, Lane 11 *emm53*, Lane 12 *emm94* , Lane 13 *emm6.4*, Lane 14 *emm12*, Lane 15 *emm4*, Lane 17 *emm89*, Lane 18 *emm89*, Lane 19 *emm53*, Lane 20 *emm116*, Lane 21 *emm 4*, Lane 22 *emm12*, Lane 23 *emm89*



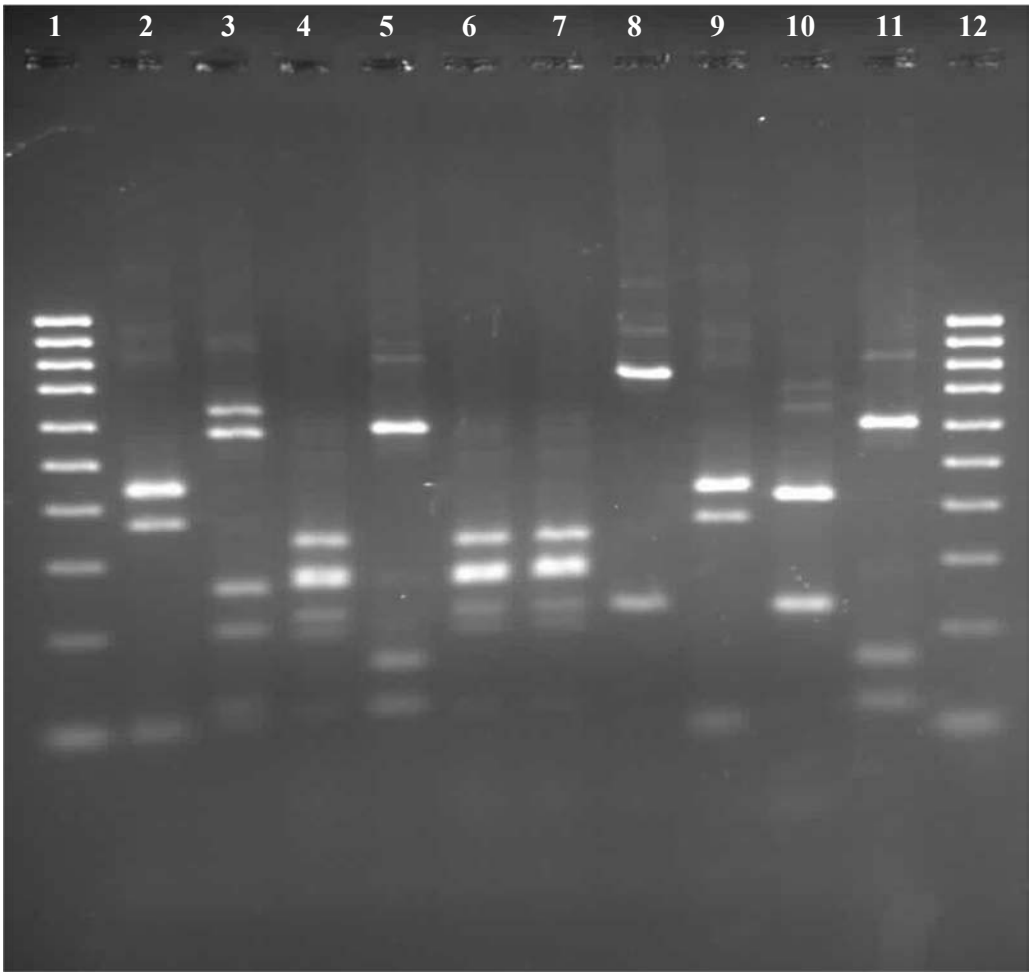
Lane 1 and 12 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm8*, Lane 3 *emm118*, Lane 4 *emm58*, Lane 5 *emm4*, Lane 6 *emm82*, Lane 7 *emm77*, Lane 8 *emm49*, Lane 9, Lane 10 *emm9*, Lane 11 *emm116*



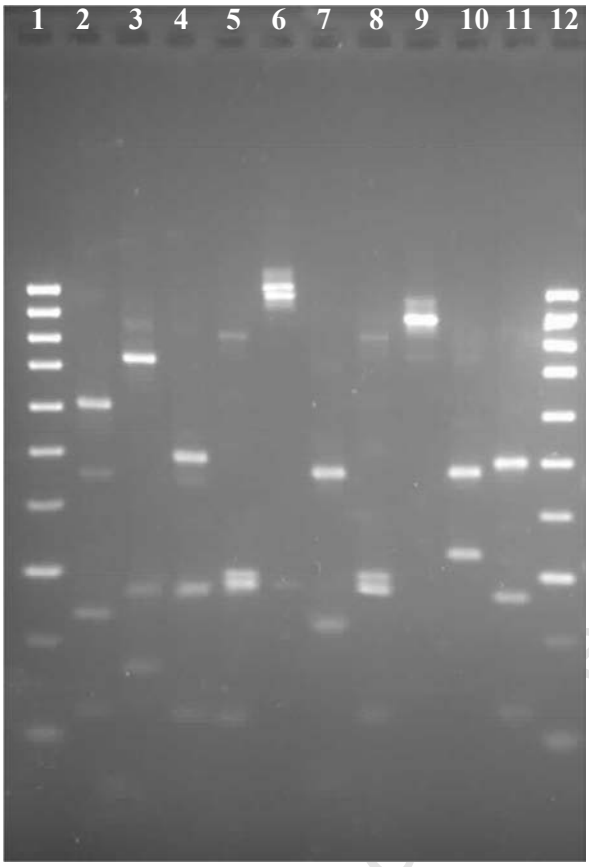
Lane 1 and 23 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm89*, Lane 3 *emm87*, Lane 4 *emm44*, Lane 5 *emm22*, Lane 6 *emm80*, Lane 7 *emm33*, Lane 8 *emm48.1*, Lane 9 *emm6.4*, Lane 10 *emm48.1*, Lane 11 *emm77*, Lane 12 *emm48.1*, Lane 13 *emm92*, Lane 14, Lane 15 *emm116.1*, Lane 16 *emm1*, Lane 17 *emm8*, Lane 18 *emm58.2*, Lane 19 *emm82*, Lane 20 *emm77*, Lane 21 *emm53*

*Mbo*I digests

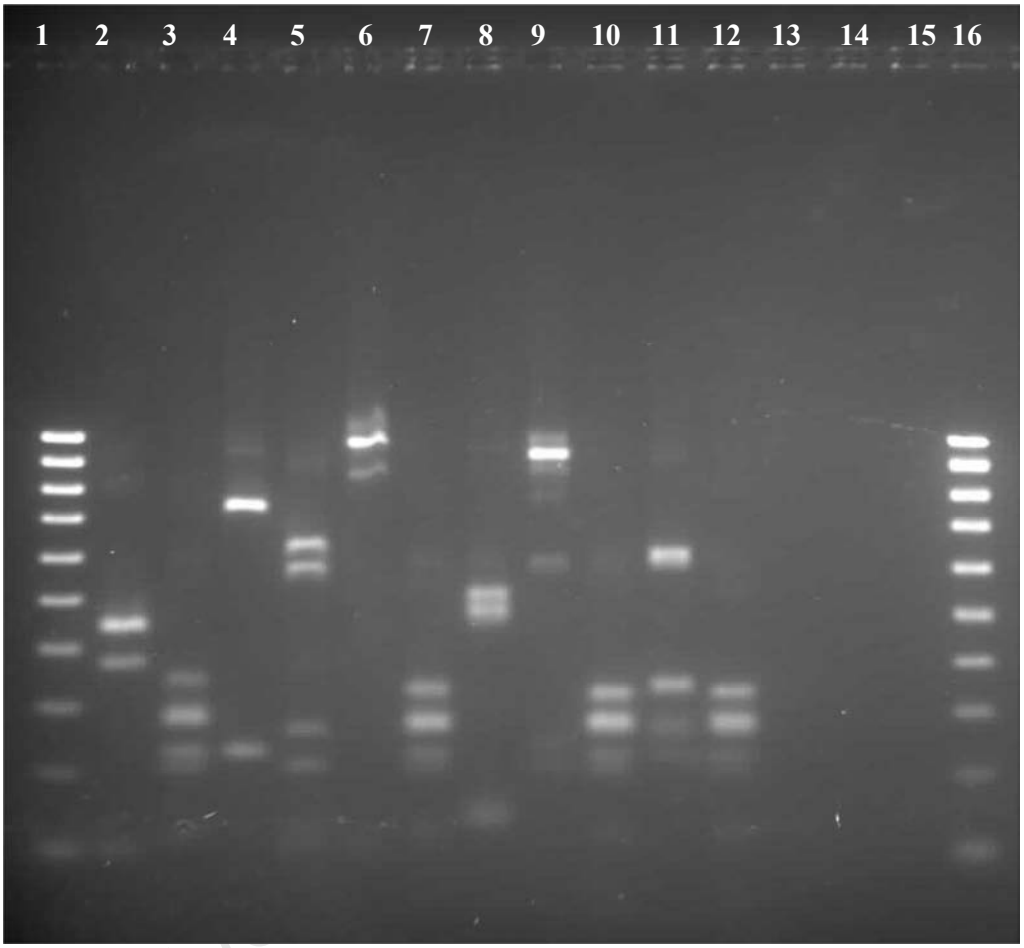
Restriction pattern of representative *emm* types following digestion with *Mbo*I, separated on a 3% agarose gel electrophoresis for 6 hours



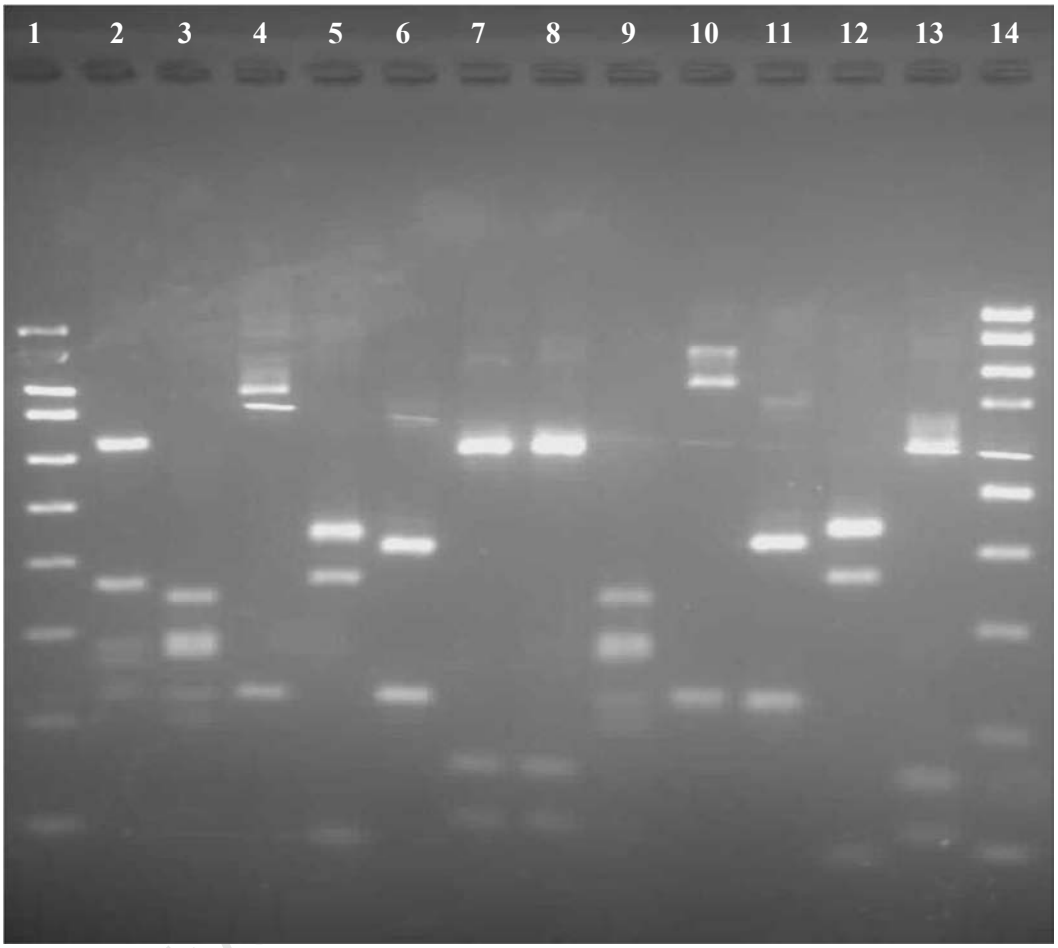
Lane 1 and 12 are molecular markers (HyperLadder IV, Appendix A). Lane 1 and 20 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm12*, Lane 3 *emm28*, Lane 4 *emm94*, Lane 5 *emm89*, Lane 6 *emm94*, Lane 7 *emm94*, Lane 8 *emm6.4*, Lane 9 *emm12*, Lane 10 *emm4* , Lane 11 *emm89*



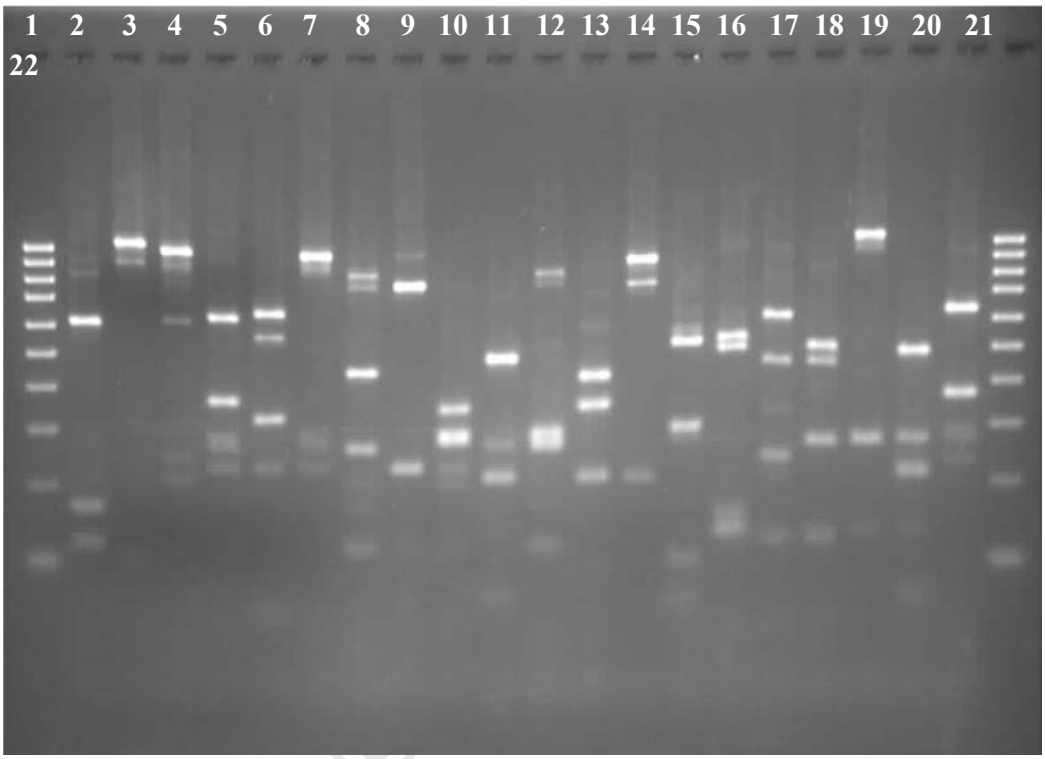
Lane 1 and 12 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm8*, Lane 3 *emm118*, Lane 4 *emm58*, Lane 5 *emm4*, Lane 6 *emm82*, Lane 7 *emm77*, Lane 8 *emm49*, Lane 9, Lane 10 *emm9*, Lane 11 *emm116*



Lane 1 and 16 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm12*, Lane 3 *st2002.2*, Lane 4 *emm116*, Lane 5 *emm28*, Lane 6 *emm97*, Lane 7 *emm94*, Lane 8 *emm1*, Lane 9 *emm89*, Lane 10 *emm94*, Lane 11 *emm22*, Lane 12



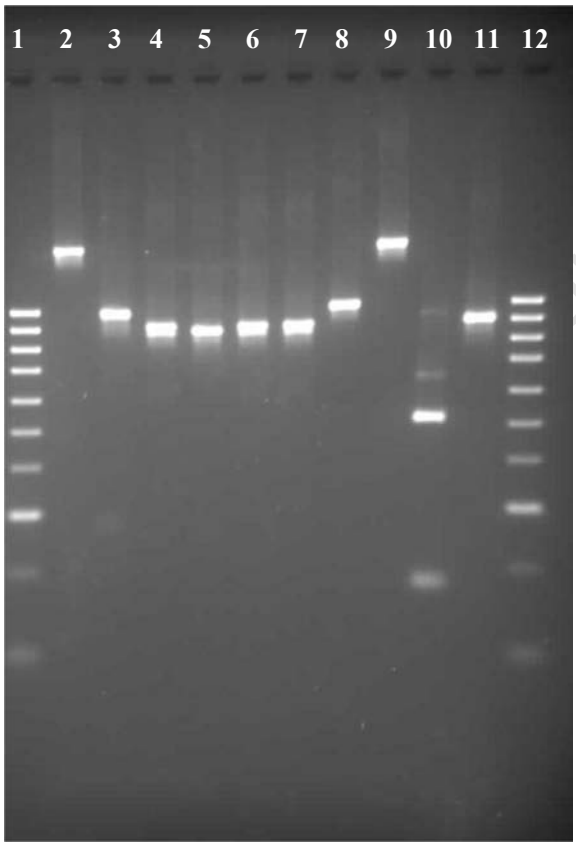
Lane 1 and 14 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm53*, Lane 3 *emm94*, Lane 4 *emm6.4*, Lane 5 *emm12*, Lane 6 *emm4*, Lane 7 *emm89*, Lane 8 *emm89*, Lane 9 *emm94*, Lane 10 *emm116*, Lane 11 *emm4*, Lane 12 *emm12*, Lane 13 *emm89*



Lane 1 and 22 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm89*, Lane 3 *emm87*, Lane 4 *emm44*, Lane 5 *emm22*, Lane 6 *emm80*, Lane 7 *emm33*, Lane 8 *emm48.1*, Lane 9 *emm6.4*, Lane 10 *emm48.1*, Lane 11 *emm77*, Lane 12 *emm48.1*, Lane 13 *emm92*, Lane 14, Lane 15 *emm116.1*, Lane 16 *emm1*, Lane 17 *emm8*, Lane 18 *emm58.2*, Lane 19 *emm82*, Lane 20 *emm77*, Lane 21 *emm53*

*Sac*I digests

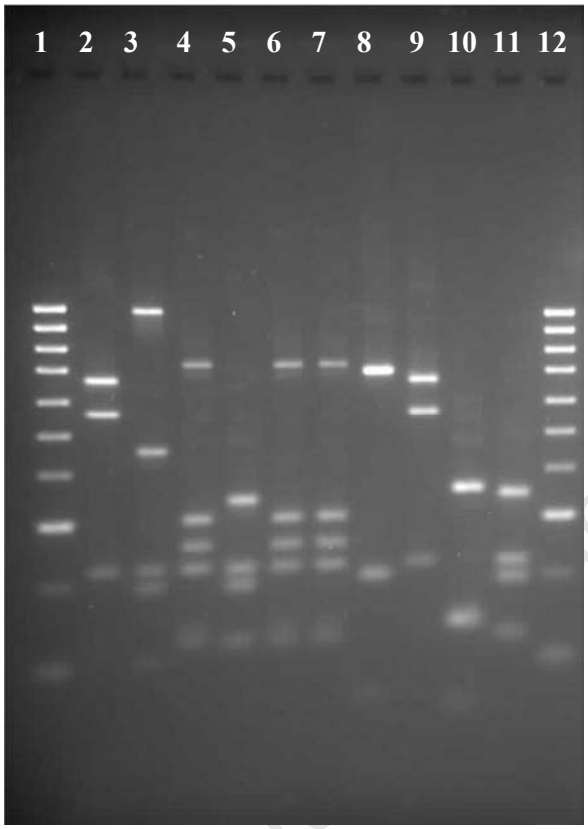
Restriction pattern of representative *emm* types following digestion with *Sac*I, separated on a 3% agarose gel electrophoresis for 6 hours



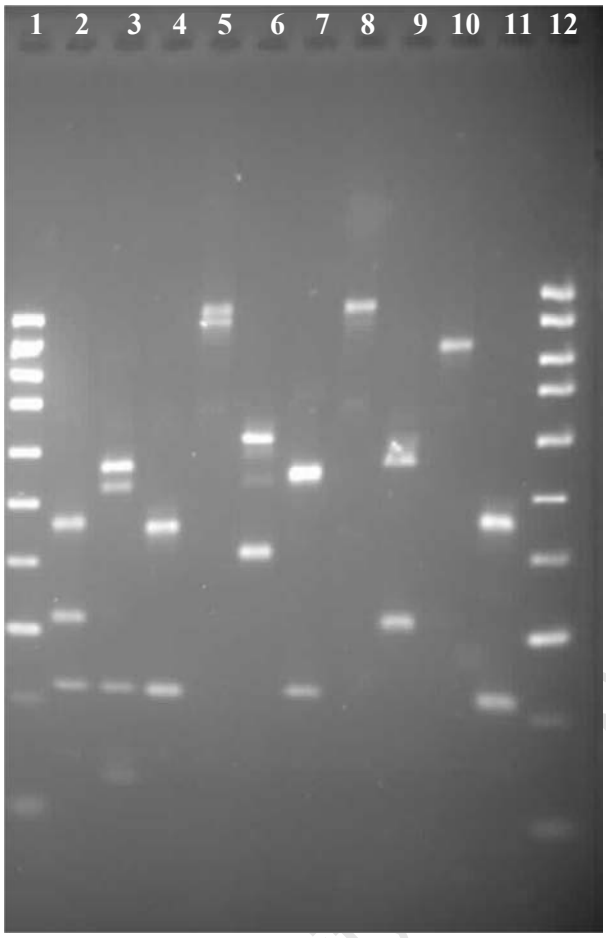
Lane 1 and 12 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm12*, Lane 3 *emm28*, Lane 4 *emm94*, Lane 5 *emm89*, Lane 6 *emm94*, Lane 7 *emm94*, Lane 8 *emm6.4*, Lane 9 *emm12*, Lane 10 *emm4* , Lane 11 *emm89*

*Hae*III and *Hinc*II digests

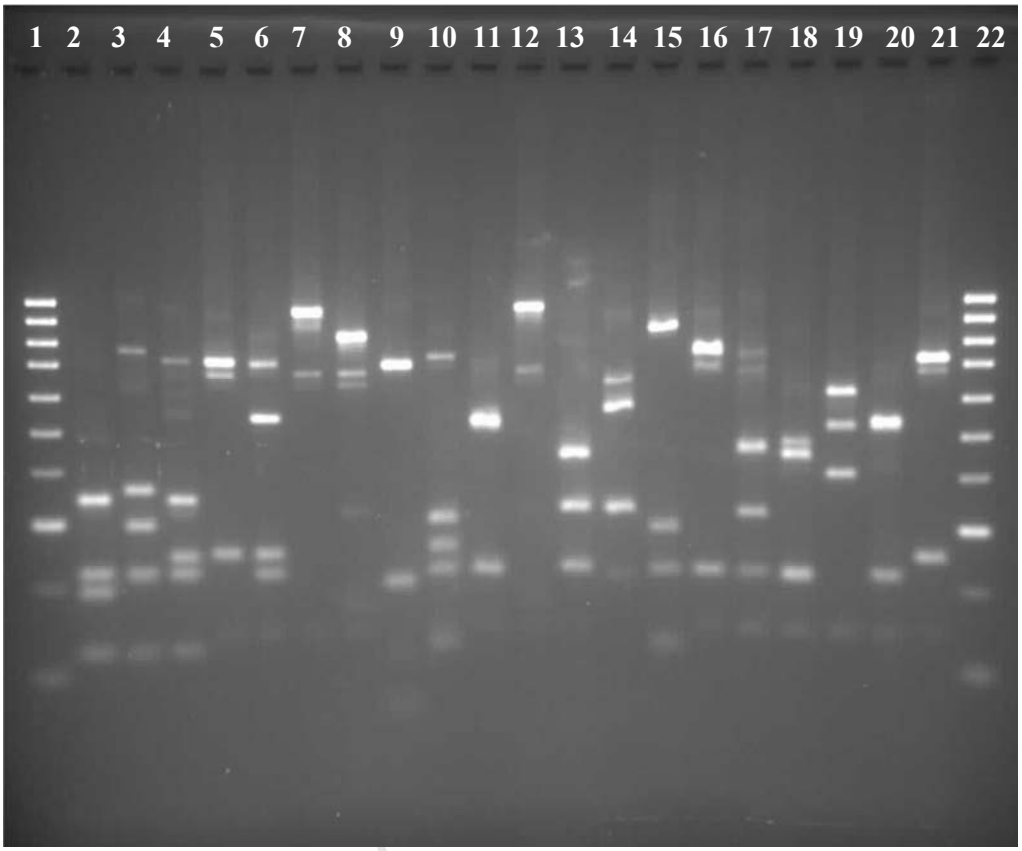
Restriction pattern of representative *emm* types following digestion with *Hae*III and *Hinc*II, separated on a 3% agarose gel electrophoresis for 6 hours



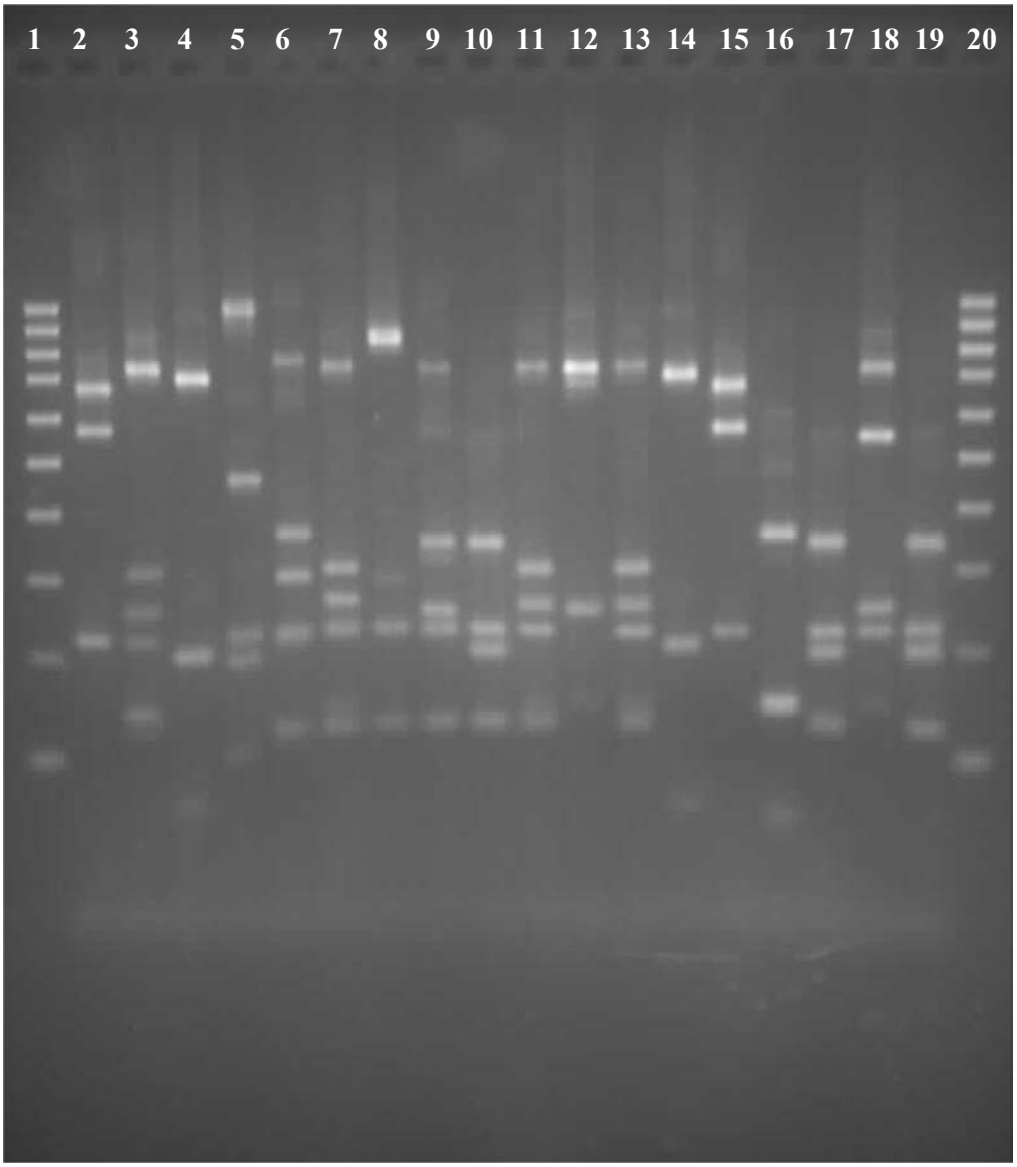
Lane 1 and 12 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm12*, Lane 3 *emm28*, Lane 4 *emm94*, Lane 5 *emm89*, Lane 6 *emm94*, Lane 7 *emm94*, Lane 8 *emm6.4*, Lane 9 *emm12*, Lane 10 *emm4* , Lane 11 *emm89*



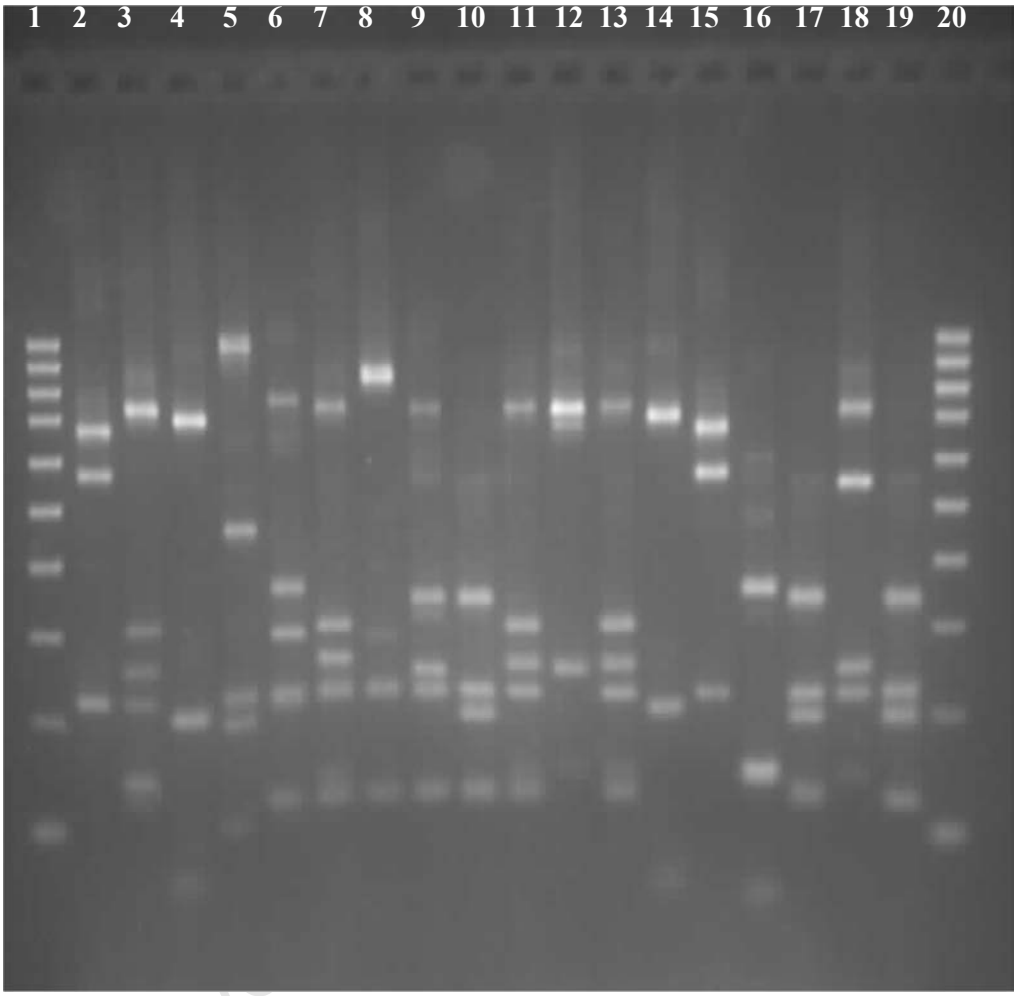
Lane 1 and 12 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm8*, Lane 3 *emm118*, Lane 4 *emm58*, Lane 5 *emm4*, Lane 6 *emm82*, Lane 7 *emm77*, Lane 8 *emm49*, Lane 9, Lane 10 *emm9*, Lane 11 *emm116*



Lane 1 and 22 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm89*, Lane 3 *emm87*, Lane 4 *emm44*, Lane 5 *emm22*, Lane 6 *emm80*, Lane 7 *emm33*, Lane 8 *emm48*.1, Lane 9 *emm6.4*, Lane 10 *emm48.1*, Lane 11 *emm77*, Lane 12 *emm48.1*, Lane 13 *emm92*, Lane 14, Lane 15 *emm1*, Lane 16 *emm116.1*, Lane 17 *emm8*, Lane 18 *emm58.2*, Lane 19 *emm82*, Lane 20 i, Lane 21 *emm53.*



Lane 1 and 20 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm12*, Lane 3 *st2002.2*, Lane 4 *emm116*, Lane 5 *emm28*, Lane 6 *emm87*, Lane 7 *emm94*, Lane 8 *emm1*, Lane 9 *emm44*, Lane 10 *emm89*, Lane 11 *emm94*, Lane 12 *emm22*, Lane 13 *emm94*, Lane 14 *emm6.4*, Lane 15 *emm12*, Lane 16 *emm4*, Lane 17 *emm89*, Lane 18 *emm80* and Lane 19 *emm89*.



Lane 1 and 20 are molecular markers (HyperLadder IV, Appendix A). Lane 2 *emm12*, Lane 3 *st2002.2*, Lane 4 *emm116*, Lane 5 *emm28*, Lane 6 *emm87*, Lane 7 *emm94*, Lane 8 *emm1*, Lane 9 *emm44*, Lane 10 *emm89*, Lane 11 *emm94*, Lane 12 *emm22*, Lane 13 *emm94*, Lane 14 *emm6.4*, Lane 15 *emm12*, Lane 16 *emm4*, Lane 17 *emm89*, Lane 18 *emm80* and Lane 19 *emm89*.

Appendix D

Amino acid (single letter and representative codons)

Amino acid	Symbol	Codons
Isoleucine	I	ATT, ATC, ATA
Leucine	L	CTT, CTC, CTA, CTG, TTA, TTG
Valine	V	GTT, GTC, GTA, GTG
Phenylalanine	F	TTT, TTC
Methionine	M	ATG
Cysteine	C	TGT, TGC
Alanine	A	GCT, GCC, GCA, GCG
Glycine	G	GGT, GGC, GGA, GGG
Proline	P	CCT, CCC, CCA, CCG
Threonine	T	ACT, ACC, ACA, ACG
Serine	S	TCT, TCC, TCA, TCG, AGT, AGC
Tyrosine	Y	TAT, TAC
Tryptophan	W	TGG
Glutamine	Q	CAA, CAG
Asparagine	N	AAT, AAC
Histidine	H	CAT, CAC
Glutamic acid	E	GAA, GAG
Aspartic acid	D	GAT, GAC
Lysine	K	AAA, AAG
Arginine	R	CGT, CGC, CGA, CGG, AGA, AGG
Stop codons	Stop	TAA

Alignment of sequences generated from the C-region of the M-protein

Sequence alignment of C1-repeat

Name	Sites	
emm89	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASRAAKKDLEAEHQKLKE
emm118	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASRAAKKDLEAEHQKLKE
emm15		QLETEKQISEASRKSLSRDLEASRAAKKDLEAEHQKLKE
emm82	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASRAAKKDLEAEHQKLKE
emm118	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASRAAKKDLEAEHQKLKE
emm89	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASRAAKKDLEAEHQKLKE
emm58	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASRAAKKDLEAEHQKLKE
emm8	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASRAAKKELEAEHQKLKE
emm44	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASRAAKKELEAEHQKLKE
emm77	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASRAAKKELEAEHQKLKE
emm44	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASRAAKKELEAEHQKLKE
emm8	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASRAAKKELEAEHQKLKE
emm22	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASRAAKKGLEAEHQKLKE
emm22	QKQQ	QLETEKQISEASRKSLSRDLEASRAAKKKVEADLAALNA
emm82	KDTLAEKAK	KLEEDKQISDASRKSLSRDLEASRAAKKELEAKXQKLET
emm49	KAK	KLEEDKQISDASRKSLSRDLEASRAAKKELEANHQKLET
emm94	EKDTLAEKAK	KLEEDKQISDASRKSLSRDLEASRAAKKELEAKHQKLET
emm4	AK	KLEEDKQISDASRKSLSRDLEASRAAKKELEANHQKLET
emm94	EKDTLAEKAK	KLEEDKQISDASRKSLSRDLEASRAAKKELEAKHQKLET
emm48	EKDTLAEKAK	KLEEDKQISDASRKSLSRDLEASRAAKKELEANHQKLET
emm48	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASREAKKKVEADLAALTA
emm49	RYQEQLQKQQ	QLETEKQISEASRKSLSRDLEASREAKKKVEADLAALTA
emm92	RYQEQLQKQQ	QLEKEKQISEASRKSLSRDLEASRAAKKDLEAEHQKLKE
emm92	RYQEQLQKQQ	QLEKEKQISEASRKSLSRDLEASRAAKKDLEAEHQKLKE
emm1	EKEQLTIEKA	KLEEEKQISDASRQSLRRDLDASREAKKQVEKDLANLTA
emm1	EKA	KLEEEKQISDASRQSLRRDLDASREAKKQVEKDLANLTA
emm116	ERKEKEAEQK	KLKEEKQISDASRQGLRRDLDASREAKKQVEKDLANLTT
emm116	ERKEKEAEQK	KLKEEKQISDASRQGLRRDLDASREAKKQVEKDLANLTT
emm116	ERKEKEAEQK	KLKEEKQISDASRQGLRRDLDASREAKKQVEKDLANLTT
emm80	DLANLTAELG	KVKEEKQISDASRQGLRRDLDASREAKKQVEKDLANLTA
emm53	DLANLTAELD	KVKEEKQISDASRQGLRRDLDASREAKKQVEKGLANLTA
emm53	DLANLTAELD	KVKEEKQISDASRQGLRRDLDASREAKKQVEKGLANLTA
emm22	Q	KLAKDKQISDASRQGLSRDLEASRAAKKELEAKHQKLEA
emm4	KHQKLEAENK	KLTEANQVSEASRKGLSNDLEASRAAKKELEAKHQKLEA
emm4	KHQKLEAENK	KLTEANQVSEASRKGLSNDLEASRAAKKELEAKHQKLEA
emm48	KHQKLEAENK	KLTEANQVSEASRKGLSNDLEASRAAKKELEAKHQKLEA
emm4	KHQKLEAENK	KLTEANQVSEASRKGLSNDLEASRAAKKELEAKHQKLEA

Sequence alignment of C2-repreat

Name	Sites		
emm49	KELEANHQKL	ETEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm4	KELEANHQKL	ETEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm48	KELEANHQKL	ETEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm48	KKVEADLAAL	TAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm49	KKVEADLAAL	TAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm94	KELEAKHQKL	ETEHQKLKEDKQISDASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm94	KELEAKHQKL	ETEHQKLKEDKQISDASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm89	EASRAAKKDL	EAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm118	EASRAAKKDL	EAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm14	EASRAAKKEL	EAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm89	EASRAAKKDL	EAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm44	EASRAAKKEL	EAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm8	EASRAAKKEL	EAEHQKLKEEKQISDASRKSLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm15	EASRAAKKDL	EAEHQKLKEEKQISDASRKSLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm22	EASRAAKKGL	EAEHQKLKEEKQISDASRKSLSRDLEASREAKKKVEADLAA	LNAEHQKLKE
emm92	EASRAAKKDL	EAEHQKLKEEKQISDASRKSLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm92	EASRAAKKDL	EAEHQKLKEEKQISDASRKSLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm58	EASRAAKKDL	EAEHQKLKEEKQISDASRKSLSRDLEASREAKKKVEADLAA	LNAEHQKLKE
emm8	EASRAAKKEL	EAEHQKLKEEKQISDASRKSLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm118	KQVEKDLANL	TTELDKVKEEKQISDASRQGLRRDLDASREAKKKVEADLAA	LTAEHQKLKE
emm116	KQVEKDLANL	TTELDKVKEEKQISDASRQGLRRDLDASREAKKKVEADLAA	LTAEHQKLKE
emm82	EASRAAKKDL	EAEHQKLKEEKQISDASRKSLSRDLEASREAKKKVEADLAT	LNAEHQKLKE
emm82	KELEAKXQKL	ETEXQKLKEDKQISDASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm118	EASRAAKKDL	EAEHQKLKEEKQISXASRQGLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm80	NEKEAERKEK	EAEQKKLKEEKQISDASRQGLRRDLDASREAKKQVEKDLAN	LTAELGKVKE
emm53	NGALRYINEK	EAEQKKLKEEKQISDASRQGLRRDLDASREAKKQVEKDLAN	LTAELDKVKE
emm53	RYINEK	EAEQKKLKEEKQISDASRQGLRRDLDASREAKKQVEKDLAN	LTAELDKVKE
emm1	KQVEKDLANL	TAELDKVKEDKQISDASRQGLRRDLDASREAKKQVEKDLAN	LTAELDKVKE
emm1	KQVEKDLANL	TAELDKVKEDKQISDASRQGLRRDLDASREAKKQVEKDLAN	LTAELDKVKE
emm9	REIDKRYQEQ	LQKQQQLETEKQISEASRKSLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm9	REIDKRYQEQ	LQKQQQLETEKQISEASRKSLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm9	REIDKRYQEQ	LQKQQQLETEKQISEASRKSLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm9	REIDKRYQEQ	LQKQQQLETEKQISEASRKSLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm9	REIDKRYQEQ	LQKQQQLETEKQISEASRKSLSRDLEASREAKKKVEADLAA	LTAEHQKLKE
emm22	REIDKRYQEQ	LQKQQQLETEKQISEASRKSLSRDLEASRAAKKKVEADLAA	LNAEHQKLKE
emm22	REIDKRYQEQ	LQKQQQLETEKQISEASRKSLSRDLEASRAAKKKVEADLAA	LNAEHQKLKE
emm80	NEKEAERKEK	EAEQKKLKEEKQISDASRQGLRRDLDASREAKXQVEKDLAN	LTAELGKVKE
emm116	KQVEKDLANL	TTELDKVKEEKQXSDASRQGLRRDLDASREAKKKVEADLAA	LTAEHQKLKE
emm12	EAVRKAKAQV	EAALKQLEEQNKISEASRKGLRRDLDASREAKKQVEKDLAN	LTAELDKVKE
emm12	EAVRKAKAQV	EAALKQLEEQNKISEASRKGLRRDLDASREAKKQVEKDLAN	LTAELDKVKE
emm3	EAVRQAKAQV	EAALKQLEEQNRISEASRKGLRRDLDASREAKKQVEKDLAN	LTAELDKVKE

Continued from previous page

emm28	REIDKRYQEQ	LQKQQQLETEKQISEASRKSLSRDLEASRAAKKDLEAEHQK	LKEEKQISDA
emm28	REIDKRYQEQ	LQKQQQLETEKQISEASRKSLSRDLEASRAAKKDLEAEHQK	LKEEKQISDA
emm4	KKHQQEQQQL	EAEKQKLAKDKQISDASRQGLSRDLEASRAAKKELEAKHQK	LEAENKKLTE
emm4	KKHQQEQQQL	EAEKQKLAKDKQISDASRQGLSRDLEASRAAKKELEAKHQK	LEAENKKLTE
emm48	KKHQQEQQQL	EAEKQKLAKDKQISDASRQGLSRDLEASRAAKKELEAKHQK	LEAENKKLTE
emm6	QKSKQDIGAL	KQELAKKDEGNKVSEASRKGLRRDLDASREAKKQVEKDLAN	LTAELDKVKE
emm6	QKSKQDIGAL	KQELAKKDEGNKVSEASRKGLRRDLDASREAKKQVEKDLAN	LTAELDKVKE
emm6	QKSKQDIGAL	KQELAKKDEGNKVSEASRKGLRRDLDASREAKKQVEKDLAN	LTAELDKVKE
emm2	REVEKRYQEQ	LQKQQQLEKEKQISEASRKSLSRDLEASRAAKKDLEAEHQK	LKEEKQISEA
emm2	REVEKRYQEQ	LQKQQQLEKEKQISEASRKSLSRDLEASRAAKKDLEAEHQK	LKEEKQISEA
emm48	NKLTTEKDTL	AEKAKKLEEDKQISDASRKSLSRDLEASRAAKKELEANHQK	LETEHQKLKE
st2002	QEQQQKQQQL	EAENQKLTEANKVSEASRKGLSNDLEASRAAKKELEAKHQK	LEADHQALEA

Sequence alignment of C3-repeat

Name	Sites		
emm2	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm2	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm49	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm116	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm8	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm9	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm89	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm118	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm17	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm44	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm94	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm92	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm9	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm116	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm9	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm9	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm49	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm118	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm94	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm92	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm89	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm48	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm44	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEES
emm9	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm8	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm22	KKKVEADLAA	LNAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm22	KKKVEADLAA	LNAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm58	KKKVEADLAA	LNAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm22	KKKVEADLAA	LNAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm28	LEASRAAKKD	LEAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm77	LEASRAAKKE	LEAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm28	LEASRAAKKD	LEAEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm4	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGXSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm116	KKKVEADLAA	LTAEHQKLKEEKQISDASRQGXSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG

emm48	KKELEANHQK	LETEHQKLKEEKQISDASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm2	LEASRAAKKD	LEAEHQKLKEEKQISEASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm2	LEASRAAKKD	LEAEHQKLKEEKQISEASRQGLSRDLEASREAKKKVEADLAEANSKLQAL	E
emm22	KKKVEADLAA	LNAEHQKLKEEKQISDASRKSLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm82	KKKVEADLAT	LNAEHQKLKEEKQISDASRKSLSRDLEASREAKKKVEADLAEANSKLQAL	EKLNKELEEG
emm80	KKQVEKDLAN	LTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLAAL	EKLNKELEES
emm53	KKQVEKGLAN	LTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLAAL	EKLNKELEES

Continued from previous page

emm6	KKQVEKDLAN	LTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLAAL	EKLNKELEES
emm1	KKQVEKDLAN	LTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLAAL	EKLNKELXES
emm12	KKQVEKDLAN	LTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLAAL	EKLNKELEES
emm1	KKQVEKDLAN	LTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLAAL	EKLNKELEES
emm6	KKQVEKDLAN	LTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLAAL	EKLNKELEES
emm12	KKQVEKDLAN	LTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLAAL	EKLNKELEES
emm6	KKQVEKDLAN	LTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLAAL	EKLNKELGRK
emm53	KKQVEKGLAN	LTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLAAL	EKLNKELEES
emm3	KKQVEKDLAN	LTAELDKVKEEKQISDASRQGLRRDLDASREAKKQVEKALEEANSKLAAL	EKLNKELEES
emm1	KKQVEKDLAN	LTAELDKVKEDKQISDASRQGLRRDLDASREAKKQVEKDLANLTAELDKV	KEEKQISDAS
emm80	KXQVEKDLAN	LTAELGKVKEEKQISDASRRGLRRDLDASREAKKQVEKDLANLTAELDKV	KEEKQISDAS
emm4	LEASRAAKKE	LEAKHQKLEADYQVSETSRKGLSRDLEASREANKKVTSELTQAKAQLSAL	EESKKLSEKE
emm4	LEASRAAKKE	LEAKHQKLEADYQVSETSRKGLSRDLEASREANKKVTSELTQAKAQLSAL	EESKKLSEKE
emm48.1	LEASRAAKKE	LEAKHQKLEADYQVSETSRKGLSRDLEASREANKKVTSELTQAKAQLSAL	EESKKLSEKE
emm4	LEASRAAKKE	LEAKHQKLEADYQVSETSRKGLSRDLEASREANKKVTSELTQAKAQLSAL	EESKKLSEKE
st2002	HQKLEADHQA	LEAKHQKLEADHQVSETSRKGLSRDLEASREANKKVTSELTQAKAQLSVL	EESKKLSEKE
emm2	LEASRAAKKE	LEAKHQKLEADYQVXETSRKGLSRDLEASREANKKVTSELTQAKAQLSAL	EESKKLSEKE

Appendix E

Basic Alignment Search Tool (BLAST) results of streptococcal superantigens

SpeA

> gb|CP000017.1| Streptococcus pyogenes MGAS5005, complete genome

Length=1838554

Features in this part of subject sequence:

enterotoxin

Score = 385 bits (208), Expect = 6e-107

Identities = 212/214 (99%), Gaps = 1/214 (0%)

Strand=Plus/Minus

```
Query  6       ATATC-AATGATAGGCTTTGGATACCATCGATTGATACTTTAACGACTATCTTTTTAGGA  64
               ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  984819  ATATCAAATGATAGGCTTTGGATACCATCGATTGATACTTTAACGACTATCTTTTTAGGA  984760

Query  65      ATTTCTAAATGATTCCCTTCATGATTTGTTACCCCTCCGTAGATACATGCACTCCTTTCT  124
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  984759  ATTTCTAAATGATTCCCTTCATGATTTGTTACCCCTCCGTAGATACATGCACTCCTTTCT  984700

Query  125     GCATTTTCACATAAATAACAGAGATGGTAATATTCTACACTATAAATATCAACGTTTTTA  184
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  984699  GCATTTTCACATAAATAACAGAGATGGTAATATTCTACACTATAAATATCAACGTTTTTA  984640
```

```
Query  185     TCCTTAAATAAAGTTGCCATCTCTTGGTTYTTAA  218
               ||||||||||||||||||||||||||| ||||
Sbjct  984639  TCCTTAAATAAAGTTGCCATCTCTTGGTTCTTAA  984606
```

SpeB

> gb|AE004092.1| Streptococcus pyogenes M1 GAS, complete genome

Length=1852441

Features in this part of subject sequence:

pyrogenic exotoxin B

Score = 1633 bits (884), Expect = 0.0

Identities = 888/890 (99%), Gaps = 1/890 (0%)

Strand=Plus/Minus

```
Query  37       ACGAAGCGCAGAAGATATTAASCTTGACAAAGTTAACTTAGGTGGAGAACTTTCTGGCTC
96

                ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||

Sbjct  1699690  ACGAAGCGCAGAAGATATTAAGCTTGACAAAGTTAACTTAGGTGGAGAACTTTCTGGCTC
1699631


Query  97       TAATATGTATGTTTACAATATTTCTACTGGAGGATTTGTTATCGTTTCAGGAGATAAACG
156

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1699630  TAATATGTATGTTTACAATATTTCTACTGGAGGATTTGTTATCGTTTCAGGAGATAAACG
1699571


Query  157      TTCTCCAGAAATTCTAGGATACTCTACCAGCGGATCATTTGACGCTAACGGTAAAGAAAA
216

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1699570  TTCTCCAGAAATTCTAGGATACTCTACCAGCGGATCATTTGACGCTAACGGTAAAGAAAA
1699511


Query  217      CATTGCTTCCTTCATGGAAAGTTATGTCGAACAAATCAAAGAAAACAAAAAATTAGACAC
276

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1699510  CATTGCTTCCTTCATGGAAAGTTATGTCGAACAAATCAAAGAAAACAAAAAATTAGACAC
1699451
```

```
Query  277      TACTTATGCTGGTACCGCTGAGATTAAACAACCAGTTGTTAAATCTCTCCTTGATTCAAA
336

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1699450  TACTTATGCTGGTACCGCTGAGATTAAACAACCAGTTGTTAAATCTCTCCTTGATTCAAA
1699391


Query  337      AGGCATTCATTACAATCAAGGTAACCCTTACAACCTATTGACACCTGTTATTGAAAAAGT
396

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1699390  AGGCATTCATTACAATCAAGGTAACCCTTACAACCTATTGACACCTGTTATTGAAAAAGT
1699331


Query  397      AAAACCAGGTGAACAATCTTTTGTAGGTCAACATGCAGCTACAGGATGTGTTGCTACTGC
456

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1699330  AAAACCAGGTGAACAATCTTTTGTAGGTCAACATGCAGCTACAGGATGTGTTGCTACTGC
1699271


Query  457      AACTGCTCAAATTATGAAATATCATAATTACCCTAACAAAGGGTTGAAAGACTACACTTA
516

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1699270  AACTGCTCAAATTATGAAATATCATAATTACCCTAACAAAGGGTTGAAAGACTACACTTA
1699211


Query  517      CACACTAAGCTCAAATAACCCATATTTCAACCATCCTAAGAACTTGTTTGCAGCTATCTC
576

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1699210  CACACTAAGCTCAAATAACCCATATTTCAACCATCCTAAGAACTTGTTTGCAGCTATCTC
1699151


Query  577      TACTAGACAATACAACTGGAACAACATCTTACCTACTTATAGCGGAAGAGAATCTAACGT
636

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1699150  TACTAGACAATACAACTGGAACAACATCTTACCTACTTATAGCGGAAGAGAATCTAACGT
1699091
```

SpeC

> gb|AE004092.1| Streptococcus pyogenes M1 GAS, complete genome

Length=1852441

Features in this part of subject sequence:

pyrogenic exotoxin C precursor, phage associated

Score = 985 bits (533), Expect = 0.0

Identities = 544/550 (99%), Gaps = 3/550 (1%)

Strand=Plus/Minus

```
Query  9       TCT-AGAAGAGACATTTCGAATGTTAAAAGTGATTTACTTTATGCATACACTATAACTCC
67
                       ||| ||||
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  569204  TCTAAGAA-AGACATTTCGAATGTTAAAAGTGATTTACTTTATGCATACACTATAACTCC
569146


Query  68      TTATGATTATAAAAATTGCAGGGTAAATTTTTCAACGACACACACATTAAACATTGATAC
127
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  569145  TTATGATTATAAAAATTGCAGGGTAAATTTTTCAACGACACACACATTAAACATTGATAC
569086


Query  128     TCAAAAATATAGAGGGAAAGACTATTATATTAGTTCCGAAATGTCTTATGAGGCCTCTCA
187
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  569085  TCAAAAATATAGAGGGAAAGACTATTATATTAGTTCCGAAATGTCTTATGAGGCCTCTCA
569026


Query  188     AAAATTTAAACGAGATGATCATGTAGATGTTTTTGGATTATTTTATATTCTTAATTCTCA
247
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  569025  AAAATTTAAACGAGATGATCATGTAGATGTTTTTGGATTATTTTATATTCTTAATTCTCA
568966
```

```
Query  248     CACCGGTGAGTACATCTATGGAGGAATTACGCCTGCTCAAAATAATAAAGTAAATCATAA
307

               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  568965  CACCGGTGAGTACATCTATGGAGGAATTACGCCTGCTCAAAATAATAAAGTAAATCATAA
568906


Query  308     ATTATTGGGAAATCTATTTATTTCGGGAGAATCTCAACAGAACTTAAATAACAAAATTAT
367

               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  568905  ATTATTGGGAAATCTATTTATTTCGGGAGAATCTCAACAGAACTTAAATAACAAAATTAT
568846


Query  368     TCTAGAAAAAGATATCGTAACTTTCCAGGAAATTGACTTTAAAATCAGAAAATACCTTAT
427

               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  568845  TCTAGAAAAAGATATCGTAACTTTCCAGGAAATTGACTTTAAAATCAGAAAATACCTTAT
568786


Query  428     GGATAATTATAAAATTTATGACGCTACTTCTCCTTATGTAAGCGGCAGAATCGAAATTGG
487

               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  568785  GGATAATTATAAAATTTATGACGCTACTTCTCCTTATGTAAGCGGCAGAATCGAAATTGG
568726


Query  488     CACAAAAGATGGGAAACATGAGCAAATAGACTTTATTTGACTCACCAAATGAAGGGACTA
547

               |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||

Sbjct  568725  CACAAAAGATGGGAAACATGAGCAAATAGACTT-ATTTGACTCACCAAATGAAGGGACTA
568667
```

SpeF

> gb|AE004092.1| Streptococcus pyogenes M1 GAS, complete genome

Length=1852441

Features in this part of subject sequence:

mitogenic factor

Score = 1354 bits (733), Expect = 0.0

Identities = 743/747 (99%), Gaps = 3/747 (0%)

Strand=Plus/Minus

```
Query  10       GCTAGTAAAATTTTTC-ATGGTAGCTCTTGTATCAGCCACAATGGCTGTAACAACAGTCA
68

                ||||||||||| ||||| ||||||||||||||||||||||||||||||||||||||||||

Sbjct  1702652  GCTAGTAAAA-TTTTCAATGGTAGCTCTTGTATCAGCCACAATGGCTGTAACAACAGTCA
1702594


Query  69       CACTTGAAAATACTGCACTGGCACGACAAACACAGGTCTCAAATGATGTTGTTCTAAATG
128

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1702593  CACTTGAAAATACTGCACTGGCACGACAAACACAGGTCTCAAATGATGTTGTTCTAAATG
1702534


Query  129      ATGGCGCAAGCAAGTACCTAAACGAAGCATTAGCTTGGACATTCAATGACAGTCCCAACT
188

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1702533  ATGGCGCAAGCAAGTACCTAAACGAAGCATTAGCTTGGACATTCAATGACAGTCCCAACT
1702474


Query  189      ATTACAAAACCTTAGGTACTAGTCAGATCACTCCAGCACTCTTTCCTAAAGCAGGAGATA
248

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1702473  ATTACAAAACCTTAGGTACTAGTCAGATCACTCCAGCACTCTTTCCTAAAGCAGGAGATA
1702414
```

```
Query  249      TTCTCTATAGCAAATTAGATGAGTTAGGAAGGACGCGTACTGCTAGAGGTACATTGACTT
308

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1702413  TTCTCTATAGCAAATTAGATGAGTTAGGAAGGACGCGTACTGCTAGAGGTACATTGACTT
1702354


Query  309      ATGCCAATGTTGAAGGTAGCTACGGTGTTAGACAATCTTTCGGTAAAAATCAAAACCCCG
368

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1702353  ATGCCAATGTTGAAGGTAGCTACGGTGTTAGACAATCTTTCGGTAAAAATCAAAACCCCG
1702294


Query  369      CAGGCTGGACTGGAAACCCTAATCATGTCAAATATAAAATTGAATGGTTAAATGGTCTAT
428

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1702293  CAGGCTGGACTGGAAACCCTAATCATGTCAAATATAAAATTGAATGGTTAAATGGTCTAT
1702234


Query  429      CTTATGTCGGAGATTTCTGGAATAGAAGTCATCTCATTGCAGATAGTCTCGGTGGAGATG
488

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1702233  CTTATGTCGGAGATTTCTGGAATAGAAGTCATCTCATTGCAGATAGTCTCGGTGGAGATG
1702174


Query  489      CACTCAGAGTCAATGCCGTTACAGGGACACGTACCCAAAATGTAGGAGGTCGTGACCAAA
548

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1702173  CACTCAGAGTCAATGCCGTTACAGGGACACGTACCCAAAATGTAGGAGGTCGTGACCAAA
1702114


Query  549      AAGGCGGCATGCGCTATACCGAACAAAGAGCTCAAGAATGGTTAGAAGCAAATCGTGATG
608

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1702113  AAGGCGGCATGCGCTATACCGAACAAAGAGCTCAAGAATGGTTAGAAGCAAATCGTGATG
1702054
```

SpeG

> gb|AE009949.1| Streptococcus pyogenes MGAS8232, complete genome

Length=1895017

Features in this part of subject sequence:

exotoxin G precursor

Score = 206 bits (111), Expect = 3e-53

Identities = 111/111 (100%), Gaps = 0/111 (0%)

Strand=Plus/Plus

```
Query  9       ACAGTTTACTTTACAGGAATTTGATTTTAAAATAAGAAAATTTCTAATGGAAAAATACAA
68

               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  185048  ACAGTTTACTTTACAGGAATTTGATTTTAAAATAAGAAAATTTCTAATGGAAAAATACAA
185107


Query  69      TATCTATGATTCGGAATCGCGTTATACATCGGGGAGCCTTTTCCTTGCTAC  119

               ||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  185108  TATCTATGATTCGGAATCGCGTTATACATCGGGGAGCCTTTTCCTTGCTAC  185158
```

SpeH

> emb|AM295007.1| Streptococcus pyogenes Manfredo complete genome

Length=1841271

Features in this part of subject sequence:

streptococcal exotoxin H precursor

Score = 697 bits (377), Expect = 0.0

Identities = 381/383 (99%), Gaps = 0/383 (0%)

Strand=Plus/Minus

```
Query  5        GAAGTAGATATTTATGCTCTATCTGCACAAGAGGTTTGTGAATGTCCAGGGAAAAGGTAT
64

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1030920  GAAGTAGATATTTATGCTCTATCTGCACAAGAGGTTTGTGAATGTCCAGGGAAAAGGTAT
1030861


Query  65       GAAGCGTTTGGTGGAATTACATTAACTAATTCAGAAAAAAAAAGAAATTAAAGTTCCTGTA
124

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1030860  GAAGCGTTTGGTGGAATTACATTAACTAATTCAGAAAAAAAAAGAAATTAAAGTTCCTGTA
1030801


Query  125      AACGTGTGGGATAAAAGTAAACAACAGCCGCCTATGTTTATTACAGTCAATAAACCGAAA
184

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1030800  AACGTGTGGGATAAAAGTAAACAACAGCCGCCTATGTTTATTACAGTCAATAAACCGAAA
1030741


Query  185      GTAACCGCTCAGGAAGTGGATATAAAAGTTAGAAAGTTATTGATTAAGAAATACGATATT
244

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1030740  GTAACCGCTCAGGAAGTGGATATAAAAGTTAGAAAGTTATTGATTAAGAAATACGATATC
1030681
```

```
Query  245      TATAATAATCGGGAACAAAAATACTCTAAAGGAACTGTTACCTTAGATTTAAATTCAGGT
304

                ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1030680  TATAATAACCGGGAACAAAAATACTCTAAAGGAACTGTTACCTTAGATTTAAATTCAGGT
1030621


Query  305      AAAGATATTGTTTTTGATTTGTATTATTTTGGCAATGGAGACTTTAATAGCATGCTAAAA
364

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1030620  AAAGATATTGTTTTTGATTTGTATTATTTTGGCAATGGAGACTTTAATAGCATGCTAAAA
1030561


Query  365      ATATATTCCAATAACGAGAGAAT  387

                ||||||||||||||||||||||

Sbjct  1030560  ATATATTCCAATAACGAGAGAAT  1030538
```

SpeJ

> gb|CP000017.1| Streptococcus pyogenes MGAS5005, complete genome

Length=1838554

Features in this part of subject sequence:

exotoxin type J precursor

Score = 905 bits (490), Expect = 0.0

Identities = 492/493 (99%), Gaps = 0/493 (0%)

Strand=Plus/Minus

```
Query  8       ATTTAAGACGTTAAGCTACAATTAAATTACGCATACGAAATCATACCAGTAGATTATACG
67

               ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  362651  ATTAAAGACGTTAAGCTACAATTAAATTACGCATACGAAATCATACCAGTAGATTATACG
362592


Query  68      AATTGTAATATTGATTACTTGACTACTCATGATTTTTATATTGATATTTCCAGTTATAAA
127

               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  362591  AATTGTAATATTGATTACTTGACTACTCATGATTTTTATATTGATATTTCCAGTTATAAA
362532


Query  128     AAGAAAAATTTTTCAGTTGATTCTGAGGTCGAGAGCTATATTACAACAAAGTTTACGAAA
187

               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  362531  AAGAAAAATTTTTCAGTTGATTCTGAGGTCGAGAGCTATATTACAACAAAGTTTACGAAA
362472


Query  188     AATCAAAAAGTAAATATTTTTGGTCTTCCGTACATATTTACTCGTTATGATGTTTATTAT
247

               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  362471  AATCAAAAAGTAAATATTTTTGGTCTTCCGTACATATTTACTCGTTATGATGTTTATTAT
362412
```

```
Query  248     ATATATGGTGGGGTTACACCATCAGTAAACAGTAATTCGGAAAATAGTAAAATTGTAGGT
307

               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  362411  ATATATGGTGGGGTTACACCATCAGTAAACAGTAATTCGGAAAATAGTAAAATTGTAGGT
362352


Query  308     AATTTACTAATAGATGGAGTCCAGCAAAAAACACTAATAAATCCCATAAAAATAGATAAA
367

               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  362351  AATTTACTAATAGATGGAGTCCAGCAAAAAACACTAATAAATCCCATAAAAATAGATAAA
362292


Query  368     CCTATTTTTACGATTCAAGAATTTGACTTCAAAATCAGACAATATCTTATGCAAACATAC
427

               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  362291  CCTATTTTTACGATTCAAGAATTTGACTTCAAAATCAGACAATATCTTATGCAAACATAC
362232
```

SmeZ

> gb|AE004092.1| Streptococcus pyogenes M1 GAS, complete genome

Length=1852441

Features in this part of subject sequence:

mitogenic exotoxin Z

Score = 608 bits (329), Expect = 6e-174

Identities = 335/340 (99%), Gaps = 0/340 (0%)

Strand=Plus/Plus

```
Query  8        TATTTTTTCTTTTACTTCAATATTCATTGCAATAATTTCTCGTCCTGKGTTTGGATTAGA
67

                ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||

Sbjct  1665203  TATTTTTTCTTTTACTTCAATATTCATTGCAATAATTTCTCGTCCTGTGTTTGGATTAGA
1665262


Query  68       AGTAGATAATAATTCCCTTCTAAGGAATATCTATAGTACGATTGTATATGAATATTCAGA
127

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1665263  AGTAGATAATAATTCCCTTCTAAGGAATATCTATAGTACGATTGTATATGAATATTCAGA
1665322


Query  128      TACAGTAATTGATTTTAAAACCAGTCATAACTTAGTGACTAARAAACTTGATGTTAGAGA
187

                |||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||

Sbjct  1665323  TACAGTAATTGATTTTAAAACCAGTCATAACTTAGTGACTAAGAAACTTGATGTTAGAGA
1665382


Query  188      TGCTAGAGATTTTTTTATTAACTCCGAAATGGATGAATATGCAGCCAATGATTTTAAAGA
247

                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1665383  TGCTAGAGATTTTTTTATTAACTCCGAAATGGATGAATATGCAGCCAATGATTTTAAAGA
1665442
```

```
Query  248      TGGARATAAAATAGCTATGTTCTCCGTCCCATTTGATTGGAACTACTTGTCAGAAGGAAA
307

                ||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||

Sbjct  1665443  TGGAGATAAAATAGCTATGTTCTCCGTCCCATTTGATTGGAACTACTTGTCAGAAGGAAA
1665502


Query  308      AGTCATAGCATATACCTATGGCGGAATGACSCCTTATCAA  347

                ||||||||||||||||||||  ||||||||  |||||||||

Sbjct  1665503  AGTCATAGCATATACCTATGGTGGAATGACGCCTTATCAA  1665542
```