MOLECULAR CHARACTERISATION AND VIRULENCE PROFILES OF GROUP A STREPTOCOCCI CAUSING HUMAN INFECTION IN A SOUTH INDIAN COMMUNITY.

Thesis

submitted to

The Tamil Nadu Dr. M. G. R. Medical University, Chennai

In partial fulfillment for the degree of

DOCTOR OF PHILOSOPHY

By

J. JOHN MELBIN JOSE

DEPARTMENT OF MICROBIOLOGY

CHRISTIAN MEDICAL COLLEGE

VELLORE - 632 004, TAMIL NADU, INDIA.

July 2009

CERTIFICATE

This is to certify that the thesis entitled "**Molecular Characterisation and virulence Profiles of Group A Streptococci causing Human Infection in A south Indian community**" is based on the results of the work carried out by me under the supervision of my guide, **Dr. K. N. Brahmadathan** and co-guide **Dr. Mary S. Mathews**. This thesis is submitted under the regulations of the Tamil Nadu Dr. M. G. R. Medical University, Chennai for the degree of DOCTOR OF PHILOSOPHY. This work has not formed the basis of any fellowship, degree or diploma of any other University.

J. John Melbin Jose
Candidate

Station: Vellore
Date: July 2009

ACKNOWLEDGMENTS

First of all, I give all the glory and praise to the **Almighty God**, through my **Lord Jesus Christ**, who, by his sovereign plan, has helped me to do this thesis work successfully.

I elucidate my deep sense of gratitude to **Prof. K. N. Brahmadathan**, my supervisor and guide for accepting me as his research student. I am very much indebt to him for his guidance, encouragement, extreme patience and support all through this doctoral program. His esteemed personality is a source of great inspiration to me and my association with him has developed a great confidence and understanding in the field of streptococcal research. Working with him had been an extremely interesting and a pleasant experience for me.

I deem it as a privilege to thank **Prof. Mary S. Mathews**, the Head of the Microbiology department and my co-guide for her supervision, advice, encouragement and continuous support that have helped me greatly during the course of this doctoral research program.

I like to give special thanks to **Dr. Thangam menon,** Professor of Microbiology, IBMS, Tharamani, the member of the doctoral committee, whose suggestions and comments were of great value to my thesis work.

I wish to express my unfeigned thanks to **Dr. Vinod Abraham** of Community health department for his favors during the course of my research program and for his key role in the effective management of the project. I also appreciate and thank **Dr. Muthulakshmi**, **Dr. Rashme**, **Mr. Pandiarajan** and **Mr. Jayakumar** for their contribution.

I also like to thank **Dr. Shaji, Dr. Jayandran, Dr. Eunice, Dr. Priya,** my friend **Salamun** and others in the Department of Haematology for helping me with DNA sequencing.

I sincerely thank **Dr. Richard M. Krause, Dr. Jim Musser** and **Dr. Nancy Hoe** for their concern and help regarding the partial work done in the Microbial Pathogenesis Lab of Rock Mountain Laboratories, NIAID, NIH, Mountana, USA. I also extend my particular thanks to **Mr. Kent Barbian** and others who helped me with their expertise.

I like to mention a word of thanks to **Dr. Bernard Beall,** the Curator of the *emm* typing Protocol, (Center for Disease Control and prevention) who helped me by spending his valuable time in assigning the new *emm* types encountered in this study and for his mastermind behind the provision of access to the *emm* database and protocol for researchers like me.

I owe my special thanks to **Prof. G. Sridharan**, **Dr. Rajesh Kannangai** and **Mr. Abraham Joseph** of Virology department, CMC for their concern and help regarding the phylogenetic part of this research.

I wish to record my gratitude to **Prof. Mary V. Jesudason,** who, as the Head of the Department, gave me an opportunity to work in the project.

I wish to record my sincere thanks to the past and present faculty Members of the Microbiology department especially, **Prof. M.K. Lalitha, Prof. Elizabeth Mathai, Dr. Balaji, Dr. Joy, Dr. John Jude, Dr. Shalini, Dr. Rani** and **Dr. Padmaja** and all who have helped me and supported me all through the program.

I specially thank my colleague **Mrs. Reena Doss** for her friendly help and moral support towards this work, which is unforgettable. I also have to give a special thanks to **Mr. Jonathan** especially for helping me in times of need.

Special word of thanks are also due to **Mr. Paul Gladstone, Mrs. Promila, Mrs. Inbamalar, Mrs. Ismath, Mrs. Priya, Dr. Sangeetha** and **Dr. Pradeepa** for their fellowship at various times. Thanks are also due to **Ms. Angelin, Mrs. Anne Rebecca, Mr. Charles, Mrs. Dhaya Rani, Mr. Dhinakaran** and all members of Microbiology department who have helped me in many ways.

I also like to acknowledge the friendly favors and encouragement extended by **Dr. Hubert**, **Manu**, **Magesh**, **John**, **Joel** and **Ishmael of Virology Department** and **Sasirekha** of Welcome Microbiology section.

I remember **Mr. Solomon**, **Mr. Prasana** and **Ms. Nithya** for their help regarding statistics.

The grant provided by **Department of Biotechnology**, **Government of India**, for the "Indo-US VAP program" on which I worked is thankfully acknowledged. I also have to thankfully acknowledge the financial support provided by my Guide **Dr. K. N. Brahmadathan,** for the phylogenetic part of this research program.

I want to specially thank my beloved **Father**, and my loving **Brothers**, **Relatives** and **Friends** for their prayers, continuous help, cooperation, patience and encouragement, all through the way.

I do like to bring back the memory of my beloved **Mother**, who would have been the happiest of all, about this accomplishment.

Mere words alone cannot express my gratitude to all whose brilliant efforts and contribution have made this study a reality.

J. John Melbin Jose

CONTENTS

Chapter	Titles	Page No.
1	Introduction	1
2	Aim & Objectives	8
3	Review of Literature	9
4	Scope and Plan of Work	53
5	Materials & Methods	55
6	Results	72
7	Discussion	106
8	Summary & Conclusion	149
9	Recommendations	158
10	Bibliography	-

1. INTRODUCTION

Since the formation of life on earth, complementation and competition for survival between organisms is a usual phenomenon in an ecosystem. There is no place in the earth where there are no microorganisms. We human beings are any time exposed to microorganisms from the environment. Most of them are harmless saprophytes while few of them could colonise the outer surface of the body such as skin and few others can colonise the inner surface such as mucosal lining on the upper respiratory tract, gastrointestinal tract and even in the urogenital tract. Some of the microorganisms during the course of time adapt themselves to cross the physical barrier and overcome the immune system for their own benefit while leading to illness and discomfort for the host which may even result in death. Those organisms are termed as pathogens and the disease they cause are termed as infectious diseases. They are the major cause of morbidity and mortality in human beings world wide [1].

One such pathogen is Group A Streptococcus (GAS), consisting of the species *Streptococcus pyogenes*. The term streptococcus was coined by Billroth in 1874 from two Greek words *streptos*, meaning twisted or chain, and *kokhos*, meaning a berry or seed based upon its appearance. GAS is a versatile species of gram-positive extracellular bacterial pathogen causing diseases ranging from uncomplicated infections to life threatening invasive diseases and sequelae in humans. GAS has also gained notoriety as “flesh-eating bacterium” which invades skin and soft tissues and in severe cases leaves infected tissues or limbs destroyed [2]. GAS ranks fourth among the bacterial cause of global mortality as estimated by World Health Organization (WHO) in 2004 [3].

As pathogens, GAS has developed complex virulence mechanisms to avoid host defense mechanisms. GAS is the most pathogenic group among many other groups of the genus *Streptococcus* classified based on the specific carbohydrate antigen present on the cell surface. GAS colonizes the throat or skin and is responsible for a number of suppurative infections and non-suppurative sequelae [4]. The increased interest in GAS research is due to the unanswered questions about the influence of genetic mechanisms, including gene exchange on the pathogenesis and various clinical manifestations of GAS infections, which include toxic shock syndrome and necrotizing fasciitis in the developed countries and the continuing incidence of post-streptococcal sequelae such as Rheumatic Fever (RF) and Rheumatic Heart Diseases (RHD) in the developing world [5]. Despite intense research, the molecular mechanisms of GAS disease remains unclear [6]. Therefore it becomes important to plan the research within the framework of the molecular and genetic understanding about GAS and the evolutionary events that influence the emergence and demise of epidemics in time and place [5].

Reported estimates state that there are at least 5,17,000 deaths each year due to various GAS diseases worldwide [3]. Although RF and RHD have declined in many parts of the world due to effective antibiotic therapy and increased standards of living, it is still a problem of major concern in developing countries including India [4, 5, 7-9]. It is estimated that the occurrence of new cases of RF is around 50,000 and 5,00,000 per year in India. It is also estimated that approximately 6,00,000 children, less than 15 years of age suffer from RHD and that 1,21,000 children are newly diagnosed for RF each year [7]. The burden of other sub clinical GAS infections should be certainly more than these estimates. Although timely benzathine penicillin prophylaxis can prevent the onset of these complications, their use has not been

widely adopted in developing countries [10]. Moreover the intake of phenoxymethyl penicillin fails to eradicate 30 % of the *Streptococci* from throat due to protective internalization into the epithelial cells [11, 12]. Thus, there is a strong case for the use of GAS vaccine in such countries.

Studies on GAS isolates from different geographical locations have shown substantial chromosomal and allelic diversity in relation to diverse infection types [13, 14]. So, application of a valid typing system is essential to address the epidemiological problems concerning GAS infection and diseases. The methods used for typing of GAS can be divided into phenotypic and genotypic procedures. Phenotypic procedures take advantage of biochemical, physiological and biological phenomena, whereas genetic procedures aim to detect polymorphisms at the level of nucleic acids. Conventional typing methods mostly depend on the phenotype of the test strains and differentiate between them based on the growth characteristics, antigenic structure, biochemical character, antimicrobial susceptibility pattern or susceptibility to bacteriophages etc. Previously, serotyping was the universally used phenotypic method for the discrimination of GAS. Among many techniques, M typing was considered to be the gold standard, excellently developed by the pioneering work of Rebecca Lancefield in 1943, which is based on the antigenic variation in the amino-terminal portion of M protein [5, 15, 16].

The M protein is a major virulence factor and 83 different M serotypes were designated with in the sequence of M1- M93 [17]. This technique was able to generate more knowledge on the epidemiology of GAS infections in the past [18]. However, this method requires skill, intensive labor, consumes time and is applicable to a narrow range of types for which they were developed. Consequently, due to the

inherent difficulties encountered by limited array of typing antisera, there were always a group of isolates remained as nontypable [19]. Therefore, sequencing of the gene that encodes the type specific hyper-variable region of M protein known as "*emm* typing" was developed [19-21]. Using this more sensitive and less demanding sequence based *emm* typing, an extension of Lancefield's M serotypes have been established and are now well defined and widely used. This method has led to the identification of an increasing number of *emm* types, with more than 170 *emm* types and 750 subtypes from GAS [22]. This technique would not have been successful without the open access to the online database maintained by Centers for Disease Control & Prevention (CDC) at http://www.cdc.gov/ncidod/biotech/strep/strepblast.html [17, 19, 20].

Many a times, GAS occurs as carriers in the upper respiratory tract of healthy children without any symptoms [23]. Although many hypothesis were proposed, the mechanisms that promote translation of carriage to disease are still not clearly understood [6]. Epidemiological understanding of the ubiquitous occurrence and versatile nature of this pathogen is essential for the development of novel preventive or therapeutic schemes against this prevalent human pathogen [13]. The concept of distinct throat and skin specific types of GAS is evident from several epidemiological reports [4, 24-26]. A similar concept of relationship between M types with RF and pharyngitis has been documented [4, 27]. Clonal expansion of certain GAS causing invasive diseases was also reported [28, 29]. On the other hand, certain studies have documented a contrasting epidemiology of RF/RHD in developing countries and certain indigenous populations. The contrasting association between throat specific strains infecting skin and the skin specific strains infecting throat has been reported from Ethiopia [16], Australian aboriginals [30] and Nepal [22]. Moreover the types

not implicated with RF were isolated from the RF cases in Hawaiian ethnic group, USA [31], Ethiopia [16] and Thailand [32]. The association between pharyngitis and RF has been made questionable due to certain reports reflecting the high prevalence of GAS in throat carriage with low incidence of RF [33] as well as a high rate of RF with low incidence of pharyngitis, but high incidence of impetigo [34, 35]. GAS diseases in many developed countries is limited to a very few types [28, 36, 37] while it is highly diverse in many developing countries including India [22, 38, 39]. Reported literature on invasive diseases from India although very scanty are associated with various predisposing factors [40].

The phenomenon of tissue specificity is further correlated with the genetic marker exerted by the arrangement pattern of *emm* like genes in the *vir* locus into 5 main family groups namely A through E. The *emm* family pattern A-C is reported to be throat specific, D is skin specific and E is common for both the sites [41]. This association of the *emm* family pattern with the site of infection is very arbitrary and not consistent with many other reports [16, 30]. The *emm* Pattern E is always associated with its ability to produce serum opacity factor (SOF) which can be identified using PCR as well [42]. The complex epidemiology of GAS infection in endemic region highlights the necessity to understand the molecular epidemiology of GAS infection in such region to design proper prevention strategies.

GAS have adopted specific virulence mechanisms through an evolutionary process of selection of the fittest due to distinct spatial and temporal epidemiological trends [43]. The resolving power of epidemiological typing of microorganisms has been greatly expanded by molecular analysis of DNA [44]. Since DNA represents a fundamental molecule of cellular identity, assessing genetic similarity is a measure of

epidemiological relatedness between the study isolates [45]. The *emm* typing system does not have any relationship in terms of its numerical order but is designated based on the order of identification and confirmation of genotypes. So the knowledge about *emm* typing alone does not give any idea about the genetic relatedness between types. Till date no phylogenetic study has been reported based on the complete sequence of *emm* gene which will demonstrate a better relationship between *emm* types. Population genetic studies of GAS, reported from different regions are based on techniques such as Multilocus Sequence Typing (MLST) or other less variable genes that show limited relationship between *emm* types and are more complicated. The present study will serve as a background to understand the development of clones in an endemic situation and in relation to various clinical manifestations and differential tissue tropism. This mode of systematic approach to decode the pathogenic mechanism in relation to the presence of virulence factors has not yet been undertaken for GAS so far. Further, analysis of the virulence factors with the knowledge of evolutionary cluster and the *emm* family pattern will reveal more about the pathogenic potential of specific *emm* types with reference to disease and site of infection.

A major driving force in the emergence and evolution of pathogenic isolates is the horizontal acquisition of virulence factors [46]. Most of the virulence gene inheritance was reported to be associated with phage-mediated lysogenic conversion, thereby providing a potential molecular basis for characteristic variation in the disease manifestations [14]. Complete genome sequencing of serotypes M1, M3 and M18 has revealed that the majority of differences among them correspond to prophages and other exogenous elements. The strains that were sequenced contained many prophages (polylysogenic), each encoding for one or two putative virulence factors [47-49]. Most of these have super antigenic properties and can activate large subsets

of T cells with massive cytokine release. So looking into these virulence factors among GAS isolates from our region may explain the possible mechanism(s) for generating new strains with increased pathogenic potential. In the present study, the contemporary isolates of GAS collected from school children have been compared with invasive isolates, with the objective of assessing the presence of unique or high prevalence of any of the known phage associated virulence factors.

The extra cellular proteins and toxins that are encoded in the chromosomal genome of GAS play a major role in the successful pathogenesis of the disease. These include SLO, SKA, SIC, SPE-B, SPE-F, SPE-G, SPE-J, and SMEZ. Bacterial adherence to host tissues is the first step towards colonization which leads to further invasion and severe complications. Binding of GAS to fibronectin or collagen in the epithelial cells or basement membrane is essential for adherence. Such adherence mechanisms have two important implications: firstly, it might confer certain site or tissue specificity to the GAS. Secondly it may facilitate invasion of GAS into deeper tissues. Several studies have shown that different repertoire of adhesion genes and their interactions with each other may determine tissue tropism [50] and thus the pathogenicity of GAS [51-53]. Therefore profiling of toxin and surface proteins genes that help in the adhesion and pathogenesis of GAS strains appears to be a valuable tool to define the ability of diverse M types in causing various diseases and preference for site of infection.

2. AIM AND OBJECTIVES

2.1. OVERALL AIM

To know the genotypic distribution and understand the pathogenic determinants of GAS causing pharyngitis and impetigo seen during a rural school survey and compare it with invasive isolates from patients attending a tertiary hospital in southern India.

2.2. SPECIFIC OBJECTIVES

1. To determine the distribution of *emm* types prevalent among GAS isolated from children aged 7 – 11 years, who are asymptomatic carriers of GAS in throat or with pharyngitis and impetigo, over a period of 24 months, by gene sequencing.

2. To study the relative distribution of genetic marker for tissue tropism by *emm* family patterns and serum opacity factor gene detection among different *emm* types of GAS.

3. To determine the genetic similarity and differences between the various *emm* types using cluster analysis of the full *emm* gene sequence based phylogenetic tree.

4. To study the PAVF gene profiles of GAS isolated from the school survey (community strains) and compare them with invasive isolates from patients in a tertiary care hospital.

5. To study chromosomally coded virulence factor (CCVF) gene profiles of community GAS isolates.

3. REVIEW OF LITERATURE

Group A Streptococci is a very common pathogen responsible for wide variety of human diseases that even causes life threatening invasive diseases and post streptococcal sequelae. GAS ranks ninth among all the infectious diseases and fourth among the bacterial cause of global mortality as estimated by WHO in 2004 [3]. The first good work on GAS was recorded as early as 1879 by Louis Pasteur. The modern classification of hemolytic streptococci, beginning in 1918 by Lancefield and others led to the recognition that only GAS is responsible for pharyngitis that causes Acute Rheumatic Fever (ARF). Despite intensive study for nearly a century, the infections caused by GAS, still remains to be a formidable problem [5]. Yet, there is no successful licensed vaccine available until this day. The manifestations of GAS disease continues especially in the developing countries and the indigenous population of the developed countries [3]. The epidemiology and the pathogenesis are not completely understood, but the GAS research is getting intensified as the knowledge and technology increases [54]. In this contest the review of Literature will help to recollect what is known about this pathogen and will help to direct and formulate strategies to focus on the issues in order to rectify the problem.

3.1. HISTORY OF GAS

The term streptococcus was first applied by Billroth in 1874 to a chain-forming coccus he saw in wound infections. The first work on the morphology, cultural characters and the pathogenesis of streptococcus on mice and rabbits was done by Pasture in 1879 and by Rosenbach in 1884. The classification became a problem in the early days of 20th century because streptococci were classified according to the source from which they were recovered. Classification by Schotmuller in 1903,

determined that streptococci produced various kinds of hemolysis on blood agar plates. In 1919, the different hemolytic pattern was described by Brown as alpha (partial), beta (complete) and gamma (no) hemolysis as it is known even today. Rebecca Lancefield classified the organism according to the group specific polysaccharide led to the recognition that GAS is responsible for most of the diseases and has categorized GAS into specific types according to the M protein in the cell wall [4]. At about the same time, Griffith differentiated GAS into different types based on T antigen by slide agglutination [5]. In 1932 Todd developed a test for measuring antibodies to streptolysin O, demonstrating the host response to streptococcal infection. There were several other important works contributed to what we know about GAS today, which will be described below in the contest of the scope and plan of this research topic.

3.2. THE GENUS STREPTOCOCCUS

The genus *Streptococcus* consists of several Gram positive, non-sporing, non-motile chain forming spherical or ovoid cells, arranged in chains or pairs. They are facultatively aerobic but most of them require carbon dioxide for growth. All the species in this genus fail to reduce nitrate. They are homo-fermentative, meaning that the sole product of glucose fermentation is lactic acid with no gas formation. Many species are members of commensal flora on mucosal membranes of human or animal, while some are highly pathogenic. The G+C content of DNA of this genus ranges from 36 to 46 %. The name of the genus was coined by Billroth in 1874. The streptococcus genetics has a long history and has gained importance with the development of genetics and biology in general. Transformation of a bacterial property (capsule production) was first demonstrated in pneumococci by Griffith in

1928. These studies subsequently led to the identification of DNA as the genetic material [5]. Molecular genetic studies on streptococci have provided substantial evidence that horizontal gene transfer and recombination play an important role in the evolution of bacterial species [55-57]. All the members of the genus *Streptococcus* were also classified into 20 groups, named as A, B, C etc. through V except I and J, based on the serological difference in the polysaccharide antigen present on the cell wall. The group antigens of groups D, N and Q are teichoic acid [58, 59]. All the species are obligate parasites of the mucosal membrane and some species resides on tooth surface of humans and several animals [60]. Many of the species are dominant members of the upper respiratory tract and few others can colonize the intestinal and genital tracts as well. Most of the species are the normal flora which may cause infection when introduced into normally sterile site of the body or in immunocompromised patients. Other species are true pathogens which spread from person to person and cause infection in non-immune individuals [61].

Based upon the modern Taxonomy in Bergey's manual, S.pyogenes (GAS) comes under the catalase-negative, gram-positive cocci of human origin in the phylum Firmicutes under the proposed class Bacilli are now in the order Bacillales. Among six families in the proposed order Bacillales, the genus Streptococcus comes under the family "lactobacillales". The genus Streptococcus, which contains the most important human pathogens, can be operationally divided into seven groups, consisting of 51 species including 4 species with subspecies (3 having 2 subspecies and 1 has 3 sub species). *S. pyogenes* comes under pyogenic group of the Genus Streptococcus [62].

3.3. MORPHOLOGY OF GAS

GAS may be more or less spherical cells that are arranged in chains of up to 50 cells or more. Chain formation is most pronounced in broth media. The individual cells are typically 0.5-1.0 µm × 1.0-2.0 µm. Growth occurs by elongation on the axis parallel to the chain. Cross walls form at right angles to the chain and after division an appearance of pairing may remain. Capsulation is not a regular feature of GAS but some form a capsule of hyaluronic acid. They are readily stained with ordinary dyes and are gram positive although old cultures may show variable staining [58].

3.4. CULTURAL CHARACTERISTICS OF GAS

GAS is fastidious and growth on ordinary nutrient media is generally poor. On media enriched with blood, serum or a fermentable carbohydrate, growth is more profuse. In addition to complex nutrient source, such as meat infusion, and glucose source, culture media should contain buffers to prevent acidification and subsequent killing during growth. Colonies on blood agar seldom exceed 1 mm in diameter after 24 hours growth at 37˚C. Some strains characteristically form relatively large colonies whereas others form small colonies. GAS induces zones of complete hemolysis around colonies on blood containing agar media. Crystal violet can be used in situations where cultures are likely to be overgrown with staphylococci to select out streptococci because high concentration of staphylococci may completely mask streptococcal hemolysis and may be difficult for selecting out Streptococci. Supplementation with sulfamethoxazole (23.75 µg/ml) and trimethoprim (1.25 µg/ml) may increase the recovery of GAS [58].

3.5. CLINICAL INFECTIONS

GAS has long been recognized as a human pathogen causing an exceptionally broad range of diseases. GAS disease may be categorized into ***superficial***, ***invasive*** and ***immunologically mediated*** [4, 63].

3.5.1. The Carrier state

GAS is exclusively a human pathogen and hence it must survive in a human reservoir. Although they are not considered normal flora, GAS can occur without clinical symptoms of disease. Many individuals carry them asymptomatically in their upper respiratory tract and other anatomic sites [6, 23]. Studies of the natural history of RF development and transmission of GAS to others are substantially lower among asymptomatic carriers than among individuals with symptomatic pharyngitis, however, this situation is a potential source of infection to others. Outbreaks of food borne infection and nosocomial puerperal infection have been traced to asymptomatic carriers who harbor GAS in the throat or on the skin. The carrier rate may vary from place to place but may be higher before or during outbreaks [64]. Over the years several hypothesis have been proposed to explain the asymptomatic carriage of GAS [23]. In principle, it may be due to (i) mutations in the pathogen that down regulate virulence, (ii) a productive immune response by the host that constrains pathogen proliferation, (iii) internalization into the host cells or some combinations of any of the three reasons [6]. High rate of carriage of GAS in the throat influences the diagnosis of pharyngitis, as many of the GAS obtained during pharyngitis were confirmed as viral etiology [7].

3.5.2. Superficial infections

Group A streptococci are extracellular bacterial pathogens that causes a variety of pyogenic infections involving the mucous membranes, tonsils, skin, and deeper tissues resulting in pharyngitis, impetigo/pyoderma, erysipelas, cellulitis, necrotizing fasciitis, toxic streptococcal syndrome, scarlet fever, septicemia, pneumonia, and meningitis. Infections may be mild to extremely severe. Complications such as sepsis, pneumonia, meningitis and several specific clinical syndromes, such as toxic streptococcal syndrome and necrotizing fasciitis can be propagated as the result of superficial infections [65].

3.5.2.1. Pharyngitis

Streptococcal pharyngitis is generally characterized by inflammation of the pharynx. While numerous bacterial and viral pathogens cause pharyngitis, GAS is the most common source of pharyngitis primarily affects school children of age 5 to 15years [66]. All ages are susceptible to spread of the organism under crowded conditions, such as those at schools and military facilities [67]. Pharyngitis and its association with rheumatic fever are seasonal, occurring in the fall and winter [27]. Organisms which colonize the skin can also colonize the throat, but streptococcal strains which commonly produce skin infections do not lead to rheumatic fever. Groups C and G can also cause pharyngitis and must be distinguished from group A organisms after throat culture and serology [68]. Certain M protein serotypes, such as M types 1, 3, 5, 6, 14, 18, 19, and 24 are found associated with throat infection and rheumatic fever [69]. These pharyngitis-associated serotypes do not produce opacity factor [4, 70].

3.5.2.2. Impetigo (Pyoderma)

Streptococcal impetigo, otherwise known as pyoderma which occurs on the exposed areas of the skin [66]. Pyoderma occurs in the summer and can be associated with the production of acute glomerulonephritis [71]. M serotypes such as 2, 49, 57, 59, 60, and 61, which are associated with pyoderma and acute glomerulonephritis. Differentiations of GAS into skin and throat M types are not clear cut. While many of the so called skin types are also capable of initiating throat infection, streptococci of the so-called throat M types are also implicated in skin infections, primarily trauma sites [72]. The infection is limited to the epidermis, usually on the face or extremities, and is highly contagious [27, 73]. Group A streptococcal strains may enter the skin through abrasions and other types of lesions to penetrate the epidermis and produce erysipelas or cellulitis [71].

3.5.3. Invasive Infections

GAS have adopted complex virulence mechanisms to avoid host immune response to penetrate in the tissues of the host [74]. Severe infections such as necrotizing fasciitis are characterized by extensive tissue damage, vascular dissemination and systemic disease manifestations that result in high morbidity and mortality. Invasion may occur either by progression of a superficial infection such as pharyngitis or direct inoculation on a penetrating injury [75]. Severe invasive infection followed by GAS with shock like syndrome accompanied by multi-organ dysfunction called streptococcal toxic shock syndrome (STSS) and septicemia have increased world wide, probably because of the increased prevalence of more toxigenic GAS strains.

3.5.4. Post Streptococcal Sequelae

Serious life threatening complications arise from recurring GAS infection due to the immunologically mediated sequelae of Glomerulonephritis or RF resulting in RHD [63]. These conditions are caused by cross-reactivity between an immune response against GAS and the host tissue [27, 76].

3.5.4.1. Acute Glomerulonephritis

Acute Glomerulonephritis is a representative disease of acute nephritic syndrome characterized by the sudden appearance of edema, hematuria, proteinuria and hypertension. It was found to be associated with poststreptococcal infection, which related to tropical climate and demonstrated peak occurrence in the summer [77]. The nephritogenicity of GAS appears to be related with specific M protein serotypes M2, M49, M42, M56, M57 and M60. The pathogenic events that lead to glomerulonephritis are related to an immunological phenomenon involving immune complexes. A latency period, decrease in serum complement, and observed effects on the glomeruli suggest an immune mediated event. Several mechanisms have been proposed, including immune complex deposition, reaction of antibodies cross-reactive with streptococcal and glomerular tissues by streptococcal products such as proteinase or streptokinase, and direct complement activation by streptococcal components deposited in the glomeruli [4]. The general absence of individual recurrence may be due to the presence of type specific antibodies against the nephritogenetic factors.

3.5.4.2. Acute rheumatic fever

Rheumatic fever is non-suppurative sequelae which may develop following repeated exposure to GAS. It is a major health concern in developing countries and a

preventable childhood heart disease worldwide [27, 78]. The rheumatogenicity of GAS appears to be related to specific M protein serotypes M1, M3, M5, M6, and M 18. Rheumatic fever is a delayed sequel to group A streptococcal pharyngitis. The disease manifests as an inflammation of the joints (arthritis), heart (carditis), central nervous system (chorea), skin (erythema marginatum), and/or subcutaneous nodules [79]. These five major clinical manifestations, any of which may be seen in rheumatic fever, were established by the Jones criteria and revised by the American Heart Association. The disease is autoimmune in nature and most likely results in part from the production of auto-reactive antibodies and T cells shown to cross-react with components of the group A streptococcus and host tissues. The medical importance of rheumatic fever is serious cardiac involvement, with myocarditis or valvulitis leading to death or valve replacement [79]. Outbreaks of rheumatic fever have also been observed in the United States since 1983, when Utah reported an outbreak associated with children from middle to upper income families with good access to medical care, but crowding was a factor associated with disease [80]. Mucoid M18 strains have been associated with the outbreak. Although progress has been made in the understanding of rheumatic fever and its pathogenesis as an autoimmune disease, there is still much to be elucidated about the disease process [81].

3.6. TREATMENT FAILURE

Disease caused by GAS became a major health issue due to increased number of reports of GAS that failed to be eradicated by treatment with standard regimens [82]. The failure of eradication may be reflected in the clinical failure of the patients. The cause of this failure may be due to patient noncompliance, antimicrobial tolerance or co-infection with other pathogens. Till date none of the isolates from these patients

were resistance to penicillin. But there are reports about increasing resistance to erythromycin [83]. Now it is known that the treatment failure may be mainly due to the internalization of GAS into the epithelial cells that protect them from the reach of antibiotics [12]. Treatment with antibiotics is not always effective though the mortality rate is decreased [84]. The failure of penicillin to treat severe invasive streptococcal infections successfully is attributed to the phenomenon that a large inoculum reaches stationary phase quickly and penicillin is not very effective against slow-growing bacteria [85].

3.7. DISEASE BURDEN OF GAS

GAS is a human specific pathogen transmitted from one to another by direct contact with infected skin or respiratory aerosols, which gives rise to infections of varying severity. They can remain dormant in the human oro-pharynx, which serves as the reservoir in asymptomatic carriers [23]. They are the common cause of bacterial pharyngitis and impetigo. The concept of distinct throat and skin strains became evident from decades of epidemiological studies favoring certain serotypes of GAS with a tendency to cause throat infection and similarly other serotypes were associated with impetigo [69].

In the past, GAS was a common cause of puerperal sepsis or child bed fever. GAS is also responsible for streptococcal toxic shock syndrome and invasive diseases resulting in the destruction of soft tissues and limbs. GAS has been investigated for its development of post streptococcal infection sequelae, including acute rheumatic fever, acute glomerulonephritis and reactive arthritis. GAS have been associated with Tourette’s syndrome, tics and movement and attention disorders [4].

There is accumulating evidence that streptococcal disease and its complications thrive in many of the developing areas of the world where people have not had the advantage of advanced health care systems [86]. There were also many other factors suggested as responsible for the incidence of GAS diseases [3, 7]. Nasopharyngeal carriage and infection of GAS varies widely according to the age, geography, socioeconomic status and hygiene [7].

It was concluded that, approximately 15 % of school-age children suffer from pharyngitis episode of GAS confirmed by culture, each year in many developed countries and similarly 4-10 % of adults are affected. The incidence in less developed countries may be 5-10 times greater. With possible lower estimate, over 600 million cases of GAS pharyngitis occur annually among people aged over 4 years, and over 550 million of these occur in less developed countries [3].

The pyoderma (impetigo) prevalence ranged from approximately 1 to 20 % in less developed countries and 40 – 90 % in pacific regions. So it was estimated that over 111 million children aged under 15 years have pyoderma in less developed countries at any time [3].

ARF cases were more than 336,000 in 5-14 year old children. Based on this it was calculated that 471,000 cases occur in all age groups. It was also estimated that 282,000, ie, 60 % of all ARF will develop RHD. Due to ARF and RHD, 233,000 to 294,000 deaths occur each year [3].

In India, ARF and RHD continues to be a major cause of mortality and morbidity [7]. According to this estimate, if 0.3 – 3 % of all pharyngitis develops in to RF, then 50,000-5000,000 new cases of ARF occurs each year in India.

According to one of the old studies conducted in south Indian children, 4.2 % of the children had pharyngitis and 14.9 % had pyoderma. The prevalence rate of RF was 0.5/1000 and RHD was 4.4/1000, given an overall rate of 4.9/1000 [87]. One of the reports published in 2003 points out the decline of RHD in the rural school children with 0.68/1000 [88]. Another study conducted quiet recently in a very small group of children, in Chennai, shows that only 10% of all the respiratory infections are caused by GAS [89]. However, the lacunae in the diagnosis of streptococcal pharyngitis in the Indian contest [90] should be considered while determining the disease burden due to GAS.

3.8. REVIVAL AND RESURGENCE OF SEVERE GAS DISEASES AND SEQUELAE

A decline in the prevalence of serious GAS infection in the western world was observed since the mid 1980s [91]. Possible reason for this decline may be due to improved socio-economic conditions, better hygiene and due to treatment with antibiotics [92]. Research on GAS led to the successful clinical trials to prevent GAS infections and ARF by treating the streptococcal infected with penicillin [93]. Prevention of ARF has led to almost complete disappearance of RHD in the industrial countries [5]. Out breaks were also reported from Pennsylvania, Ohio, Tennessee and West Virginia and at the Naval Training center in California at San Diego [8]. During the same time streptococcal toxic shock syndrome, bacteremia and severe invasive GAS infections were reported in United States, Europe and many other countries. Increased bacteremic infections were reported in Colorado, Sweden and the United Kingdom. The resurgence was associated with the increase in the number and clinical severity of GAS infections [8]. An epidemiological shift in the distribution of M serotypes had been associated with the increase in severe GAS infections. Although

several different M serotypes have been isolated from severe invasive diseases, M1 and M3 serotype were predominating. The sudden change in the epidemiology of GAS disease might have resulted by change in the virulence properties of the pathogens. Host susceptibility was also suggested as an important factor in the immunological diseases like toxic shock syndrome [4]. Though the increased virulence profiles of those isolates termed as highly virulent clones have been shown [94-96], the return of GAS serious infections as the cause of rheumatic fever outbreaks and toxic shock syndrome reminded that medical sciences knows little about the evolutionary and epidemiological forces that drive temporal changes in disease frequency and severity [97]. The resurgence combined with increasing antibiotic resistance increased anxiety, which led to the description of this pathogen as "superbugs".

3.8.1. Emergence of highly invasive clones of GAS

The resurgence of GAS as a cause of serious human infection in US and Europe, starting from 1980s has heightened awareness about this organism. The examination of 108 invasive isolates by multilocus electrophoretic typing (MLEE) during the time led to the recognition of allelic variation and clonal relationship in the emergence of severe GAS diseases [28]. The study found extensive variations in GAS and, importantly more than half of the isolates were members of clonally related lineages designated as multi locus electrophoretic type 1 and 2 (ET1 and ET2), represented by serotypes M1 and M3 respectively. Further analysis identified a strong statistical relationship in the presence of the gene, *speA* encoding the superantigen streptococcal pyrogenic exotoxin A. This was confirmed by another study which shows 90% of the isolates from serious diseases had a characteristic restriction fragment profile and

were positive for *speA* gene [94]. This study led to the concept about the existence of two distinct clones of serotype M1 with the difference of approximately 70 kb of additional DNA fragment in the *speA* positive strains. The enhanced virulence of this clone was also proved in the mouse model [98]. It was then discovered that some contemporary M1 isolates contain DNase encoding in the prophage, which is crucial in assisting GAS to avoid destruction by neutrophils [99]. Sequencing one of the M1 strain has revealed that, the *speA* positive genotype had a 36 kb recombination region involving genes encoding several GAS toxins including streptolysin O and NAD glycohydrolase [100]. These studies has shown that a highly invasive clone has emerged, which is responsible for much of the severe diseases across the world.

3.9. MOLECULAR EPIDEMIOLOGY

Epidemiology is defined as the study of the distribution and determinants of infectious diseases. The study of molecular epidemiology uses molecular techniques to investigate the distribution and determinants of infectious diseases [101]. The infectious disease distribution in a community may be determined by the pathogen's own genetic material that evolved in response to hosts antimicrobial machinery as well as to the environment in which the pathogen resides. The practical goals of molecular epidemiology are to identify the organism responsible for the disease and determine their physical sources, their biological relationships, and their route of transmission and those of the genes responsible for their virulence, vaccine relevant antigens and drug resistance [102]. Therefore, the study of molecular epidemiology of any pathogen should include the genetic factors that determine the specific pattern of disease [103]. The general definitions as recommended by the European Study Group on Epidemiological Markers [104] and the Molecular Typing Working Group of the

Society for Healthcare Epidemiology of America [105], Which is followed in this study are given below.

Isolate: A population of microbial cells in pure culture derived from a single colony on a solid medium and identified to the species level.

Strain: An isolate or group of isolates exhibiting phenotypic and/or genotypic traits belonging to the same lineage, distinct from those of other isolates of the same species.

Clone: An isolate or a group of isolates descending from a common precursor strain by nonsexual reproduction exhibiting phenotypic or genotypic traits characterized by a strain typing method to belong to the same group.

Type: A specific and discrete unit of information or character belonging to a strain displayed upon application of a strain typing procedure.

Outbreak: An acute appearance of a cluster of illness caused by a pathogen that occurs in number in excess of what is expected for that time and place.

Epidemic: Occurrence of a cluster of illness caused by a pathogen in numbers in excess of what is expected for that time and place. An epidemic can be an outbreak, although outbreaks usually refer to an abrupt appearance of a cluster of illness over a brief period of time in a population, geographic area, or institution.

Pandemic: An epidemic, if it spreads globally and lasts for years (e.g., AIDS)

Endemic: Occurring in a population or a geographic area or institution with no apparent deviation (increase or decrease) from the expected number of cases over time for that place.

3.9.1. Molecular typing

Distinctive characters, called epidemiological markers, are scored by typing systems that are designed to optimize discrimination between epidemiologically related and unrelated isolates of the pathogen of interest. The ability to discriminate between genomes is essential to several disciplines of microbiology [106]. Genetic typing is the means by which the microbiologist is provided with the ability to discriminate between catalogue microbial nucleic acid molecules. Genotyping methods overcome the problems of typeability, reproducibility and discrimination and enable the establishment of large databases of the organisms characterized [105, 107]. The most detailed form of typing to study the genetic relatedness at the subspecies level would be the determination of the complete sequence of the bacterial genome, which may not be always possible [104]. Hence the decision to use a molecular biology technique should be based on the following consideration; i) simplicity, ii) high throughput, iii) cost, and iv) appropriateness. Speed may be another criterion to consider in special circumstances, such as epidemiologic investigation involving clinical management of patients in a nosocomial setting or in an outbreak or if the epidemic is unknown. The severe acute respiratory syndrome (SARS) pandemic of 2003 is an example of need for speed in a molecular strain-typing test which led to the differentiation and identification of a new strain of corona virus as the possible etiologic agent [108].

3.9.1.1. Criteria for evaluating typing methods

There is no gold standard by which different typing methods can be judged and consequently the terminology usually used to describe the operating characteristics of the laboratory tests such as sensitivity and specificity are not strictly applicable for typing methods. Typing systems are often proposed or applied without adequate

knowledge of their effectiveness. The results of typing systems must therefore be considered relatively to the available epidemiological data or to the results of other typing systems [44].

3.9.1.2. Purpose of epidemiological typing

Typing methods are essential because several species of microorganisms share overlapping niches or thrive under identical environmental conditions [106]. Typing provides the means to discriminate between catalogue microbial nucleic acids. In general terms, epidemiological typing systems are used to study the population dynamics and spread of bacteria from a single host to world wide population ecosystems [104]. The specific purpose of the epidemiological typing is to address the epidemiological problems, which include the study of i) Determining dynamics of disease transmission ii) Detecting the modes of disease transmission iii) Addressing hospital infectious disease and out breaks iv) Identification of risk fractions in infectious diseases v) Distinguishing Pathovars and non-pathovars vi) Stratification of Typing data vii) Bacterial Population genetics.

i) Determining dynamics of disease transmission

Given an opportunity, infectious agents will spread and does not obey national boundaries. Therefore, one obvious application of a strain typing technique is in tracking pathogens geographically. The application of molecular biology techniques requires detailed attention to both laboratory experimental and epidemiologic study designs. The laboratory experimental design includes criteria used to collect and transport specimens, isolate and identify microorganisms, extract and quantitate nucleic acid materials, prevent contamination by other organisms and DNA, storage of isolates and DNA, standardize protocols and select appropriate control strains.

Epidemiologic study designs include proper selection of study subjects from which the isolates are obtained with appropriate control subjects and statistical tests. The questions need to be asked before a study is; How does an organism get introduced into the community, and how and why does it spread? What are the risk factors for infection and transmission in a community and how is the infection maintained in a community of hosts? How and why does a particular strain suddenly emerge as the predominant strain in a community, and why does it disappear? How do these fluctuations in clonal distribution affect the prevalence of drug resistant disease and the spectrum of clinical manifestations in a community? How do we use the knowledge we gain about the dynamics of disease transmission to implement effective disease control programs? How do we measure the effect and impact of the intervention made based on the knowledge acquired from the use of a particular laboratory test? These are some of the questions that need to be asked before using the molecular biology techniques for determining dynamics of disease transmission [103].

ii) Detecting Modes of disease Transmission

Studies about the transmission of infectious agents are a major component of infectious disease epidemiology. Most pathogens do not rely on a single host to perpetuate themself. To ensure their long term survival, transmission to new hosts is a necessary biological activity of these pathogens. Pathogens have developed distinct strategies to facilitate the mode of transmission within the context of their environmental setting. Studies that attempt to identify a pathogen's genetic determinants of transmission have focused on factors responsible for the pathogen to establish either a type of infection in a host that enhances opportunities for transmission to new hosts or a unique niche or reservoir that becomes a major source of transmission [103].

iii) Addressing hospital infectious disease and out breaks

The issues related to infectious diseases problems in hospital or a closed settings are distinct from those related to field or community. It is very difficult to characterize outbreaks from the episodes and even clusters of infectious diseases observed in hospitals. One frequent question asked in the hospital is, “Is the cluster of an illness observed in a particular ward or service caused by a single pathogen or a closely related set of pathogens, or is it due to changes in the host factors such as immunosuppression or antibiotic exposure shared by these patients which predispose them to be infected with a commensal organism from their or some one else’s body, or with saprophytes from the environment? It is also becoming recognized that organisms traditionally associated with hospital infections have begun to appear in the community settings. Hence, a rigorous epidemiologic analysis using molecular techniques will only help to answer these types of questions [103].

iv) Identification of Risk Fractions in Infectious diseases

The opportunity to assess the risk factor is provided from the identified outbreaks. A case control study can be designed to implicate a source or reservoir of an infectious agent responsible for the outbreak. But most of the diseases are sporadic especially in the case of food borne diseases [109]. In such situation, a strain that has previously been characterized shown to be linked to a particular vehicle or reservoir may provide a clue about the sources of infection for sporadic cases of an illness. Such information can then be used to generate a hypothesis and to design cross sectional or prospective studies to determine the proportion of sporadic infections caused by such a strain. Thus an attributable risk fraction can be calculated for a particular risk factor for the sporadic infections in a community [103].

v) Distinguishing Pathovars from Non-pathovars

A pathogenic variant of an organism can be referred to as a pathovar or pathotype. The defining feature of a pathovar in a human is that, it colonizes the same site in which non-pathovars establish themselves as commensal, but it needs additional mechanism to cause severe disease [103]. While traditional methods are available, molecular techniques are also limited in their ability to discreately distinguish the pathotypes that colonize the pharynx of the human host [23]. Even though many environmental and host related factors have been attributed for the risk of diseases onset, significant questions remain regarding how GAS causes human diseases [54]. The specific factor responsible for the pathogenicity of a severe disease should be a common marker, which may or may not be present in other isolates but exclusively should be significantly associated with the isolates of the particular disease manifestation [110].

vi) Stratification of Typing data

Molecular typing methods can group data related to infectious agents by subtypes. The subtype information itself can be used to group the isolates into discrete units. Then the patients infected with the discrete units of the isolates can be reclassified and look forward to develop new case definition. This would be particularly useful for patients who have sporadic illness in a community or hospital. Refinement of data into discrete units creates the opportunity to perform risk assessment using an analytical epidemiologic study design. This provides an opportunity to generate new hypotheses about cause-effect relationship. Stratification of data is one of the major applications of molecular typing system [103]. The ability to stratify data creates opportunity to identify new risk factors.

vii) Bacterial Population genetics

Several typing techniques have been applied to understand the intra-species population structure, and to determine the phylogenetic relationship within and between species [36, 111, 112]. These studies have also been used to determine the genetic variability and the evolutionary processes responsible for it [106]. Fundamental mechanisms that have been studied extensively by the use of these typing techniques to understand genetic variability are mutation, hyper-mutation, genetic recombination and natural selection [45]. Recently, the application of the evolutionary principles to study the population structure using the sophisticated molecular biology tools has increased the understanding of the bacterial population structure [113].

3.10. CLASSIFICATION OF TYPING METHODS FOR GAS

Traditionally, bacterial differentiation has relied on phenotypic characteristics. However, the techniques based on phenotypic features occasionally show poor reproducibility, insufficient discrimination and non-typeability of many isolates. Genotyping does not have these inherent problems and provides reliable and more discriminatory results to study every isolate. Direct genetic information is provided by sequencing of the gene of interest, which have increased the knowledge of GAS during the past. Recently, due to the advancement in sequencing and comparison of the whole genome, associated with severe diseases have resulted in the understanding of molecular basis of GAS pathogenesis [6]. Whole genome is the best method to discriminate between strains, but it is not always feasible for all research groups due to many reasons. So, many of the studies are restricted with less expensive and available techniques according to their economical capacity. The molecular typing

systems can be broadly classified into three groups based on **i)** direct comparison of nucleotide sequences, **ii)** gel electrophoretic finger printing patterns **iii)** hybridization matrix patterns (e.g., microarray) [103].

3.10.1. Phage typing for GAS

Phage typing classifies bacterial organisms according to susceptibility of the bacteria to lysis by a panel of bacteriophages. This was the method used in earlier days before the establishment of the sophisticated and more reliable methods for GAS [58]. A strong correlation between three of the four serological types classified by phage specificity and those identified by M protein serotyping, suggesting that patterns of lysogenization were related to serologic variants of the M protein [114].

3.10.2. Serotyping For GAS

3.10.2.1. M typing

M typing is a highly discriminatory method which is based on the antigenic variations in the M protein (Fig. 1) which is a surface protein that was well studied than any other molecules of GAS. The N-terminus of the M protein has been demonstrated to contain the type specific moiety and is recognized by type specific sera [115, 116]. Accumulated evidence indicates that the M protein putatively responsible for at least 100 known M sero-specificities of GAS, identified as a result of the antigenic variation in the N-terminal regions of M proteins. There are limitations in M serotyping, including ambiguities in the results, difficulty and high cost in preparing and maintaining a comprehensive set of typing sera, lack of reactivity with available antisera, unavailability of antisera for newly encountered M types and difficulty in interpretation [19, 20, 42, 107]. M typing antisera are not commercially available and

are very tedious to prepare and maintain even for the reference laboratories [117]. Therefore this method was replaced by molecular characterization of the M protein gene (*emm*) known as *emm* typing [20]. A non-serological method for the identification of specific M proteins using oligonucleotide probes has also been described for typing GAS [118].

3.10.2.2. T Typing

T typing is an agglutination pattern technique which requires antisera production using the extracts of T protein. T protein is a surface antigen which exhibit variations making it suitable for typing. This typing is generally used as an alternative typing method for M typing [5, 58]. The role of T protein in virulence is not known until recently [119]. It is common for a single GAS strain to have more than one T antigen and a single T type/T pattern may be found in strains of different M types, while strains of a given M type may have one or another totally unrelated T antigen. Hence, T pattern is not an M-type specific marker [58, 119].

3.10.3. Genotyping Methods for GAS

Serotyping was a universally used phenotypic method for discrimination of GAS. However these methods are labor intensive, require skill and are applicable to a narrow range of types for which they were developed. Genotyping does not have these inherent problems and provide reliable and more discriminatory results. Direct genetic information could be obtained by sequence determination, but this is usually limited to small part of the genome. Therefore, it is common practice to try with other molecular techniques [104].

Most of the techniques developed before the advent of *emm* typing were based on the restriction analysis as an alternative to serotyping. These techniques require isolation of the genomic component followed by the use of restriction enzymes which cut the DNA at specific sites. DNA fragments of different lengths are produced, depending on the distribution of restriction sites in the genome. These fragments are resolved according to their lengths on electrophoresis. Based upon the different combinations of the restriction enzymes, the gel types and the electrophoretic conditions, fragments between different size ranges can be studied. Pulsed-field gel electrophoresis (PFGE) for bacterial DNA provides genotypic information with a resulting pattern of DNA fragments to assess strain similarity among group A streptococcus has been demonstrated [120-122]. Another method, Restriction Fragment Length Polymorphism (RFLP) which is based on the restriction analysis has shown to be useful [123]. SOF typing is generally attempted when M typing is unavailable or difficult to perform. Strains belonging to certain M types produce opacity factor (OF), which is a type-specific enzyme. The specificity of OF parallels that of M protein types [124]. Therefore, OF typing can determine the M types of GAS strains by using an OF inhibition test.

3.10.3.1. Ribotyping

The RFLP method has been also described by using restriction enzyme patterns of ribosomal RNA gene to characterize GAS [125]. Published reports have suggested that this method was less effective than the standard total DNA based RFLP method in discriminating subtle differences between GAS isolates [126, 127].

3.10.3.2. Vir typing

A method used for characterization of GAS isolates based up on long PCR to amplify 4-7 kb of the entire virulence regulon (vir), followed by restriction digestion of the amplified product with HaeII. The cleaved DNA was then subjected to agarose gel electrophoresis. Vir typing is useful for classifying large number of GAS isolates into distinct genotypes and is suitable for studies in regions where GAS is endemic and predominately M-nontypable [63, 117, 128]. Vir typing is highly concordant with *emm* sequence [107].

3.10.3.3. M protein gene sequence typing (*emm* typing)

The *emm* typing is based on the sequence analysis of the portion of the *emm* gene that encodes M protein serospecific hyper variable region (Fig. 1). Thus, the problems associated with M serotyping were solved by extending it to the typing of its gene by sequencing. This system of *emm* typing, relies upon the use of two highly conserved primers, primers 1 and 2, to amplify a large portion of the *emm* gene [20]. There are more than 170 *emm* types and 750 subtypes from GAS have been characterized [22]. A sequence is considered to be a given *emm* gene or sequence type if it has greater than or equal to 95 % identity over the first 160 bases of sequence obtained with primer 1 or *emm* seq2 with that reference *emm* gene sequence in the CDC database, allowing for one frame shift of up to seven codons or one in-frame insertion or deletion of up to seven codons [19]. In general, there is a good correlation between the M serotype and the *emm* gene type [17, 129].

3.11. POPULATION GENETICS SYSTEM FOR GAS

Population genetics is the study of structure, function and inheritance of genes and genomes in natural populations [45]. The phenotype of living organisms is always a result of the genetic information that they carry and pass on to the next generation and of the interaction with the environment. Mutations in a gene that are passed on to the progeny and that coexist with the original gene result in **polymorphisms**. Evolution is always a result of changes in **allele frequencies**, also called **gene frequencies**. Whereby some alleles are lost over time and other alleles increase their frequency. If a particular allele is more fit than its polymorphic allele in a particular environment, it will be subjected to **positive selective pressure**; if it is less fit, it will be subjected to **negative selective pressure**. **Natural selection** is the effect of positive and negative selective pressure, accounts entirely for the changes in frequencies. When random fluctuations determine in part for the allele frequencies, chance events play a role and allele frequencies or population distributions cannot be entirely predicted [130]. The ***neutral theory of evolution*** follows a more stochastic approach. Kimura (1983) advocated that the majority of gene substitutions were the result of random fixation of neutral or nearly neutral mutations. Positive selection does operate, but the effective population size is generally so small in comparison with the magnitude of the selective forces that the contribution of positive selection to evolution is too weak to shape the genome. According to the neutral theory, only a small minority of mutations become fixed because of positive selection. Organisms are generally so well adapted to the environment that many non-synonymous changes are deleterious and, therefore, quickly removed from the population by negative selection. Stochastic events and substitutions, which are leading to fixed mutations, are mainly the results of random genetic drift [131].

Population genetics about the pathogens could reveal the differences that exist in concordance with the disease manifestation or ecological niche [132]. The population structure for GAS has been studied first by the sequence obtained from the hyper variable region of the *emm* gene which have shown a non-congruent relationship between the *emm* sequence based phylogeny and the over all genetic relationship among strains determined by Multi-locus enzyme electrophoresis [111]. Latter the *emm* family pattern (described below) was used to study the population structure which could classify GAS strains with specificity for site of isolation [133]. Now a days the MLST based studies are being conducted to determine the population structure of GAS [134]. These studies provide insights into the nature and extent of GAS variation and describe how these variations influence the population genetic structure of GAS lineages. The complexity of GAS relationships revealed by this analysis has significant implications for understanding evolutionary events generating strain diversity and the epidemiology of GAS diseases.

3.11.1. *emm* family Pattern

M protein genes (*emm*) and M-like genes share a high degree of sequence homology within the signal peptide and conserved C-terminal regions. The *emm* like genes must have occurred by gene duplication from an ancestral gene and are located in tandem in a single locus [135]. The content and relative chromosomal arrangements of the 4 *emm* SF genes in the vir regulon (*mga*) are found to exist in 5 basic patterns, A–E with very few exceptions. The phenomenon of tissue specificity is further correlated with the genetic marker exerted by the arrangement pattern namely A through E (Fig. 2) [30]. The *emm* family pattern A-C is reported to be throat specific, D is skin specific and E is common for both the sites [41]. The genes comprising the subfamily

may be mapped on the GAS chromosome by PCR amplification with specific primers [30, 56, 135, 136]. In western countries, pattern A-C is associated with throat and pattern D is associated with skin, from where the method was developed [41], but this association of the *emm* family pattern with the site of infection is very arbitrary and not consistent with many other reports [16, 30]. The *emm* Pattern E is always associated with its ability to produce serum opacity factor (SOF) [42].

3.11.2. Multi Locus Sequence Typing (MLST):

The recent development of multi locus sequence typing (MLST) schemes is a new approach to molecular epidemiology that can identify and track the global spread of any bacterial pathogens. Different methods may be appropriate for investigating local and global epidemiology, and the isolates assigned to the same molecular type are likely to be descended from a recent common ancestor, but the relatedness between the isolates that belong to different types can not be known in other methods. In contrast MLST is a highly discriminatory unambiguous method for identifying clusters of isolates with identical or closely related genotypes and is highly suitable for the analysis of bacterial populations [137]. The slow accumulation of variation within 7 housekeeping loci provides framework for long term global studies and reveals information concerning evolutionary relationship [134]. Moreover the sequence data can be compared readily between laboratories, such that a typing method based on the sequences of gene fragments from a number of different housekeeping loci [multilocus sequence typing (MLST)] is fully portable and data stored in a single expanding central multilocus sequence database can be interrogated electronically via the Internet to produce a powerful resource for global epidemiology

[138]. The MLST scheme has been associated with *emm* types and *emm* family pattern based genetic structure for GAS [139].

3.11.3. Genome sequencing and Microarray analysis for GAS

The complete genome sequence of GAS M1 serotype was first reported in 2001 by Ferretti et al. [140]. Since then many more strains of GAS were sequenced and much information about the genetic organization of GAS is being coming out. Complete genome sequencing of serotypes M1, M3 and M18 has revealed that the majority of differences among them correspond to prophages and other exogenous elements. All strains studied thus far have a genome size of _1.8 to 1.9 Mbps with five or six highly conserved rRNA operons, and a core group of proven and putative virulence genes [47]. The strains that were sequenced contained many prophages (polylysogenic) each encoding for one or two putative virulence factors (Fig. 3) [47-49]. Genetic diversity in M18 strains was investigated by several state-of-the-art methods, including DNA-DNA microarray and whole-genome PCR tiling [141]. The core genome of all strains in M18 serotype was identical in their gene content. That is, variation in prophage content accounted for all differences in gene content between strains. An iterative expression microarray method was used to gain enhanced understanding of the molecular mechanisms used by GAS to interact with the host [142].

DNA microarray technology allows a parallel analysis of RNA abundance and DNA homology for thousand of genes in a single experiment [143]. Over the past few years, this technology has been used to explore transcriptional profiles and genome differences for a variety of microorganisms, greatly facilitating the understanding of microbial metabolism. With the increasing availability of complete microbial

genomes, DNA microarray are becoming a common tool in many areas of microbial research, including microbial physiology, pathogenesis, epidemiology, ecology, phylogeny, pathway engineering and fermentation optimization [143].

3.12. PATHOGENESIS

GAS is exclusively a human pathogen but the pathogenesis varies from strain to strain, and not all the strains cause RF. So the identification of the difference in the pathogenesis becomes very important. The introduction and propagation of GAS on the surface of the mucous membrane or skin within these tissues results in destruction of host cells, and is accompanied by an intense inflammatory response [144]. Such infections can range from mild and self limiting infections of the throat or skin to infections which involve deeper layers of tissues. GAS also has the capacity to penetrate into the blood stream and cause life threatening sepsis and multi-organ failure. Non-suppurative sequelae of GAS infection including RF and Glomerulonephritis are also due to adaptive pathogenic mechanisms of this organism [145].

To cause wide spectrum of disease, GAS has adapted sophisticated strategies and complex regulatory mechanisms that enable it to thwart host defenses and successfully colonize, thrive and persist in the host [6]. Advances in genome sequencing, microarray technology and proteomic analysis, in combination with the development of more suitable animal models, have markedly increased our knowledge of the mechanisms underlying GAS pathogenesis. This regulatory system controls the expression of GAS genome differently at different conditions which is very essential to control the virulence mechanisms at different stages of infection. The pathogenesis of GAS has been divided into three stages (Fig. 4). The first step of GAS

in the interaction with its host is the **adherence**, and the second step of its infection is **avoidance of the host immune response** resulting in proliferation and finally the **invasion** into the tissues causes the manifestation of the disease which may be even an immunological shock or autoimmunity [4, 6].

3.12.1. Adherence: The interaction of the infecting GAS with the receptors on the surface of the host cells has been extensively investigated [4, 146]. The attachment to the pharyngeal or dermal epithelial cells is the most important initial step in colonization of GAS to establish successful infection. Without strong adherence mechanisms, GAS could not attach to host tissues and would be removed by mucous and salivary fluid flow mechanisms and exfoliation of the epithelium. In skin, a site of previous damage may be important in overcoming the dermal barrier. Adherence has been demonstrated as an initial weak interaction with the mucosa which is followed by a second adherence event which confers tissue specificity and high avidity [147].

In early studies, M protein was suggested as the adhesin which is necessary for the attachment to the host cell [148]. The M protein is also important for attachment to keratinocytes in skin infections [149]. A rat model of infection has demonstrated that M protein was required for the persistence of GAS infection [150]. Later on the Lipoteichoic acid (LTA) was identified as the adhesin by reacting with molecules on the streptococcal surface through its negatively charged polyglycerol phosphate backbone and positively charged residues of surface proteins [4, 151]. The lipid moiety of LTA projected outward and interacted with fatty acid-binding sites on fibronectin and epithelial cells [152]. After that, fibronectin was identified as the epithelial cell receptor binding LTA [153]. Evidence suggested that LTA accounted for approximately 60 % of adhesion to epithelial cells, indicating that other adhesions

wee involved. Later on, the multiple adhesions, of at least 11 different proteins were described for GAS, including M protein [4].

3.12.2. Immune evasion: The successful establishment of infection by GAS in the host pharynx needed numerous mechanisms for evading the host innate immune response. The presence of capsule and M protein are considered as the virulence factor conferring antiphagocytic properties upon the streptococcal cell [154, 155]. Streptococcal C5a Peptidase (*ScpA*) is the other immune modulating protein that cleaves C5a of the complement mediated immune system [156]. Streptococcal inhibitor of complement (*sic*) also inhibits the complement mediated membrane attack complex [157]. The Molecular mimicry and antigenic variation are other means by which GAS evade from the immune response of the host.

3.12.3. Proliferation: Development of GAS pharyngeal infection occurs following transmission of relatively small numbers of organisms from an infected or colonized host. Following adhesion, proliferation in the oropharynx triggers the signs and symptoms of pharyngitis. Correlation of organism density with pharyngeal evaluation demonstrated that GAS proliferation preceded the clinical development of pharyngitis due to the lack of nutrients needed for proliferation. The concentration of glucose in fluid lining the human oropharynx is too low to support GAS growth, which shows that other sources are needed for proliferation [158]. GAS produces many extracellular products that favor proliferation by damaging the host cells.

3.12.4. Intra-cellular Invasion: Though GAS was known as an extracellular pathogen, many reports shows that the organism may invade and persist within epithelial cells [159]. The exact role of this event is not clear. Many of the proteins involved in invasion also participate in adherence, including Fibronectin Binding

proteins (FBP), M protein and Streptococcal collagen-like proteins. The host signaling pathways mediating GAS invasion begin with bacterial binding to cell surface integrins, which results in the cytoskeletal rearrangement and GAS internalization [160]. The ability of GAS to invade and persist intracellularly has been associated with penicillin treatment failure and recurrent pharyngitis, but a definitive link between GAS invasion and pharyngeal pathogenesis has not yet determined [161].

3.12.5. Autoimmunity and Systemic Toxicity: Many studies support and demonstrates the hypothesis that ARF has an autoimmune origin which is due to the sharing of epitopes between the human host and the infecting GAS, a phenomenon called Molecular Mimicry [4]. Many cell wall antigens have been proposed as the trigger of auto-antibodies, including group A carbohydrate N-acetylglucosamine, M protein, 60 kDa actin-like protein, 67 kDa protein and collagen [162-165]. Although these studies provide evidence that anti streptococcal antibodies could be potentially damaging to cells or tissues of the host, multiple factors have to be considered and no specific factor has been definitively proved with clinical correlation.

The systemic toxicity is associated with the **pyrogenic exotoxins** exhibiting superantigenic activity. Pyrogenic exotoxins binding to the beta chain (Vβ) of a characteristic set of T-cell receptors and also to the MHC class II molecule expressed on B cells, monocytes and dentritic cells. Binding to the T-cell receptor and to MHC class II molecules causes the T cells to proliferate with subsequent high expression of inflammatory cytokines [166]. By virtue of their potent immune-stimulatory capacity, superantigens have been implicated as central mediators of the systemic effects seen in severe invasive GAS infections [167].

GAS phages also encode non-pyrogenic exotoxins that are believed to contribute to host-pathogen interactions, including newly discovered phospholipase (designated as Sla) identified during sequencing the genome of M3 strain, an organism recovered from patient with streptococcal toxic shock syndrome. This enzyme has sequence homology to potent snake venom, phospholipase. The role of Sla in the patho-physiology of GAS infection has not been elucidated. But it was speculated that it contributes to inflammation and coagulopathy and may participate in the initial stages of GAS interactions with the host cells [168].

3.12.6. Phages in the pathogenesis of GAS: Many other sequelae of GAS infection were also reported which were not directly proved in the laboratory but have strong association with GAS. The severity of all the sequelae is through immunologic complications induced by the involvement of various components and extracellular products of GAS [4]. Advances in the knowledge about molecular basis of GAS pathogenesis have led to the hypothesis that more virulent strains could emerge through recent acquisition of novel virulence factors [8, 47, 94]. Many of the virulence traits can evolve in quantum leaps through horizontal gene transfer rather than accumulation of single nucleotide mutations [14]. A major driving force in the emergence and evolution of pathogenic isolates is the horizontal acquisition of virulence factors [47]. Most of the virulence gene inheritance was reported to be associated with bacteriophage lysogenic conversion and therefore provide a potential molecular basis for characteristic variation in the disease manifestations [29]. Complete genome sequencing has revealed that majority of the differences in GAS correspond to prophages and other exogenous elements. The strains that were sequenced contains many prophages (polylysogenic) each encoding for one or two putative virulence factors with very few exceptions [47, 81, 169]. These Phage

associated Virulence factors (PAVF) thought to contribute significantly for the pathogenesis and diversification of GAS.

Few differences in gene content were identified between the strains but the largest overall difference was 3.0 % of the genes. Importantly, prophages were responsible for all variations in gene content among the 36 strains. Similar study was also conducted in 255 M3 invasive isolates [170] which provided a framework for examining the relationship between a strain's phage genotype and the patient's disease phenotype. Comparative genomics with three sequenced strains indicated that they share approximately 90 % of their chromosomally encoded genes and the remaining 10 % of the genomes are composed of prophage-encoded genes and other similar elements [47]. Though the prophage comprise a minority of the total GAS genome, the are responsible for up to 74 % of the variation in gene content within open reading frames (ORF) of different GAS strains, indicating that phages have contributed significantly to the evolution of GAS pathogenicity [54].

3.13. VIRULENCE DETERMINANTS

3.13.1. M protein

M protein is a helical coiled-coil surface protein anchored to the bacteria by its carboxy terminal end. This region is divided into three repeating regions; A, B, C and D repeat blocks where all of them differing in size or sequence (Fig. 1). The sequence from the N terminus to the first A repeat is a short non-repeat domain, which is hyper-variable and serotype specific [171]. The N-terminal region, which extends from the surface of the cell wall, contains the non-helical region, A and B repeat regions. The A repeat region is highly variable while the B repeat region is less variable. The C repeats are conserved but varies among two different classes of M types. The D repeat

region is highly conserved and the proline / glycine rich region is located within the peptidoglycan layers [172]. During infection, antibodies against the N-terminal part of this protein are rapidly generated, leading to protection against GAS infection of the same type [173]. Therefore, M protein undergoes frequent size variations due to high selection pressure [174]. GAS has also been classified into different serotypes based on the protein that can be extracted from the bacteria with boiling hydrochloric acid. It is an important protein for the virulence, which helps in adhesion and even invasion in to the host cells [175]. It is under the regulatory control of Mga regulon known to control several other virulence related genes [176]. However, the ability of M proteins to elicit host-crossreactive antibodies is thought to be involved in post-streptococcal autoimmune diseases [163].

M proteins have been divided into class I and class II molecules. The division of the M proteins into two classes is based on their reaction with antibodies against the C repeat region of M protein. Class I M proteins are reported to contain a surface-exposed epitope on whole group A streptococci that reacts with the antibodies against the C repeat region. Streptococcal strains containing the class II M proteins do not react with these antibodies and do not contain the class I epitope [177, 178]. In addition, the class I M protein serotypes were opacity factor negative, while the class II serotypes were opacity factor positive. One of the studies shows strong correlation between serotypes known to produce rheumatic fever and the presence of the class I epitope. A study shows that the antigens to Class I antigen is identified in both RF patients and the controls [179]. Another study show serologic evidence that rheumatic fever patients were recently infected with a class I group A streptococcal strain [180].

The ability of GAS to persist in the infecting tissues is primarily due to the M protein, which is a major virulence factor that inhibits phagocytosis, [4, 154]. Absence of the *emm*gene allowed rapid phagocytosis of the streptococcus. Introduction of the *emm*gene into an M negative strain converted it to an M-positive strain and restored resistance to phagocytosis [181]. The antiphagocytic activity of M protein is due to its binding of complement regulatory protein factor H and fibrinogen [182]. The binding of fibrinogen also leads to activation of plasminogen, which is then converted by streptokinase that activate plasmin [183]. The immune response against the M protein leads to production of protective antibody which promotes phagocytosis and killing as well as antibodies which may react with host tissues due to molecular mimicry [154, 184]. Immunity to the M protein is protective against group A streptococcal infection and has led to the study of M protein vaccines [185].

3.13.2. Hyaluronic acid capsule

The Capsule in GAS is a major virulence determinant along with the streptococcal M protein. It may be an important adherence factor in the pharynx on epithelial cells [186]. The mechanism may be due to the physical barrier of the capsule in preventing access of phagocytes to opsonic complement proteins bound to the bacterial surface. Studies provide evidence that the capsule is a major virulence determinant involved in resistance to phagocytosis [155, 187]. It is composed of a polymer of hyaluronic acid containing repeating units of glucuronic acid and *N*-acetylglucosamine. Acapsular mutant strains were altered in their virulence and colonization capacities in animal models [188, 189]. mimicry enables GAS to avoid detection by the host immune system [4]. Epidemiologic evidence linking highly mucoid strains with rheumatic fever and severe invasive streptococcal disease suggests that the capsule could play an

important role in invasive infections in humans [58]. Acapsular mutants of serotypes M18 and M24 had drastically reduced virulence in mice after intraperitoneal challenge [189, 190]. But acapsular strains enter cultured human keratinocytes with higher efficiency than mucoid strains, which suggests that capsule may interfere with GAS internalization by blocking the interaction between adhesions on the bacterial surface and host cell receptors [191].

3.13.3. Streptolysins

GAS and other streptococci produce two distinct hemolysins, one which is hemolytic in the reduced form and the other is oxygen stable which is released in serum containing media, known as streptolysin S (serum soluble) and streptolysin O (oxygen labile). Streptolysin O is a protein of 60 kDa that binds to cholesterol on the surface of erythrocytes, polymorphonuclear leukocytes and platelets resulting in lysis [192]. The exact biological significance is not known. However, intravenous injection of streptolysin O into mice, rabbits and guinea-pigs causes death immediately. In murine model, GAS infected with *slo* mutant had decreased mortality compared to wild type [193]. It has been also suggested that streptolysin O plays a role in the development of RF [194]. Serum antibodies to streptolysin O can be demonstrated after streptococcal infection, with a frequency that increases with the severity of the infection. The highest titers are found in the patients with RF [195].

3.13.4. Streptokinase (ska)

Streptokinase produced by GAS, also known as fibrinolysin is a 46 kDa secreated protein binds to human plasminogen which is than converted to active plasmin. Plasmin lyses blood clots and tissue barriers and thus facilitates the systemic dissemination of GAS [196]. The streptokinase from GAS encoded by *ska* gene,

which has been cloned and sequenced shows substantial genetic diversity giving rise to antigenic and chemical diversity [197]. It is found in most of the GAS isolates. The expression of *ska* has been associated with the pathogenesis of acute Glomerulonephritis and contributes to the virulence and invasion of GAS [198].

3.13.5. Streptococcal C5a Peptidase (SCP)

The C5a peptidase produced by GAS is a highly conserved and immunogenic enzyme with proteolytic activity located in the cell surface of all GAS isolates [199, 200].C5a peptidase in GAS is encoded by the *scpA* which is located in the Mga regulon. C5a peptidase acts on the C terminus of C5a component of the complement system and thus inactivates the chemo-attractant property, resulting in the impairment of phagocytes recruitment in phagocytosis [63, 201]. GAS M1 and M49 strains, inactivated for C5a peptidase enhances the clearance of these strains in intranasally challenged mice [202]. These studies suggest that C5a peptidase is also a promising vaccine candidate [203].

3.13.6. Serum Opacity factor (SOF)

Serum opacity factor is an approximately 1000 amino acid residue cell surface bound apoproteinase named for its property of rendering opaque nature to serum [204]. Serum opacity factor is a major fibronectin-binding protein and a virulence determinant of GAS. SOF has the fibronectin-binding activity that resides in the short C-proximal domain that is distinct from the large opacity-factor conferring segment [205]. Several studies have also suggested that SOF production has a high correlation with M protein class. The strains that do not produce SOF tend to possess class I M proteins and the SOF producing strains carry class II M proteins [177, 178]. The

presence or absence of *sof* indicates fundamental differences in the M protein structure and differences in the vir locus gene composition and arrangement [206].

3.13.7. Streptococcal inhibitor of compliment mediated lysis (SIC)

The streptococcal inhibitor of complement mediated lysis is a secreated protein which is encoded by the highly polymorphic *sic* gene present in the vir regulon, located nearer to the *emm* gene and C5a peptidase. This protein interferes directly with the complement cascade and is expressed in large amount in a very few virulent types such as M1 and M57 serotypes [207, 208]. SIC is a highly polymorphic and immunogenic protein, that exhibits high variations than any other streptococcal proteins, suggesting that *sic* is under strong seletive pressure in the mucosa of the upper respiratory tract [209]. SIC inhibits complement mediated cell lysis by incorporating it into the complement componants [207]. The knockout mutants of GAS shows a reduced ability of colonising in mice following intranasal innoculation. These data suggests that SIC may be involved in the initial stage of infection, in strains which harbor this protein [210]. Another protein named as distantaly related gene encoded by agene *drs* was also reported in *emm* 12 and *emm* 55 types, located far [211].

3.13.8. Collagen binding proteins of GAS (Cpa, CpaI)

The collagen binding protein is a 57-kDa protein encoded by the gene located in the FCT (Fibronectin – collagen – T antigen) region of GAS chromosome [212]. As it is named, it binds to the human type I collagen present in the basement membrane with high affinity. The contribution of GAS virulence is yet to be fully elucidated, however the binding of GAS to type I collagen may cause pathogenesis of bone and joints

[213]. CpaI is a novel collagen binding protein, for which much of the property is not known [214].

3.13.9. Streptococcal collagen like proteins (Scl, SclA, SclB)

Streptococcal collagen like protein is an extracellular matrix protein consisting of 348 amino acid residues, containing a signal sequence, a surface exposed hyper variable domain with a collagen like region. The number of contiguous Gly-X-X motif ranged from 14 – 62 in different M serotypes [215]. This gene is transcribed abundantly in the logarithmic phase of growth. Isogenic mutant M1 strains were not attenuated for mouse virulence as assessed by intraperitoneal inoculation but were less virulent when inoculated subcutaneously into mice. It also had reduced adherence to the human epithelial cells grown in culture, which shows that this protein partipates in adherence and hence play an important role in the pathogenesis of GAS infection [215].

Two other streptococcal collagen like proteins (SclA and SclB) were identified at the same time of the identification of the first protein [216, 217]. The SclA (also known as Scl1) is a cell wall associated protein of GAS with a surface exposed portion containing a distal variable region forming its globular domain, with a central rod-shaped collagen-like (CL) domain, which is composed of repeating GXY sequences. The transcription of the SclA gene (*sciA*) is positively regulated by Mga, and the expression occurs in the logarithmic phase [216], and the SclB gene (*sciB*) has a unique regulation [217].

3.13.10. Protein F (*prtf1, prtf2, prtf15*)

Protein F is an adhesin that binds fibronectin which is also known as SfbI. Protein F and Fibronectin each have two domains which was involved in binding to dermis and Langerhans cells [50]. Protein F was expressed in 75 to 80 % of the streptococci investigated [218]. The availablity of superoxide appears to be an environmental signal for Protein F expression which is regulated by rofA, a potential regulator of adhesion [219]. The protein F mediate adherence and inhibits C3 deposition on the GAS cell surface and thus confers resistance to phagocytosis. GAS strains that lack PrtF1 may have PrtF2 [220]. Antibodies aginst Protein F block GAS attachment and internalisation into epithelial cells [221]. Another adhesin protein F identified in M15 that binds fibronectin has also been identified and named as PrtF15 [222].

3.13.11. Fibronectin Binding Protein (Fba)

Fba, a novel fibronectin-binding protein from Streptococcus pyogenes, promotes bacterial entry into epithelial cells, and the *fba* gene is positively transcribed under the Mga regulator [223]. Fba plays an important role in the adherence to host cells, confirmed by a mutant strain having significantly reduced levels of adhesion and invasion in HEp2 cell lines [223]. It is located on the same chromosomal location within the fibronectin-collagen-T antigen (FCT) locus [224].

3.13.12. Fibronectin binding protein 54 (FBP54)

The streptococcal fibronectin-binding protein (FBP54) is a cell surface protein that lacks LPXTG motif, but associates with cell surface of GAS [225]. FBP54 mediates the adhesion of group A streptococci to human buccal cells and not on HEp-2 cell lines [226]. *fbp54* is represented more frequently among invasive-disease isolates in

Australia [227]. Systemic and mucosal immunizations with FBP54 induce protective immune responses against GAS challenge in mice [228].

3.13.13. *S. pyogenes* fibronectin binding protein (PFBP)

PFBP is a cell wall associated protein with 127 kDa with high sequence homology with Protein F2, with LGXTG motif [53].

3.13.14. Streptococcal Fibronectin Binding Protein (SFB, SFBI)

Sfb and sfb1 were major fibronectin binding proteins that mediates adherence and invasion to epithelial cells [229, 230].

3.13.15. Pyrogenic Exotoxins

There were 11 distinct exotoxins produced by GAS, including the streptococcal pyrogenic exotoxins (Spe) A to M except D and E. Among these, SpeB, SpeF, SpeG and SpeJ were encoded in the native chromosomal genome and the others were found in the prophage of sequenced GAS strains and detected among many other GAS isolates [231-233]. SpeG is the pyrogenic exotoxin named as it was first identified in Group G Streptococcus [234]. The pyrogenic exotoxins SpeA, SpeC, SpeH, SpeI, SpeK, SpeL, SpeM, streptococcal superantigen (SSA), were phage mediated [47].

3.13.16. Mitogenic Factors

Streptococcal mitogenic exotoxin Z (SmeZ) is a chromosomal gene located in the pathogenic island, associated with the regulatory region of Mga regulon [140, 235]. The DNases encoded by GAS also have been referred to as mitogenic factors (MF, MF2, MF3, and MF4) and streptodornases (Sda, Sdb, and Sdn). Various functions

have been ascribed to these putative virulence factors, such as promoting GAS dissemination in host tissues and assisting phage dispersal after bacterial lysis [236].

3.13.17. Secreted Phospholipase

GAS phages encode a secreted protein **phospholipase** (SlaA) identified in serotype M3 [170]. This protein contributes to the adherence to host pharyngeal epithelium [237]. The mechanism of action of SlaA appears to involve entry into host cells. A strain without *slaA* was significantly impaired in its ability to cause pharyngitis in the cynomolgus macaque, providing strong evidence of the key role of SlaA in GAS pharyngeal pathogenesis [238].

3.14. VACCINES

The development of a safe and efficacious vaccine would be a major drive to move forward, in order to combat the public health problem caused by GAS infection. Immunization or exposure of humans to microbial adhesions may induce antibodies which concentrate in the mucosal layer and block adherence and colonization at the mucosal epithelium. An ideal vaccine candidate should be confirmed for their function and immunogenicity, should be well conserved in the species, and preferably, have an essential role in the host-pathogen interaction [239]. The development of multivalent vaccine by incorporating amino-terminal M protein fragment from 26 different serotypes of GAS were shown to be broadly opsonic and bactericidal against majority of the types included [240]. There have been some reports that doubt the usage of the type specific vaccines due to high diversity of GAS [39]. The other approach is the development of a mucosal vaccine containing the conserved region of the M protein [173, 241] . Other vaccines candidates under trial include C5a peptidase, SpeB and Fibronectin binding proteins [200, 203, 221].

Despite many decades of research, there is still no licensed vaccine to protect against GAS infections. It is recognized that a successful GAS vaccine must be idle to protect humans from pharyngitis caused by extensive array of genetically diverse strains. Many novel surface and extra cellular proteins were also being studied [132]. Combined biochemical, proteomic and bioinformatics strategy to identify fragments of proteins displayed on the surface of GAS has revolutionized be genome scale strategies for discovering candidate vaccine [242].

4. SCOPE AND PLAN OF WORK

4.1. STUDY HYPOTHESIS

It has been our experience that GAS infections of the throat seen in rural school children are clinically milder while those seen in the hospital are much more acute and invasive. A pilot study done on 12 clinical strains isolated from acute pharyngitis in the hospital group, revealed 11 different *emm* types among them. Only one of these 12 types was seen among 100 GAS strains isolated from rural school children. Based on these observations, we hypothesize that the GAS strains causing infections in children seen in a school survey (community strains) are less virulent than those associated with clinical conditions seen in a hospital (invasive). We plan to test this hypothesis, by studying certain virulent profiles of these strains using molecular methods and comparing the community and invasive strains in the context of these characters. If our hypothesis holds well, we expect the pathogenic gene profile of these two sets of strains to be significantly different from one another.

The surface proteins are related to the adhesion of any bacteria to the host for the successful infection. So if there is truly any difference in the ability of GAS strains in infecting the throat or skin, we expect specific surface protein(s) associated with it. Hence a set of fibronectin binding proteins were analyzed among GAS strains recovered from throat and skin infections, in order to determine the factor(s) associated with tissue tropism.

4.2. RESEARCH QUESTIONS

1. What is the association of *emm* types of normal throat, pharyngitis throat and impetigo skin isolates?

2. Does the *emm* family pattern of GAS strains show any association between strains causing pharyngitis and those causing impetigo?

3. Is there any similarity between *emm* types circulating in our community? In other words, are they clustered together based on full *emm* gene sequences?

4. What is the gene profiles of the prophage associated virulence factors (PAVF) in the *emm* types circulating in the community? How do they differ from the invasive strains?

5. Is there any toxin or fibronectin binding protein gene encoded in the chromosome associated with skin or throat isolates?

4.3. STUDY DESIGN

<u>Phase I</u>: To characterize three different categories of GAS isolates, namely pharyngitis associated isolates, impetigo associated isolate and those colonize throat of asymptomatic children by *emm* typing

<u>Phase II</u>: To develop a cluster analysis system based upon *emm* family pattern and phylogeny

<u>Phase III</u>: To compare the virulence factors of *emm* types in relation to site and symptoms.

5. MATERIAL AND METHODS

5.1. SELECTION OF GROUP A STREPTOCOCCI FOR *emm* TYPING

Isolates of group A streptococci (GAS) selected for *emm* typing included the following:

a) Those isolated from throat cultures of children with GAS pharyngitis (PT isolates)

b) Those isolated from skin lesions of children with GAS impetigo (SK isolates) and

c) Those isolated from throat cultures of normal healthy children (NT).

5.2. ISOLATION OF GAS

GAS isolates selected for the study were recovered from school children aged 7-11 years and attending a rural school situated 10 Kms south west of Vellore town, during a prospective cohort study conducted among 237 school children, each year during March 2002 – March 2004. This project was supported by the Department of Biotechnology (DBT) in collaboration with National Institutes of Health (NIH), Bethesda, Maryland, USA under the auspices of INDO-US Vaccine Action Programme, co-sponsored by ICMR, New Delhi and NIH, USA.

Throat and skin cultures were collected from children with clinical symptoms of pharyngitis and impetigo on filter paper strips (FPS) and transported to the Laboratory within 3 to 4 hours as per standard procedures [58]. In the Laboratory, the FPS were processed for the isolation of beta hemolytic streptococci (BHS) according to standardized procedures [58]. BHS were characterized by identification of group specific antigen as given below.

5.3. IDENTIFICATION AND PRESERVATION OF GAS

GAS were identified by micro-nitrous acid extraction – coagglutination method using antisera and reagents prepared and standardized in our laboratory [243]. All GAS isolates were preserved by lyophilization and stored at +4°C until use.

5.4 CALCULATION OF SAMPLE SIZE FOR *emm* TYPING:

Sample size was calculated based on the prevalence of most common *emm* types encountered. In a preliminary study done in our laboratory this was 8 %. ie. of 100 GAS strains typed by *emm* gene sequencing, 33 different types were identified. Among these, five types, namely, 49, 77, 53, 1 and 118 accounted for 35% of the types. Among them, *emm* type 49 and 77 accounted for 8 % each. In a very similar study, McGregor et al., (2004) reported a prevalence of 9.56 %. Based on this difference of 1.56 % (taken as 2 %), the sample size was calculated by the equation,

$$n = \frac{4pq}{d^2}$$

where p = Prevalence; q = 1-p and d = Expected variation (ie. 2 %)

Ie. $$n = \frac{1.96^2 * 0.08 * 0.92}{0.02^2} = 707$$

Hence, the sample size was calculated as 707.

5.5 *emm* TYPING

The *emm* typing was performed according to the protocol described by the Centers for Disease Control and Prevention (CDC) (www.cdc.gov/ncidod/biotech/strep /assigning.htm) with some modifications. The *emm* typing relies upon the use of two highly conserved primers [20] to amplify a large portion of the *emm* gene. The primer 1 (*emm* seq2), specific for the conserved 3' portion of the signal sequence, which is located out side the 5' region (amino-terminal) of the gene encoding for mature M protein and the primer 2 targets a highly conserved sequence within the 3' portion (carboxy-terminal) of the gene encoding for mature M protein. The hyper-variable sequence encoding M serospecificity lies adjacent to primer 1, allowing for direct sequencing with out any loss in the sequence required for *emm* typing. This method could be used to type the unknown M serotype isolates also [17, 19].

5.5.1 DNA extraction

The cell lysate method was performed according to the protocol for *emm* typing described by CDC in their website (http://www.cdc.gov/ncidod/biotech/ strep/protocols.htm) with some modifications [244]. Single colony of the each GAS isolate was sub cultured on to 5 % sheep blood agar plate and incubated at 37°C in the presence of 5 % CO_2 for 16 hours. One loopful of the pure culture was picked up from the plate and suspended in 300 µl of 0.85 % of NaCl. This suspension was heated for 15 minutes at 70°C and centrifuged for 2 minutes at 13,000 RPM. The pellets were resuspended in 50 µl TE buffer (10mM Tris, 1mM EDTA, PH 8) containing 300 U/ml of mutanolysin and 30 µg / ml of hyaluronidase (Sigma-Aldrich, Bangalore, India) and were allowed to react for 30 minutes at 37°C. Samples were then heated at 100°C for 10 minutes in a water bath and cooled immediately in ice or stored in -20°C for

long term use. The supernatant of the lysate was centrifuged for 1 minute at 13,000 RPM contains the DNA. Not more than 0.5 µl is taken for 20 µl of Polymerase Chain Reaction (PCR).

5.5.2 PCR Technique and detection of amplicon

PCR master mix was prepared as given below for five samples.

10 µl 10X buffer containing 15mM MgCl2

2.0 µl of dNTP mixture (10mM)

2.0 µl of each primer 1 and primer 2 (70 pico mole/µl)

primer 1: TATT(C/G)GCTTAGAAAATTAA

primer 2: GCAAGTTCTTCAGCTTGTTT

0.5 µl Taq (3U/µl)

82 µl dH_20

The PCR programming conditions included initial denaturation at 94°C for 1 minute followed by 94 °C for 15 seconds, 49°C for 30 seconds and 72°C for 75 seconds repeated for 10 cycles and followed by the same conditions but with a 10 seconds increment for 72°C for each of the subsequent 19 cycles. The amplicons were then maintained at 4°C (http://www.cdc.gov/ncidod/biotech/strep/protocol_*emm*-type.htm).

The purity of the product was checked by running 5µl of the amplified product with 15 µl of loading dye in 1.5 % agarose gel with 10 µg/ml ethidium bromide set at 100 volts for 1½ hours along with an appropriate base pair marker.

The illuminated gel picture was documented using gel doc system (BioRad Laboratories, Hercules, CA). The single band in the range of 700 to 1200 base pairs showed that the amplified product was specific and pure.

5.5.3 Pre-sequencing clean up

a) The 15 µl of the remaining PCR product was mixed with 85 µl of Millipore water and filtered through Millipore PCR product cleanup kit.

b) Again filtered with 100 µl of Millipore water.

c) The filtered PCR product was eluted with 25 µl of the Millipore water

d) The purified product was checked in 1.5 % agarose gel electrophoresis for its quality and quantity.

5.5.4 Sequencing reaction

Depending upon the concentration of purified PCR products, 1µl to 2µl were used for 10 µl reaction with primers *emm* seq2 (5' TATTCGCTTAGAAAATTAAAAA CAGG 3') using the Big Dye terminator mix (Applied Biosystems, Warrington. UK). The cycling parameters were 25 cycles of 96°C for 15 seconds, 50°C for 12 seconds and 60°C for 4 minutes.

5.5.5 Post-sequencing clean up

To 10 µl of the reaction product, 30 µl of injection solution was added and then filtered using Millipore post sequencing cleanup kit (Millipore systems, Bedford, MA, USA) under suction pressure. Filtration was done by adding 30 µl of injection solution. The purified product was eluted using 25 µl of the injection solution and transferred to the sequencing tube for sequencing. The tube was covered with the septum and placed on the machine which was set ready as per manufacturer's

instructions. The purified products were subjected to automated sequence analysis on ABI 310 genetic analyzer (PE Applied Biosystems, Foster City, CA,USA).

5.5.6 Assigning *emm* type and subtypes

The extracted sequences were subjected to online homology search in the CDC database (http://www.cdc.gov/ncidod/biotech/strep/strepblast.htm) by Basic Local Alignment Search Tool (BLAST) analysis. The *emm* type was determined as described by the CDC method published in1996 [20] and modified as given in http://www.cdc.gov /ncidod/biotech/strep/assigning.htm the CDC website. For a given *emm* sequence, an *emm* type is assigned if it has more than 92 % homology with the first 90 bases of the *emm*type reference strain in the database. New types are identified on the basis of sharing less than 92 % sequence identity over the first 90 bases.

For the subtype assignment, the exact homology of the submitted sequences of a given strain with the trimmed 180 base entries corresponding to the first 50 residues of the mature M protein and the adjacent 10 C-terminal residues of the signal sequence of *emm* types in the database is used. If there is a perfect match to bases 31-180 along with 3 or fewer mismatches to bases 1-30, this indicates identification of a specific subtype. If there is any mutation in the DNA sequence corresponding to the first 50 residues of the mature M protein, it is considered as a new subtype.

5.6 DETERMINATION OF *emm* FAMILY PATTERN

5.6.1 Selection of strains

Representative *emm* types identified among PT, NT and SK isolates were chosen for determining the *emm* family pattern. The isolates were selected after completing the study on *emm*typing of GAS isolates selected for the study.

5.6.2 DNA extraction, PCR & Electrophoresis

- Single colony from the confirmed *emm* type stock was grown in 8 ml Todd Hewitt broth for about 16 hours.
- Pellet the broth culture.
- Add 20mg/ml lysozyme to the enzymatic lysis buffer.
- Re-suspend the pellet in 180 µl of enzymatic lysis buffer.
- Incubate for 30 minutes at 37°C
- Add 20 µl of 10 mg/ml proteinase K, 200 µl of buffer AL and vortex for 5 seconds.
- Incubate at 56°C for 30 minutes and then at 95°C for 15 minutes
- Add 200 µl of buffer AL and vortex for 5 seconds.
- Incubate at 70°C for 10 minutes.
- Repeat the above two steps twice
- Add 200 µl of ethanol (95 %) and vortex for 5 seconds.
- Pipette the mixture into a column provided in the kit (QIAamp spin columns).
- Centrifuge for 2 minutes at 10,000rpm (repeat centrifuge to remove the liquid completely from the column).
- Discard the collection tube with the flow through.

- Place the column in a new collection tube; add 500 µl of buffer AW1 and centrifuge for 1minute at 10,000rpm.
- Discard collection tube.
- Place column in a new collection tube and add 500 µl of buffer AW2.
- Centrifuge for 3minutes at 13,000rpm.
- Discard the collection tube.
- Place the column in a 1.5ml micro-centrifuge tube.
- Add 200 µl of buffer AE, wait for 1 minute and centrifuge for 1 minute at 10,000 rpm.
- Repeat the previous step and discard the column.

The *emm* family pattern was ascertained by a PCR based chromosomal mapping technique that uses oligonucleotide primers corresponding to *emm* family specific sequences, located near the 3' terminus of each *emm* like gene [30, 41, 245]. Chromosomal DNA purified from each strain was subjected to 7 PCR reactions in which each reaction will contain a different combination of primer pair. The *emm* family pattern was established based on the profiles developed by the presence or absence of PCR amplicon for the pattern specific primer set. The *emm* pattern primers and the *emm* family pattern determining scheme is shown below.

UP-2 5' TCTGGATCCCACTCCCCCAACAAGTTGC 3,

SF4-A 5' CTCCTAGGTTCAGCTAAGCGTGAGTTG 3'

G3-F 5'CGAGAAGTAGAAAAACGTTATCAAGAAC 3'

IG-F 5' CTGGCCTTTACTCCTTTTGATTAACC 3'

SF1-R 5' GTGCTTGACCTTTACCTGGAACAGCTT 3'

SF2-R 5' GTTAGCTTGGGCTACTT 3'

SF3-R 5' GCTGTTTGAGCAGCTCTACC 3'

PCR reaction mixture consisted of 2.0 mM $MgCl_2$, 125 mM dNTPs, 35 pM of each primer, 2.5 U of Taq DNA polymerase for 25 µl reaction. The program consisted of initial denaturation of 94°C for 2 minutes, followed by 94°C for 1 minute, varied specific annealing temperature for 1 minute and 72°C for varied time depending upon the length of the target DNA repeated for 30 times. The amplicons were analyzed by electrophoretic detection.

Each test strain was subjected to PCR using seven sets of primers as given above. The *emm* family pattern was identified by the following reactions.

Forward Primer	Reverse Primer	Annealing Temperature	*emm* family pattern
UP-2	SF1-R	60°C	(A – C)
IG-F	SF3-R	55°C	C,D,E
G3-F	SF3-R	55°C	C,D,E
SF4-A	G3-R	57°C	D,E
SF4-A	SF1-R	60°C	D
G3-F	SF2-R	50°C	E
SF4-A	SF2-R	50°C	E

5.7 *sof* GENE DETECTION

5.7.1 Selection of strains

Detection of *sof* gene, responsible for the production of serum opacity factor which is associated with family pattern E or class II *emm* types [42, 177] . Thus detection of this gene by PCR technique is useful in ascertaining the class of *emm* types (belonging to *emm* pattern E) circulating in the community. In this study, *sof* gene was detected by a PCR technique on representative strain of all *emm* types that were also used for studying the *emm* family pattern.

5.7.2 DNA extraction, PCR & Electrophoresis

Single colony from the confirmed *emm* type stock was grown in Todd Hewitt broth for about 16 hours. The pellet was resuspended with 180 µl of lysis buffer containing 20 mg/ml of lysozyme was incubated at 56°C for 30 minutes. Then the Qiagen kit (Germany, Hilden) protocol was followed for extracting the DNA in pure form.

The relation of class II *emm* types of GAS belonging to *emm* family pattern E is usually detected with the phenotypic presence of opacity factor in the serum, which is detected by PCR using the primers.

sof 5' GTA TAA ACT TAG AAA GTT ATC TGT AGG 3'

5' GGC CAT AAC ATC GGC ACC TTC GTC AAT T 3'

with initial 94°C for 2 minutes, followed by 30 cycles of 94°C for 1min, 49°C for 45 s, 72°C for 2 minutes, and final elongation 72°C for 5 minutes. All PCR products are subjected to electrophoresis on 1.5% agarose gel with 10 mg/ml of ethidium bromide and visualized under UV radiation and documented.

5.8. GENERATING A PHYLOGENETIC TREE

5.8.1 Selection of strains

Seventy six GAS isolates representing 76 different *emm* types (out of the total of 77 *emm* types identified in this study) were selected to generate the phylogenetic tree. One strain representing *emm* type st3211 was not available during the study.

5.8.2 DNA extraction

DNA was extracted with the same method followed for *emm* family pattern and *sof* determination.

5.8.3 PCR and Electrophoresis

The whole length of *emm* gene extended to the conserved signal sequence on the hyper variable 5′ side and the conserved region of the 3'side was amplified using the primers.

MF2 : GGATCCATAAGGAGCATAAAAATGGCTA

MR1 : TGATAGCTTAGTTTTCTTCTTTGCGTTTT

(or)

MFall : CATAAGGAGCATAAAAATGGCTA

MRall : GCTTAGTTTTCTTCTTTGCGTTTT

with initial 94°C for 2 minutes, followed by 30 cycles of 94°C for 1 minute, 55°C (MF2, MR1) or 50°C (MallF, MallR) for 1 minute, 72°C for 2 minutes, and final elongation 72°C for 5 minutes. All PCR products are subjected to electrophoresis on 1.5 % agarose gel with 10 μg/ml of ethidium bromide and visualized under UV radiation and documented. The samples with single band are selected for sequencing using the same forward and reverse primers.

5.8.4 Construction of the phylogenic tree

The phylogenetic analysis was done by MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. [246] using Neighbor joining method with bootstrap value of 100 from 500 replicates. The evolutionary distances were computed using the Kimura 2-parameter method and the branch lengths are proportional to the difference between *emm* types sequence variation.

5.9. MEASUREMENT OF MEAN DISTANCE OF CLUSTERS IDENTIFIED IN THE PHYLOGENETIC TREE

The number of base substitutions per site from averaging over all sequence pairs between groups is shown. All results are based on the pairwise analysis of 76 sequences. Analyses were conducted using the Maximum Composite Likelihood method in MEGA4 [246]. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). There were a total of 503 positions in the final dataset.

5.10. CONSERVED REGION SEQUENCES OF *emm* TYPES

All the sequences obtained for the phylogenetic study were analyzed for the conserved region sequence based class difference. Multiple alignment was made using ClustalW software (http://www.ebi.ac.uk/Tools/clustalw2/). ClustalW2 is a general purpose multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up so that the identities, similarities and differences are seen.

5.11. DETECTION OF PAVF

5.11.1. Selection of strains

Representative isolates of 46 *emm* types from pharyngitis that are available during the time of study and the 61 representative *emm* types of normal carriage and 62 *emm* types of impetigo isolates obtained from the school children were selected for the

study. In order to investigate the involvement of PAVF in the invasive ability, 49 invasive isolates obtained during the period of this study were also included.

5.11.2. Invasive isolates

Invasive GAS isolates used for the determination of Phage Associated Virulence Factor (PAVF) genes were selected from an earlier collection from sterile body sites of patients seen at the Christian Medical College Hospital (CMC), Vellore, Tamil Nadu, India. The sterile body site samples include blood, Cerebrospinal Fluid (CSF), Peritoneal fluid, Pleural fluid and other fluids. These isolates were identified by procedures described above and were already used for other molecular studies. They were retrieved from lyophilized stock of culture collection from the Department of Microbiology, CMC, Vellore. The clinical details and the relevant personal data were collected by review of the medical charts of patients from whom the samples were obtained. Among all the GAS invasive isolates recorded within January 2000 to July 2007, only those available and that could be retrieved in viable form were included for the study.

5.11.3. DNA extraction, PCR and Electrophoresis

The direct colony PCR method is followed for the detection of these virulence factors. The single colony of the confirmed *emm* type were grown in 4 ml of Todd Hewitt broth and centrifuged. The pellets were washed with normal saline and then with TE buffer and finally resuspended in 1 ml of TE buffer. One micro-liter of the culture suspension is placed on to 96 well PCR plate and dried at 94°C for 2 minutes in the PCR machine before the addition of PCR mix [244].

A PCR based method of determining the presence of phage-associated virulence genes are employed to finely characterize the commonly found GAS *emm* types of varied disease groups. The oligonucleotide primers [247] used are:

speA	5’ ATG GAA AAC AAT AAA AAA GTA TTG 3’
	5’ TTA CTT GGT TGT TAG GTA G 3’
speC	5’ TCT AGT CCC TTC ATT TGG TG 3’
	5’ GTA AAT TTT TCA ACG ACA CA 3’
speI	5’ AAT GAA GGT CCG CCA TTT TC 3’
	5’ TCT CTC TGT CAC CAT GTC CTG 3’
speH	5’ GTG AAT GTC CAG GGA AAA GG 3’
	5’ GCA TGC TAT TAA AGT CTC CAT TG 3’
ssa	5’ TGA TCA AAT ATT GCT CCA GGT G 3’
	5’ TCC ACA GGT CAG CTT TTA CAG 3’
speL	5’ CAG CAC CTT CCT CTT TCT CG 3’
	5’ GGA AAA AGA GGG ACG CAA G 3’
speM	5’ GGA TGA GTG AAT AAA TCG GTA AAC 3’
	5’ AGT CTG GGA CGA TGA TAA 3’
spd1 (mf2)	5’ACTGTTGACGCAGCTAGGG
	5’CCCTTCAGGATTGCTGTCAT
spd3 (mf3)	5’ CAA ATT GAC TGA CGC TAA TGG 3’
	5’ CCG CTT CTT CAA ACT CTT CG 3’
spd4 (mf4)	5’ TGG CAT TGC TTC ATA GTA AAG G 3’
	5’ ATC TAC CTG AAG CTT TGT CGT G 3’
sla	5’ CTC TAA TAG CAT CGG CTA CGC 3’
	5’ AAT GGA AAA TGG CAC TGA AAG 3’

speK	5' GTG TGT CTA ATG CCA CCA CCG TCT 3'
	5' GGA ACA TAT ATG CTC CTA GAT 3'
sdn	5' AAC GTT CAA CAG GCG CTT AC 3'
	5' ACC CCA TCG GAA GAT AAA GC 3'
sda	5' AGC ATA GCC CCA AAA ATG TG 3'
	5' TGC TAC AGC ATT GAA GAC ACG 3'

with initial 94°C for 2 minutes, followed by 30 cycles of 94°C for 30 seconds, 55°C or 50 ^{o}C for 30 seconds, 72°C for 3 minutes, and final elongation 72°C for 7 minutes. All PCR products are subjected to electrophoresis on 1.5 % agarose gel with 10 μg/ml of ethidium bromide and visualized under UV radiation and documented.

5.12. DETECTION OF CCVF

5.12.1 Selection of strains

Representative 46 *emm* types from pharyngitis and 56 *emm* types from impetigo isolates available during the study were chosen to compare the association of the virulence factors to each category. Few *emm* types from children with normal throat, that were not present either in pharyngitis or in impetigo, were also chosen in order to study the association of *emm* types with the virulence factors.

5.12.2 DNA extraction, PCR and Electrophoresis

DNA was extracted with the same method followed for *emm* family pattern and *sof* determination.

A PCR based method for the detection of genes for the toxins and fibronectin binding protein are as described by Valminckx *et al*. [214]

The specific primer sequences used for toxic genes are:

speB 5’ AAG AAG CAA AAG ATA GC 3’

5’ TGG TAG AAG TTA CGT CC 3’

speF 5’ TAC TTG GAT CAA GAC G 3’

5’ GTA ATT AAT GGT GTA GCC 3’

speG 5’ AGA AAC TTA TTT GCC C 3’

5’ TAG TAG CAA GGA AAA GG 3’

speJ 5’ ATC TTT CAT GGG TAC G 3’

5’ TTT CAT GTT TAT TGC C 3’

smeZ 5’ TAA CTC CTG AAA AGA GGC T 3’

5’ CAT TGG TTC TTC TTG ATA AG 3’

slo 5’ GCC AAT GTT TCA ACA GCT ATT G 3’

5’ CGG AGC TGC ACT AAA GGC CGC 3’

ska 5’ TCC AAG CTA TTG CTG GGT A 3’

5’ GGT GTC CCT GTA TAA CGC A 3’

sic 5’ TAA GGA GAG GTC ACA AAC TA 3’

5’ TTA CGT TGC TGA TGG TGT AT 3’

The Primer sequences for various fibronectin binding proteins are

cpa 5’ CTC AAA ATG CTA TTT GGT AT 3’

5’ ATT TCC CAT CTT TAG CTA CT 3’

cpa-1 5’ TGT GAA CTT CCA TTT TTA TT 3’

5’ AGA GTA GCA CAC GAT TTA AG 3’

fba 5’ GGT GAT TCA ACA TCA GTT AC 3’

5’ CGT TTT GTG ACT AAA AGA CT 3’

fbp-54 5’ CTT CAG AAT CTG TTT CTT TG 3’

5’ AGT TCA CAG GTT GTC TAT TG 3’

pfbp 5’ CTG AAT ATG CTG CTT TTA CT 3’

5’ TTA TCC TTC GTT ACT TCT TG 3’

prtf-1 5’ CCT TTG TAG ATT ATG CTC AC 3’

5’ TTC TGT CTC AAC CAT ATT TC 3’

prtf-2 5’ AAA GCA ATT ATA TTA CTA ATG 3’

5’ TTT TGT TTC ATA CAG GTC 3’

prt-15 5’ TGG GAG TAC AGA AAC TTT TA 3’

5’ ACT AGG TAC ATA TTC AGC AC 3’

sciA 5’ TGA CAT CAA AGG AGA GAC AA 3’

5’ CAC GAG CAC CAG CTT TAC 3’

sciB 5’ TGA CAA ACA AAC AAA CTC ACT 3’

5’ ATA AAC TGC AAA ATC CCA AA 3’

sfb 5’ CAT ATC AGG CTT ATT GTT TT 3’

5’ TTC TGT CTC AAC CAT ATT TC 3’

sfb-2 5’ ATG ACA AAA GAG AAT TTT GA 3’

5’ TGT GAT ATT TTC ATT TAC CC 3’

scpA 5’ GGG GGG GGATCC TGT AAC GTG CAA TAG AC 3’

5’ GGG GGG CCG CGG GGG TGC TGC AAT AGA C 3’

scl 5’ CTC CAC AAA AGA GTG ATC AGT C 3’

5’ TTA GTT GTT TTC TTT GCG TTT 3’

Amplification conditions are 5 minutes. Initial denaturation at 96°C, followed by 30 cycles of denaturation at 96°C for 55 seconds, 65 seconds of annealing temperature for each gene and 70 seconds of extension at 72°C, with a final step at 72°C for 5 minutes. All PCR products are subjected to electrophoresis on 1.5 % agarose gel with ethidium bromide and visualized under UV radiation.

6. RESULTS

6.1. *emm* TYPING OF GAS ISOLATES

6.1.1. Community study isolates

A total of 698 group A streptococci (GAS) isolates obtained from school children with pharyngitis (*n=116*), impetigo (*n=276*) as well as those colonizing throat of asymptomatic children (*n=306*) were selected for this study (Table 1). GAS isolates with same *emm* type isolated repeatedly from the same child within consecutive weeks were considered as the same strain. Thus only 698 of the 769 isolates typed were selected for the final analysis.

Table 1: Details of GAS isolates selected for the study.

Total Number of	PT	SK	NT	Total
Visits	60	60	15	NA
Child visits	11722	11722	2823	NA
With symptoms	685	513	NA	NA
With GAS	125	339	326	NA
emm typed	125	338	306	769
Accounted for analysis	116	276	306	698

PT = Pharyngitis associated throat isolates; SK = Impetigo associated skin isolates;
NT = Normal throat isolates; NA = Not Applicable

6.1.2. Distribution of *emm* types among 698 GAS isolates

The distribution of 77 *emm* types including sequence types (st) and a nontypable strain were identified among 698 GAS isolates (Table 2). Of these, 14 types had more than one sub type and 23 subtypes were identified from 11 types of the primary reference strains. The most predominant of all type was *emm* 63 (4.6 %) followed by *emm* 81 (4.3 %) and then the *emm* 28 (4.1 %). Globally disseminated invasive and rheumatogenic types *emm* 1, 3, 12 & 18 accounted for only 4.4 % (*n=31*) of the 698 isolates.

Table 2: Distribution of *emm* types with subtypes among 698 GAS isolates

Sl. No.	*emm* type	*emm* Subtype	Total
1	63	63 (10) 63.3 (22)	32
2	81	81.1 (13) 81.2 (17)	30
3	28	28.5	29
4	49	49 (10) 49.4 (17)	27
5	122	122.1	24
6	44/61	44/61	23
7	82	82.1	22
8	st854	st854.1	22
9	42	42 (4) 42.1 (16) 42.3 (1)	21
10	118	118	21
11	st1731	st1731.0 (2) st1731.1 (19)	21
12	53	53 (13) 53.1 (6)	19
13	1	1-2.2 (14) 1.4 (4)	18
14	85	85	18
15	77	77	17
16	4	4.5	15
17	105	105	15
18	st2147	st2147	15
19	75	75.1	15
20	71	71 (13) 71.1 (1)	14
21	55	55	12
22	80	80	12
23	57	57	11
24	74	74	11
25	89	89	11
26	103	103	11
27	109	109.1	11
30	110	110	11
28	3	3.22	10
29	25	25.1 (2) 25.2 (8)	10
31	58	58 (8) 58.1 (1)	9
32	69	69.1	9
33	86	86.2	9
34	15	15 (7) 15.1 (1)	8
35	39	39.1	8
36	st1389	st1389 (1) st1389.1 (6) st1389.2 (1)	8
37	65	65.1	7
38	97	97.1	7
39	st11014	st11014	7
40	67	67	6
41	stKNB2	stKNB2 (2) stKNB2.1 (2) stKNB2.2 (2)	6
42	8	8	5
43	60	60.3	5
44	36	36.2	4
45	54	54	4
46	68	68	4
47	100	100.1	4
48	9	9	3
49	11	11.1	3
50	22	22.8	3
51	56	56	3
52	73	73	3
53	88	88.1 (1) 88.3 (2)	3
54	92	92	3
55	93	93	3
56	102	102.2	3
57	st62	st62	3
58	stD432	stD432	3
59	12	12	2
60	31	31.1	2
61	70	70	2
62	106	106	2
63	108	108	2
64	stD633	stD633	2
65	stKNB1	stKNB1	2
66	18	18.12	1
67	76	76.1	1
68	79	79.1	1
69	87	87	1
70	104	104	1
71	113	113	1
72	st212	st212	1
73	st3211	st3211	1
74	stKNB3	stKNB3	1
75	stKNB4	stKNB4	1
76	stKNB5	stKNB5	1
77	stKNB6	stKNB6	1
78	Nontypable	Nontypable	1
			698

Note: Numbers in parenthesis indicate the number of isolates in each subtype.

6.1.3. Validated and other *emm* types among 698 GAS isolates

Table 3 gives the distribution of validated *emm* types, recognized but not validated sequence types (st) and hitherto unrecognized types (new types) among the 698 GAS isolates. Sixty one (78.2 %) of the 78 types belonged to already validated *emm* types. There were 10 sequence types (st) and 6 new types while one alone was nontypable.

Table 3: Distribution of validated *emm* types, sequence types, new types and others among GAS isolates

.*emm* types	Total No. of *emm* types	Total No. of isolates
Validated types 63, 81, 28, 49, 122, 44/61, 82, 42, 118, 53, 1, 85, 77, 4, 105, 75, 71, 55, 80, 57, 74, 89, 103, 109, 110, 3, 25, 58, 69, 86, 15, 39, 65, 97, 67, 8, 60, 36, 54, 68, 100, 9, 11, 22, 56, 73, 88, 92, 93,102, 12, 31, 70, 106, 108, 18, 76, 79, 87,104,113	61 (78.2 %)	602 (86.2 %)
Sequence types st854, st1731, st2147, st1389, st11014, st62, stD432, stD633, st212, st3211	10 (12.8 %)	83 (11.9 %)
New types stKNB1, stKNB2, stKNB3, stKNB4, stKNB5, stKNB6	6 (7.7 %)	12 (1.7 %)
Nontypable	1 (1.3 %)	1 (0.14 %)
Total	**78**	**698**

6.1.4. GAS isolates and their subtypes

Four hundred and seventy one (67.5 %) of the 698 GAS isolates had no or only one subtypes identified among them (Table 4). Among these, 233 (33.4 %) had no subtypes while 238 (34.1 %) had one subtype. These represented 63 (81.8 %) 77 *emm* types.

Table 4: Distribution of *emm* types with subtypes among GAS isolates.

***emm* type with**	**Types**		**Isolates**	
	No	**%**	**No**	**%**
No Subtype	37	47.4	233	33.4
1 subtype	26	33.3	238	34.1
2 subtype	11	14.1	191	27.4
3 subtype	3	3.8	35	5.0
Nontypable	1	1.3	1	0.1
Total	**78**		**698**	

6.1.5. Distribution of *emm* types among Pharyngitis (PT) GAS isolates

Forty seven types were identified among 116 pharyngitis associated GAS isolates (PT) (Fig. 5). The most predominant type was *emm* 28 (7.76 %) while the most predominant three types (*emm* 28, 1 & 49) accounted for 17.2 % and 11 types accounted for 44.8% of the total isolates. One type, stKNB2 was a new type while one isolate was nontypable.

6.1.6. Validated and other *emm* types among PT isolates

Thirty seven (78.7 %) of the 47 types identified among PT isolates were validated *emm* types (Table 5). Only one new type was identified in this group with 2 isolates, while the only non-typable GAS isolate among the entire 698 isolates was a PT isolate.

Table 5: Distribution of validated *emm* types, sequence types, new types and others among PT Isolates

***emm* types**	**Total No. of *emm* types**	**Total No. of isolates**
Validated types 1, 3, 4, 15, 22, 25, 28, 36, 39, 42, 44/61, 49, 53, 55, 58, 63, 67, 68, 71, 73, 74, 75, 77, 80, 81, 85, 89, 97, 100, 102, 103, 105, 106, 109, 110, 118, 122	37 (78.7 %)	97 (83.6 %)
Sequence types st11014, st1389, st1731, st212, st2147, st62, st854, stD432	8 (17.0 %)	16 (13.8 %)
New type stKNB2	1 (2.1 %)	2 (1.7 %)
Nontypable	1 (2.1 %)	1 (0.9 %)
Total	**47**	**116**

6.1.7. Subtypes among PT isolates

Forty two (89.4 %) of the 47 *emm* types identified among PT isolates had no or 1 subtype and only 4 (8.5 %) of the *emm* types had 2 subtypes (Table 6).

Table 6: Subtypes among PT isolates

***emm* type with**	**Number of Types**	**Number of Isolates**
No / 1 subtype	42 (89.4 %)	100 (86.2 %)
2 subtype	4 (8.5 %)	15 (12.9 %)
Nontypable	1 (2.1 %)	1 (0.9 %)
Total	**47**	**116**

6.1.8. Distribution of *emm* types among SK isolates

Sixty two types were identified among 276 SK GAS isolates (Fig. 6). Type 122.1 (5.2 %) was the most predominant type. The most predominant 4 *emm* types (*emm* 122.1, 81, 63, 44/61) accounted for 21 % of the SK isolates. Two types, stKNB1 and stKNB2 were new types and all isolates were typable.

6.1.9. Validated *emm* and other types among SK isolates

Fifty three (85.5 %) of the 62 *emm* types identified among SK isolates were validated *emm* types (Table 7). There were 7 st types and 2 new types. Approximately there were 9 fold differences between the number of isolates in new types, st types and validated types among SK isolates.

Table 7: Distribution of validated *emm* types, sequence types and new types among SK isolates

emm types	Total *emm* types	Total isolates
Validated types 1, 3, 4, 8, 9, 11, 15, 18, 25, 28, 31, 36, 39, 42, 44/61, 49, 53, 54, 55, 56, 57, 58, 60, 63, 65, 67, 69, 70, 71, 73, 74, 75, 76, 77, 79, 80, 81, 82, 85, 86, 87, 89, 92, 93, 100, 103, 105, 106, 108, 109, 110, 118, 122	53 (85.5 %)	246 (89.1 %)
Sequence types st11014, st1389, st1731, st2147, st3211, st62, st854	7 (11.3 %)	27 (9.8 %)
New types stKNB1, stKNB2	2 (3.2 %)	3 (1.1 %)
Total	**62**	**276**

6.1.10. Subtypes among SK isolates

Fifty (80.6 %) of the 62 *emm* types identified among 276 SK isolates had no or only 1 subtype (Table 8). The *emm* types with 2 subtypes were 19.4 %. All isolates were typable.

Table 8: Subtypes among SK isolates

***emm* type With**	**Number of Types**	**Number of Isolates**
No subtype / 1 subtype	50 (80.6 %)	192 (69.4 %)
2 subtype	12 (19.4 %)	84 (30.4 %)
Total	**62**	**276**

6.1.11. Distribution of *emm* types among NT GAS isolates (*n = 306*)

Distribution of 67 types identified among 306 GAS isolates colonizing throat of asymptomatic children (NT) is given in Fig. 7. Type 28 (4.9 %) was the most predominant type followed by *emm* 118, 49, 82 which together accounted for 16.7 % of the 306 isolates. Eleven of the 16 sequence types were identified among these isolates. All strains were typable and all 6 new types were seen among them.

6.1.12. Validated and other *emm* types among NT isolates

Fifty four (80.6%) of the 67 types were validated *emm* types identified among 306 NT isolates. Six newly identified types accounted for 13.1% of the 306 isolates. (Table 9).

Table 9: Distribution of validated, sequence and new *emm* types among NT isolates.

***emm* types**	**Total *emm* types**	**Total No. of isolates**
Validated types 1, 3, 4, 8, 9, 11, 12, 15, 22, 25, 28, 36, 39, 42, 44/61, 49, 53, 54, 55, 56, 57, 58, 60, 63, 65, 67, 68, 69, 70, 71, 73, 74, 75, 77, 80, 81, 82, 85, 86, 88, 89, 92, 93, 97, 102, 103, 104, 105, 108, 109, 110, 113, 118, 122	54 (80.6 %)	259 (84.6 %)
Sequence types St11014, st1389, st1731, st2147, st854, stD432, stD633	7 (10.4 %)	40 (13.1 %)
New types stKNB1, stKNB2, stKNB3, stKNB4, stKNB5, stKNB6	6 (8.9 %)	7 (2.29 %)
Total	**67**	**306**

6.1.13. Subtypes among NT isolates

Fifty nine (88.1 %) of the 67 *emm* types identified among 306 NT isolates had either no subtype or only one subtype (Table.10). Though the NT isolates accounted for the highest number of *emm* types, only 8 (11.9 %) types had 2 subtypes.

Table 10: Subtypes among NT isolates.

***emm* type With**	**Number Types**	**%**	**Number Isolates**	**%**
No subtype / 1 subtype	59	(88.1 %)	232	(75.8 %)
2 subtype	8	(11.9 %)	74	(24.2 %)
Total	**67**		**306**	

6.1.14. Distribution of subtypes among diverse GAS isolates

Sixty seven percent of 698 GAS isolates was restricted with no or one subtype. This was highest among PT isolates (86.2 %) while 30.4 % of SK isolates had 2 subtypes. (Table 11).

Table 11: Overall distribution of subtypes among GAS isolates from different sites.

Category of Isolates	**% with No Subtype / 1 subtype**	**% with 2 Subtypes**	**% with 3 Subtypes**	**% Nontypable**	**Total**
PT (116)	86.2	12.9	0	0.9	100
SK (276)	69.6	30.4	0	0	100
NT (306)	75.8	24.2	0	0	100
Total (698)	67.5	27.4	5.0	0.1	100

Numbers in parentheses indicate the total number in each group

6.1.15. Distribution of *emm* types common to GAS isolates from different sources

Thirty seven (47.4 %) of the 78 *emm* types identified among 698 GAS isolates were seen among all three groups of isolates (Table 12). Only 17 (21.8 %) of the isolates were restricted to any one of the three sites. Most of the predominant *emm* types except *emm* 82 were found to be common in all the categories.

Table 12: Distribution of *emm* types common to GAS isolates from different sources

Common *emm*Types in	***emm* types**	**No. of *emm* types**
PT, NT & SK	63, 81, 28, 49, 122, 44/61, st854, 42, 118, st1731, 53, 1, 85, 77, 4, 105, st2147, 75, 71, 55, 80, 74, 89, 103, 109, 110, 3, 25, 58, 15, 39, st1389, st11014, 67, stKNB2, 36, 73	37
PT & NT	97, 68, 22, 102, stD432	5
PT & SK	100, st62, 106	3
NT & SK	82, 57, 69, 86, 65, 8, 60, 54, 9, 11, 56, 92, 93, 70, 108, stKNB1	16
PT only	st212, Nontypable	2
NT only	88, 12, stD633, 104, 113, stKNB3, stKNB4, stKNB5, stKNB6	9
SK only	31, 18, 76, 79, 87, st3211	6

PT = Pharyngitis associated Throat GAS Isolates; SK = Impetigo associated Skin GAS isolates; NT = Normal throat associated GAS isolates

6.1.16. Distribution of validated and other *emm* types among diverse GAS isolates

Eighty six percent of all GAS isolates belonged to validated *emm* types while 1.7 % were new types (Table 13). Only one (0.14 %) of the 698 isolates was nontypable. The percentage of new types was higher in NT isolates and lowest in SK isolates.

Table 13: Overall distribution of validated types and other types among isolates from different sites.

Category of Isolate	% Validated types	% Sequence types	% New Types	% Non typable
PT (116)	83.6	13.8	1.7	0.9
SK (276)	89.1	9.8	1.1	-
NT (306)	84.6	13.1	2.29	-
Total (698)	86.2	11.9	1.7	0.14

Numbers in parentheses indicate the total number in each group

6.1.17. Comparison of predominant *emm* types in each category of isolates

Figure 8 gives the predominant *emm* types seen among PT, SK and NT isolates. Two types, 49 and 44/61 were seen as predominant types among PT and SK isolates. Eleven types (44.8 %) of PT isolates, eleven types (44.6 %) of the SK isolates and fourteen types (50.4 %) of NT isolates together accounted for 47.1 % of the total 698 isolates typed. They also accounted for 18 (23.4 %) of the 77 types identified among the 698 isolates.

6.1.18. Distribution of common *emm* types among PT and SK isolates

A total of 40 *emm* types common to PT and SK isolates were identified (Table 14). The tissue preference index was more for PT isolates in 24 types and only 16 *emm* types were more in impetigo (Table 14). *emm* 28, *emm* 1 and *emm* 3 make up 15.6 % of the total PT isolates. The tissue preference index was calculated by dividing the percentage of PT or SK isolates of the respective *emm* types with the percentage of the SK and PT isolates respectively.

6.1.19 Distribution of Common *emm* types among GAS Isolates from Different Sources

Thirty seven *emm* types were found to be common among PT, NT and SK GAS isolates. . Among them, 16 types were predominant among PT isolates, 8 among NT isolates and 13 among SK isolates (Table 15).

6.1.20. Comparison of predominant *emm* types seen among different Site of GAS isolates

Fifteen predominant *emm* types accounting for 49.3 % of the 698 isolates and common to GAS isolates from three sites were compared (Fig. 9). Among them, eight types namely, 63, 81, 122, 44/61, 82, st854, 53 and 85 were highest among SK isolates while types 28, 42, st1731, and 1 were highest among PT isolates.

Table 14: Comparison of tissue preference in PT and SK isolates:

emm type	% PT (*n = 116*)	% SK (*n = 276*)	Tissue Preference Index for PT	Tissue Preference Index for SK
28	7.8	1.8	4.33	0.23
1	5.2	1.4	3.71	0.26
49	4.3	2.9	1.48	0.67
63	3.4	5.1	0.67	1.5
44/61	3.4	4.7	0.72	1.38
42	3.4	2.5	1.36	0.73
st1731	3.4	2.5	1.36	0.73
st2147	3.4	0.4	8.50	0.12
75	3.4	1.8	1.89	0.53
80	3.4	1.1	3.09	0.32
15	3.4	1.1	3.09	0.32
118	2.6	1.8	1.44	0.69
53	2.6	3.3	0.79	1.2
105	2.6	2.9	0.90	1.1
55	2.6	1.1	2.36	0.42
110	2.6	0.7	3.71	0.27
3	2.6	1.8	1.44	0.69
st11014	2.6	0.7	3.71	0.27
100	2.6	0.4	6.50	0.15
85	1.7	3.6	0.47	2.1
109	1.7	2.9	0.59	1.7
25	1.7	2.5	0.68	1.5
77	1.7	2.5	0.68	1.5
71	1.7	2.5	0.68	1.5
89	1.7	1.8	0.94	1.06
39	1.7	1.4	1.21	0.82
103	1.7	0.7	2.43	0.41
stKNB2	1.7	0.7	2.43	0.41
36	1.7	0.4	4.25	0.23
81	0.9	5.4	0.17	6.0
122	0.9	5.8	0.16	6.4
st854	0.9	4	0.23	4.4
4	0.9	2.2	0.41	2.4
58	0.9	1.8	0.50	2.0
st1389	0.9	1.1	0.82	1.2
st62	0.9	0.7	1.29	0.8
67	0.9	0.4	2.25	0.4
73	0.9	0.4	2.25	0.4
74	0.9	0.4	2.25	0.4
106	0.9	0.4	2.25	0.4

The tissue preference index was calculated by dividing the percentage of PT isolates of the respective *emm* types with the percentage of the SK isolates.

(REF: Shulman et al. Clin Infect Dis 2004; 39: 325-332)

Table 15: Distribution of common *emm* types in PH, NT and SK isolates.

Sl.No	*emm* type	% PT (*n = 116*)	% NT (*n = 306*)	% SK (*n = 276*)	Total (*n = 698*)
1	**63**	3.4	4.6	5.1	32
2	**81**	0.9	4.6	5.4	30
3	**28**	7.8	4.9	1.8	29
4	**49**	4.3	4.6	2.9	27
5	**122**	0.9	2.3	5.8	24
6	**44/61**	3.4	2.0	4.7	23
7	**st854**	0.9	3.3	4.0	22
8	**42**	3.4	3.3	2.5	21
9	**118**	2.6	4.2	1.8	21
10	**st1731**	3.4	3.3	2.5	21
11	**53**	2.6	2.3	3.3	19
12	**1**	5.2	2.6	1.4	18
13	**85**	1.7	2.0	3.6	18
14	**77**	1.7	2.6	2.5	17
15	**4**	0.9	2.6	2.2	15
16	**105**	2.6	1.3	2.9	15
17	**st2147**	3.4	3.3	0.4	15
18	**75**	3.4	2.0	1.8	15
19	**71**	1.7	1.6	2.5	14
20	**55**	2.6	2.0	1.1	12
21	**80**	3.4	1.6	1.1	12
22	**74**	0.9	2.9	0.4	11
23	**89**	1.7	1.3	1.8	11
24	**103**	1.7	2.3	0.7	11
25	**109**	1.7	0.3	2.9	11
26	**110**	2.6	2.0	0.7	11
27	**3**	2.6	0.7	1.8	10
28	**25**	1.7	0.3	2.5	10
29	**58**	0.9	1.0	1.8	9
30	**15**	3.4	0.3	1.1	8
31	**39**	1.7	0.7	1.4	8
32	**st1389**	0.9	1.3	1.1	8
33	**st11014**	2.6	0.7	0.7	7
34	**67**	0.9	1.3	0.4	6
35	**stKNB2**	1.7	0.7	0.7	6
36	**36**	1.7	0.3	0.4	4
37	**73**	0.9	0.3	0.4	3
	Total	**87.9**	**77.1**	**78.3**	**554**

PT = Pharyngitis associated Throat GAS Isolates; SK = Impetigo associated Skin GAS isolates;
NT = Normal throat associated GAS isolates

6.1.21. Transmission of *emm* types from one site to other

Transmission of an *emm* types from one site to other were identified in 30 children (Table 16). Among them, 14 (46.7 %) were first infected in skin and then resulted in throat colonization with the same *emm* type. In 12 students (40 %), skin infection occurred first and then led to pharyngitis with the same type. In none of the students, pharyngitis leading to skin infection or throat colonization leading to skin infection had occurred.

Table 16: Transmission of same *emm* type/s from one site to other.

Categories of Transmission	**No. of Candidates**	**%**
Skin infection leading to Throat colonization	14	46.7
Skin infection leading to Pharyngitis	12	40
Skin infection and pharyngitis at a time	2	6.7
Skin infection and throat colonisation at a time	2	6.7
Pharyngitis leading to Skin infection	0	0
Throat colonization leading to Skin infection	0	0
Total Candidates	**30**	

6.1.22. Patterns of *emm* type distribution among PT isolates

Fig. 10 shows four classical examples of varied patterns of *emm* type distribution among PT isolates causing pharyngitis seen in four of the 51 children with GAS pharyngitis. Child VAP 51 had a long standing infection of throat with type 28, interspersed with normal throat colonization. Child 110 had infection of throat with six different types (red color) and colonization with three different types (black color).

6.1.23. Patterns of type distribution among SK isolates

Fig. 11 shows three examples of varied patterns of GAS impetigo seen in three of the 88 children with GAS impetigo. Child VAP 139 had 8 episodes of GAS impetigo with an equal number of *emm* types while child VAP 195 had also multiple episodes with as many as many as 4 different types of *emm* types.

6.1.24. Patterns of distribution of types among NT isolates

Examples of colonization patterns of *emm* types of GAS isolates among asymptomatic school children are given in Fig. 12. Child VAP 57 was colonized in the throat with three different *emm* types wherein one type, st854.2 persisted on six occasions during a period of 14 months. In contrast, child VAP 78 was colonized with four different types during a period of 8 months. On the other hand, in child VAP 87, colonization was seen with only one *emm* type during the entire two year period.

6.1.25. Patterns of transmission of *emm* types from one site to the other

Four examples of diverse patterns of transmission of *emm* types from one site to another are given in Fig. 13. Child VAP194 had single episode of impetigo with type 57 following which it colonized the throat. Child VAP190 had a single episode of impetigo with type 71 followed by throat colonization and subsequent pharyngitis with the same type. Child VAP176 had simultaneous impetigo and pharyngitis with the same type, st1731. Child VAP4 had simultaneous impetigo and throat colonization with the same type 82.1.

6.1.26. Multiplicity of types causing GAS pharyngitis in school children

A total of 140 school children who were regularly surveyed every week for 60 weeks during the years, 2002 – 2004 were further analyzed for multiplicity of *emm* types of PT isolates causing pharyngitis. Among them, 51 children had GAS pharyngitis; of these, 40 (28.6 %) had infection with one *emm* type, 9 (6.4 %) with 2 types, and 1 each with three and four types respectively (Data not shown).

6.1.27. Multiplicity of types causing GAS impetigo in school children

A total of 140 school children who were regularly surveyed every week for 60 weeks during the years, 2000 – 2002, were further analyzed for multiplicity of skin infection (Table 17). Among 88 children (62.9 %) who had GAS impetigo, 38 (27.1 %) had infection with one *emm* type, 26 (18.6 %) with 2 types, 9 (6.4 %) with 4 types, and 2 (1.4%) each with five types. One lone child had 8 episodes with an equal number of types.

6.1.28. Multiplicity of *emm* types among NT isolates colonizing throat

A total of 140 school children who were regularly surveyed every month for 15 consecutive months during 2000 – 2002, were further analyzed for patterns of colonization (Table 18). Among them, 101 (72.1 %) children had GAS colonization of the throat at least once. Ninety three children (66.4 %) were colonized with one or two types while 8 (5.7 %) were colonized with three or four types.

Table 17: Frequency distribution of *emm* types among SK isolates

No of *emm* types causing Impetigo	No of Children	%
0	52	37.1
1	38	27.1
2	26	18.6
3	12	8.6
4	9	6.4
5	2	1.4
8	1	0.7

Table 18: Frequency distribution of *emm* types among NT isolates

No of *emm* types Colonising Throat	No. of Children	%
0	39	27.9
1	56	40
2	37	26.4
3	7	5
4	1	0.7

6.1.29. Monthly distribution of *emm* types identified among PT isolates

The *emm* type distribution in pharyngitis over the 2 years (Fig. 14) shows, 22 types out of 47 different types were found only in one survey. Also, of these 47 types, 27 types were found only in the first 7 surveys and 11 other types were found in the latter 8 surveys only. Among the remaining 9 types which cross over the 2 halves of the study, 6 types are spread with in 8 surveys. The other 3 *emm* type occurred wide apart over the period 12 and 13 surveys each but presented in 2 or 3 surveys only.

6.1.30. Monthly distribution of *emm* types among SK isolates

Monthly distribution of *emm* types of SK isolates for 15 consecutive months (Fig. 15) showed that 22 of 62 types were found only in one survey. Also, of these 62 types, 26 were found only in the first 7 surveys while 12 other types were found in the latter 8 surveys only. Twenty four types were found spread over many monthes. The most predominant *emm* type of impetigo is clustered together in 5 to 7 surveys only.

6.1.31. Monthly distribution of *emm* types among NT isolates

Month wise distribution of 67 *emm* types for 15 consecutive months during 2000 - 2002s showed that 27 types were found only in the first 7 surveys while 14 other types were found in the latter 8 surveys; 26 types were found in both halves of the study (Fig. 16). No type was found in all the surveys. Type 82 was the most widely distributed type and was found in 8 surveys. Type 85 was present in the first 2 surveys and then reemerged in the last survey. Few types appeared and reemerged with intervals during many surveys.

6.2 DETERMINATION OF *emm* FAMILY PATTERN

Seventy four GAS isolates representing 74 *emm* types were selected for determining the *emm* family pattern (Table 19). These included 40 from PT isolates, 29 from NT isolates and 5 from SK isolates (Data not shown). Family patterns could be established in 67 of the 74 types while it could not be determined in seven types. The latter included types 65, 87, 100, 122 as well as two new types, stKNB4 & stKNB5 and the lone nontypable isolate.

Table 19: Distribution of *emm* family pattern and *sof* in 74 representative *emm* types.

Sl.No:	*emm* type	Site of isolation	family pattern	*sof*
1	4	PT	E	+
2	25	PT	E	+
3	28	PT	E	+
4	49	PT	E	+
5	58	PT	E	+
6	63	PT	E	+
7	68	PT	E	+
8	73	PT	E	+
9	75	PT	E	+
10	89	PT	E	+
11	102	PT	E	+
12	103	PT	E	+
13	106	PT	E	+
14	109	PT	E	+
15	110	PT	E	+
16	118	PT	E	+
17	st11014	PT	E	+
18	st1389	PT	E	+
19	st1731	PT	E	+
20	st212	PT	E	+
21	st2147	PT	E	+
22	15	PT	E	-
23	9	NT	E	+
24	11	NT	E	+
25	22	NT	E	+
26	60	NT	E	+
27	82	NT	E	+
28	88	NT	E	+
29	92	NT	E	+
30	93	NT	E	+
31	104	NT	E	+
32	113	NT	E	+
33	stKNB1	NT	E	+
34	stKNB2	NT	E	+
35	stKNB3	NT	E	+
36	stKNB6	NT	E	+
37	76	SK	E	+
38	79	SK	E	+
39	st3211	SK	E	+

Sl.No:	*emm* type	Site of isolation	family pattern	*sof*
40	54	NT	D	-
41	56	NT	D	-
42	69	NT	D	-
43	70	NT	D	-
44	86	NT	D	-
45	108	NT	D	-
46	st854.1	NT	D	-
47	stD432	NT	D	-
48	stD633	NT	D	-
49	74	PT	D	+
50	81	PT	D	+
51	85	PT	D	+
52	36	PT	D	-
53	42	PT	D	-
54	53	PT	D	-
55	67	PT	D	-
56	71	PT	D	-
57	77	PT	D	-
58	80	PT	D	-
59	97	PT	D	-
60	105	PT	D	-
61	18	SK	D	-
62	3	PT	A-C	-
63	39	PT	A-C	-
64	55	PT	A-C	-
65	1	NT	A-C	-
66	12	NT	A-C	+
67	57	NT	A-C	-
68	100	PT	**Nontypable**	-
69	122	PT	**Nontypable**	-
70	new	PT	**Nontypable**	-
71	stKNB4	NT	**Nontypable**	+
72	65	NT	**Nontypable**	-
73	stKNB5	NT	**Nontypable**	-
74	87	SK	**Nontypable**	+
75	8	Not done	Not done	NK
76	31	Not done	Not done	NK
77	44/61	Not done	Not done	NK
78	st62	Not done	Not done	NK

6.2.1. Distribution of *emm* family pattern among PT, NT & SK *emm* types

Table 20 gives the family pattern of individual *emm* types identified among PT, NT and SK isolates (*n = 74*). Thirty nine (52.7 %) of the 74 types belonged to the pattern E and 25 (33.8 %) belonged to pattern D. Thus, 67 (90.5 %) of the 74 types belonged to either of these types. This pattern distribution was similar among the three groups of isolates. Further, 12 (30 %) of the 40 types associated with pharyngitis belonged to pattern D which is reportedly skin specific.

Table 20: Distribution of family patterns among PT, NT & SK *emm* types**.**

***emm* Family Pattern**	**PT *emm* types**	**NT *emm* types**	**SK *emm* types**	**Total *emm* types**
A – C	(*n=3*) 3, 39, 55	(*n = 3*) 1, 12, 57	**0**	**6**
D	(*n = 12*) 36, 42, 53, 67, 71, 74, 77, 80, 81, 85, 97, 105,	**(*n=9*)** 54, 56, 69, 70, 86, 108, st854, stD432, stD633	**(*n=1*)** 18	**22 (29.7%)**
E	**(*n=22*)** 4, 15, 25, 28, 49, 58, 63, 68, 73, 75, 89, 102, 103, 106, 109, 110, 118, st11014, st1389, st1731, st212, st2147	**(*n=14*)** 9, 11, 22, 60, 82, 88, 92, 93, 104, 113, stKNB1, stKNB2, stKNB3, stKNB6	**(*n=3*)** 76, 79, st3211,	**39 (52.7 %)**
Unknown	**(*n=3*)** 100, 122, Nontypable	**(*n=3*)** 65, stKNB4, stKNB5	**(*n=1*)** 87	**7**
Total	**40**	**29**	**5**	**74**

6.2.2. Serum opacity factor gene (*sof*) Determination

All 74 *emm* types tested for family patterns were also tested for the presence of *sof* gene (Table 19). This Table also gives the presence or absence of *sof* gene in individual *emm* types and its association with site of isolation.

Forty two (62.7 %) of the 67 *emm* types with identified family patterns were *sof* gene positive (Table 21). Thirty eight (97.4 %) of the 39 types with family pattern E were also positive for *sof* gene. On the other hand, 19 (86.4 %) of the 22 with family pattern D and 5 (83.3 %) of the 6 types with pattern A-C were negative for *sof* gene.

Table 21. Association of *sof* gene and family patterns of *emm* types.

***emm* Family Pattern**	**Total No of *emm* types**	***sof* gene**	
		+ve	**-ve**
E	39	38	1
D	22	3	19
A - C	6	1	5
Unidentified	7	2	5
Total	74	44	30

6.2.3 Distribution of *emm* types among *sof* gene positive GAS isolates

Forty four (59.5 %) of the 74 *emm* types tested, were positive for *sof* gene. The types were:

4, 25, 28, 49, 58, 63, 68, 73, 74, 75, 81, 85, 89, 102, 103, 106, 109, 110, 118, st11014, st1389, st1731, st212, st2147 (All PT 1solates; $n = 24$)

9, 11, 12, 22, 60, 82, 88, 92, 93, 104, 113, stKNB1, stKNB2, stKNB3, stKNB4, stKNB6 (All NT isolates; $n = 16$)

76, 79, 87, st3211 (All SK isolates; $n = 4$)

6.3 POPULATION GENETICS OF GAS.

6.3.1. Phylogeny of GAS *emm* types.

A phylogenic tree was developed based on whole *emm* gene sequences of 76 different *emm* types identified in this study (Fig. 17). The tree showed two major clusters. The upper cluster consisting of 53 types was larger and divided into 11 sub-clusters labeled 1 to 11. All types of sub clusters 1 to 8 belonged to sub family pattern E and Class II *emm* types. The sub cluster 9 and 10 consisted of ten types which belonged to sub family pattern D and were intermediate to Class I and Class II *emm* types. Cluster 11 had a single type which was quite distinct from the rest of the types.

The lower cluster was smaller and included 23 *emm* types and also divided into 5 sub-clusters labeled 12 to 16. All types of sub clusters 12 to 14 belonged to sub family pattern D and Class I *emm* types. Cluster 15 and 16 had six types which belonged to sub family pattern A-C. Though *emm* 1-2.2 belongs to pattern A-C, it is much closer to pattern D.

6.3.2. Analysis of distance between the clusters of phylogenetic tree.

The analysis of distance between 16 genetic clusters identified in the phylogenetic tree (Fig. 17) was done using the MEGA software (http://www.megasoftware.net/.) (Fig. 18). This matrix is used to depict the distance between the clusters which will also describe the similarity and diversity among the various *emm* types circulating in our population. For example, the distance between cluster 1 and 2 is 0.04; this means that there is 4 % diversity among the *emm* types included in these two clusters and therefore 96 % similarity among them. Further, the distance between cluster 1 and 16

is 0.31 which means a diversity of 31 % and similarity of 67 % only. The matrix also depicts the diversity of the three family patterns, namely E, D and A – C. Interestingly, the unidentified family pattern that falls into cluster 11 lies between pattern D and has a diversity of 35 % (or a similarity of 88 %) with cluster 10 (pattern D) and a 33 % diversity with cluster 12 (also pattern D). It also has a diversity of 37% with Pattern E (cluster 8) and 45 % diversity with pattern A – C (cluster 15).

6.3.3. Conserved region sequences of *emm* types

Gene sequences of the conserved regions of 76 *emm* types out of 77 types are depicted in Fig. 19. The *emm* types of the *emm* family pattern D exhibited two highly divergent genotypes, separated into the two main branches were investigated. The sequences of 10 types including 18.12 (shown in red color) indicate an intermediary position to Class II and I *emm* types.

6.3.4. Sequence comparison of new *emm* types.

The sequences given in Fig. 20 are the multiple alignments of clonally related new *emm* types and their subtypes. The sequence of stKNB2 was different from the subtype stKNB2.1 by a deletion of 9 nucleotides. The sequence of stKNB2.1 was different from stKNB2.2 by transition of just one nucleotide. stKNB1 had 9 base pairs (bp) additional deletions, 1 transition and 2 transversions. Type stKNB3 differ from stKNB1 by additional 30 bp deletions and a transversion.

6.4. PHAGE ASSOCIATED VIRULENCE FACTORS (PAVF)

A total of 218 GAS isolates from four different sources were tested for the presence of PAVF. These included 46 *emm* types from PT, 61 from NT and 62 SK isolates. In addition, 49 GAS isolates recovered from invasive GAS disease were also selected for this study. The invasive GAS strains were made available from an earlier collection of isolates that was already characterized by *emm* typing in another study.

6.4.1. Frequency distribution of PAVF

The frequency distribution of PAVF among 218 GAS isolates is given in Fig. 21. The gene for *SpeH* was seen in 70 (32.1 %) of the total isolates. Factor *sda* was found only in one single isolate while *mf4* and *sdn* were not seen in any isolate.

6.4.2. General Prevalence of PAVF

Forty seven (21.7 %) of the 216 GAS isolates had no PAVF in them (Table 22). One hundred and five isolates (48.6 %) had one or two PAVF among them while 45 (20.8%) had four or more PAVF in them. One single isolate had seven PAVF in it.

Table 22: Multiple PAVF among GAS isolates (*n = 218*)

No. of PAVF	No. of Isolates	% isolates
0	47	21.6
1	48	22
2	57	26.1
3	21	9.6
4	33	15.1
5	9	4.1
6	2	0.9
7	1	0.5

6.4.3 Comparative distribution of PAVF among GAS isolates from different sites

The distribution of PAVF among the four groups of isolates was similar (Table 23). None of the isolates had more than 7 virulence genes in them.

Table 23: Percentage prevalence of PAVF among GAS isolates from four sites

No. of PAVF	% PT Isolates	% NT Isolates	% SK Isolates	% IV Isolates	Total Isolates
0	21.7	19.3	16.4	30.6	47 (21.6%)
1	19.6	27.4	22.9	16.3	48(22.0%)
2	26.1	32.2	29.5	14.3	57(26.1%)
3	10.9	3.2	11.5	14.3	21(9.6%)
4	15.2	9.7	16.4	20.4	3315.1%)
5	4.3	6.4	3.3	2.01	9(4.1%)
6	0	1.6	0	2.0	2(0.9%)
7	2.2	0	0	0	1(0.5%)
Total	46 (21.1%)	62 (28.4%)	61 (27.9%)	49 (22.5%)	218

6.4.4 PAVF profiles among GAS isolates

A PAVF profile was prepared based on the absence or presence of various virulence factor genes (Table 24). Forty four PAVF profiles were identified among the 171 isolates that had one or more virulence factor. Forty seven of the isolates which had no virulence factors, was clustered together as Profile 1. Profiles 2, 3, 4, 5 and 6 which had only one gene each, accounted for 48 isolates (22.0 %). Among them, 22 isolates (10.1 %) belonged to Profile 4 which had *ssa* gene alone in them.

Table 24: PAVF Profiles of PT, NT SK & IV Isolates.

Profile No	Total No. of Isolate	*speC*	*mf2*	*speI*	*speH*	*ssa*	*speL*	*speM*	*mf4*	*mf3*	*sla*	*speK*	*speA*	*sda*	*sdn*	No of PAVF
1	47	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**0**
2	1	-	+	-	-	-	-	-	-	-	-	-	-	-	-	**1**
3	12	-	-	-	+	-	-	-	-	-	-	-	-	-	-	**1**
4	22	-	-	-	-	+	-	-	-	-	-	-	-	-	-	**1**
5	1	-	-	-	-	-	-	-	-	+	-	-	-	-	-	**1**
6	12	-	-	-	-	-	-	-	-	-	-	-	+	-	-	**1**
7	8	+	+	-	-	-	-	-	-	-	-	-	-	-	-	**2**
8	10	-	-	+	+	-	-	-	-	-	-	-	-	-	-	**2**
9	1	-	-	-	+	+	-	-	-	-	-	-	-	-	-	**2**
10	4	-	-	-	+	-	-	-	-	-	-	-	+	-	-	**2**
11	1	-	-	-	-	+	-	+	-	-	-	-	-	-	-	**2**
12	19	-	-	-	-	-	+	+	-	-	-	-	-	-	-	**2**
13	3	-	-	-	-	-	-	+	-	-	+	-	-	-	-	**2**
14	8	-	-	-	-	-	-	-	-	-	+	+	-	-	-	**2**
15	3	-	-	-	-	-	-	-	-	+	-	-	+	-	-	**2**
16	1	+	+	-	+	-	-	-	-	-	-	-	-	-	-	**3**
17	7	+	+	-	-	-	-	-	-	-	-	-	+	-	-	**3**
18	1	-	-	+	+	-	-	-	-	-	-	-	+	-	-	**3**
19	4	-	-	+	+	+	-	-	-	-	-	-	-	-	-	**3**
20	1	-	-	+	+	-	-	-	-	+	-	-	-	-	-	**3**
21	1	-	-	-	+	-	+	+	-	-	-	-	-	-	-	**3**
22	6	-	-	-	-	+	-	-	-	-	+	+	-	-	-	**3**
23	3	-	-	-	-	+	-	-	-	-	+	+	+	-	-	**4**
24	1	-	-	-	-	+	+	+	-	-	-	-	+	-	-	**4**
25	2	+	+	+	+	-	-	-	-	-	-	-	-	-	-	**4**
26	4	+	+	-	-	-	+	+	-	-	-	-	-	-	-	**4**
27	2	+	+	-	+	-	-	-	-	+	-	-	-	-	-	**4**
28	6	+	+	-	+	-	-	-	-	-	-	-	+	-	-	**4**
29	1	+	+	-	-	+	-	-	-	-	-	-	-	+	-	**4**
30	1	-	+	+	+	-	-	-	-	-	-	-	+	-	-	**4**
31	5	-	-	+	+	-	-	+	-	-	+	-	-	-	-	**4**
32	2	-	-	+	+	-	+	+	-	-	-	-	-	-	-	**4**
33	2	-	-	+	+	-	-	-	-	-	+	+	-	-	-	**4**
34	1	-	-	-	+	+	+	+	-	-	-	-	-	-	-	**4**
35	3	-	-	-	+	-	-	-	-	-	+	+	+	-	-	**4**
36	1	+	+	+	+	-	-	-	-	+	-	-	-	-	-	**5**
37	1	+	+	+	+	-	-	-	-	-	-	-	+	-	-	**5**
38	1	-	+	+	+	-	+	+	-	-	-	-	-	-	-	**5**
39	2	-	-	+	+	+	-	-	-	-	+	+	-	-	-	**5**
40	3	-	-	-	+	+	-	-	-	-	+	+	+	-	-	**5**
41	1	+	+	-	-	+	-	-	-	-	+	+	-	-	-	**5**
42	1	+	+	+	+	-	-	-	-	+	-	-	+	-	-	**6**
43	1	+	+	+	+		+	+	-	-	-	-	-	-	-	**6**
44	1	-	-	+	+	+	+	+	-	-	+	+	-	-	-	**7**
Total	**218**	**36**	**39**	**36**	**70**	**47**	**31**	**40**	**0**	**9**	**37**	**29**	**46**	**1**	**0**	

6.4.5 PAVF profiles and sites of GAS isolates

Analysis of various PAVF profiles in relation to the sites from which the GAS isolates were recovered did not show any unique characteristic distribution (Table 25). However, ten different profiles were found in all four groups of isolates (PT, NT, SK & IV).

Table 25: PAVF profiles of GAS isolates from three different sites

PAVF Profile in	PAVF Profile	Total Profiles (n = 44)
Only PT	16, 30, 34, 36,44	5
Only NT	5, 9, 11, 29, 37, 38, 42	7
Only SK	18, 20, 21	3
Only IV	2, 23, 24, 41, 43	5
PT & NT	13, 32	2
NT & SK	10, 15, 25, 39	4
SK & IV	26, 27, 33	3
PT, NT & SK	28, 35, 40	3
PT, SK & IV	17, 19	2
PT, NT, SK, IV	1, 3,4, 6, 7, 8, 12,14, 22, 31	10

6.4.6 Comparative distribution of individual PAVF among PT, NT, SK and IV isolates.

The number of isolates with no PAVF was highest in invasive and PT isolates (Fig. 22). Number of isolates with one or two PAVF was highest among SK and NT isolates. Number of isolates with 4 or 5 PAVF was highest among invasive isolates though not very unique. One invasive and one SK isolate had 6 PAVF in them. The only isolate with 7 PAVF was found in a PT isolate.

6.4.7. PAVF profiles of *emm* types of PT isolates

Twenty two profiles were identified among 46 *emm* types used to create a virulence gene profile (Table 26). Profile 1 with no of the virulence factors accounted for 10 types/isolates. Most frequently found four profiles, namely Profile 1, 4 and 12 (n = 18) accounted for 39.4 % of the pharyngitis isolates, each with no, 1or 2 virulence factors. Twelve different single profiles were found only once, which included the highest number of virulence factor profile with 7 virulence genes.

Table 26: PAVF Profiles of PT isolates.

Profile No	*emm* types	*speC*	*mf2*	*speI*	*speH*	*ssa*	*speL*	*speM*	*mf4*	*mf3*	*sla*	*speK*	*speA*	*sda*	*sdn*	No of PAVF
1	**28,68,77,73,85, st62,st1389,st212, st11014,stKNB2**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**0**
3	**44/61, 81**	-	-	-	+	-	-	-	-	-	-	-	-	-	-	**1**
4	**22,39,75,118**	-	-	-	-	+	-	-	-	-	-	-	-	-	-	**1**
6	**1,3,109**	-	-	-	-	-	-	-	-	-	-	-	+	-	-	**1**
7	**42**	+	+	-	-	-	-	-	-	-	-	-	-	-	-	**2**
8	**102,st1731**	-	-	+	+	-	-	-	-	-	-	-	-	-	-	**2**
12	**63,80,89,100**	-	-	-	-	-	+	+	-	-	-	-	-	-	-	**2**
13	**97,NT**	-	-	-	-	-	-	+	-	-	+	-	-	-	-	**2**
14	**36,58,106**	-	-	-	-	-	-	-	-	-	+	+	-	-	-	**2**
16	**103**	+	+	-	+	-	-	-	-	-	-	-	-	-	-	**3**
17	**15**	+	+	-	-	-	-	-	-	-	-	-	+	-	-	**3**
19	**25,105**	-	-	+	+	+	-	-	-	-	-	-	-	-	-	**3**
22	**4**	-	-	-	-	+	-	-	-	-	+	+	-	-	-	**3**
28	**67,74**	+	+	-	+	-	-	-	-	-	-	-	+	-	-	**4**
30	**71**	-	+	+	+	-	-	-	-	-	-	-	+	-	-	**4**
31	**53**	-	-	+	+	-	-	+	-	-	+	-	-	-	-	**4**
32	**49**	-	-	+	+	-	+	+	-	-	-	-	-	-	-	**4**
34	**110**	-	-	-	+	+	+	+	-	-	-	-	-	-	-	**4**
35	**122**	-	-	-	+	-	-	-	-	-	+	+	+	-	-	**4**
36	**55**	+	+	+	+	-	-	-	-	+	-	-	-	-	-	**5**
40	**st854**	-	-	-	+	+	-	-	-	-	+	+	+	-	-	**5**
44	**st2147**	-	-	+	+	+	+	+	-	-	+	+	-	-	-	**7**
Total	**46**	**6**	**7**	**9**	**17**	**10**	**7**	**10**	**0**	**1**	**10**	**7**	**9**	**0**	**0**	

NT = Nontypable

6.4.8. PAVF profiles among NT isolates

PAVF gene profiles of NT isolates were analyzed to create a virulence gene profile (Table 27). Totally 26 different PAVF gene profiles were identified among the 62 isolates representing as many *emm* types. These included those which had none of the virulence factors. Most frequently found four profiles namely, Profile 1, 4, 6 and 12 (*n = 26*) accounted for 41.9 % of the NT isolates. These had no, 1or 2 virulence factors. Fifteen profiles were found only once, which included one of the isolates with 6 virulence factors.

Table 27: PAVF Profiles of NT isolates.

Normal throat Profile No	*emm* types	speC	mf2	speI	speH	ssa	speL	speM	mf4	mf3	sla	speK	speA	sda	sdn	No of PAVF
1	8, 28, 42, 63, 65, 77, 88, 110, 113, st1389, stKNB1, stKNB2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**0**
3 (4)	44/61, 81, stKNB2, stKNB3	-	-	-	+	-	-	-	-	-	-	-	-	-	-	**1**
4 (7)	22, 56, 57, 60, 81, 118, st1731	-	-	-	-	+	-	-	-	-	-	-	-	-	-	**1**
5 (1)	st11014	-	-	-	-	-	-	-	-	+	-	-	-	-	-	**1**
6 (5)	1, 3, 11, 92, 104,	-	-	-	-	-	-	-	-	-	-	-	+	-	-	**1**
7 (3)	1, 58, 69	+	+	-	-	-	-	-	-	-	-	-	-	-	-	**2**
8 (2)	88, 93	-	-	+	+	-	-	-	-	-	-	-	-	-	-	**2**
9 (1)	stD633	-	-	-	+	+	-	-	-	-	-	-	-	-	-	**2**
10 (2)	75, 103	-	-	-	+	-	-	-	-	-	-	-	+	-	-	**2**
11 (1)	105	-	-	-	-	+	-	+	-	-	-	-	-	-	-	**2**
12 (5)	49, 54, 80, 89, stKNB6	-	-	-	-	-	+	+	-	-	-	-	-	-	-	**2**
13 (1)	97	-	-	-	-	-	-	+	-	-	+	-	-	-	-	**2**
14 (3)	36, 85, stKNB5	-	-	-	-	-	-	-	-	-	+	+	-	-	-	**2**
15 (2)	4, 63	-	-	-	-	-	-	-	-	+	-	-	+	-	-	**2**
22 (2)	70, 86	-	-	-	-	+	-	-	-	-	+	+	-	-	-	**3**
25 (1)	82	+	+	+	+	-	-	-	-	-	-	-	-	-	-	**4**
28 (1)	67	+	+	-	+	-	-	-	-	-	-	-	+	-	-	**4**
29 (1)	12	+	+	-	-	+	-	-	-	-	-	-	-	+	-	**4**
31 (1)	53	-	-	+	+	-	-	+	-	-	+	-	-	-	-	**4**
32 (1)	74	-	-	+	+	-	+	+	-	-	-	-	-	-	-	**4**
35 (1)	122	-	-	-	+	-	-	-	-	-	+	+	+	-	-	**4**
37 (1)	71	+	+	+	+	-	-	-	-	-	-	-	+	-	-	**5**
38 (1)	49	-	+	+	+	-	+	+	-	-	-	-	-	-	-	**5**
39 (1)	st2147	-	-	+	+	+	-	-	-	-	+	+	-	-	-	**5**
40 (1)	st854	-	-	-	+	+	-	-	-	-	+	+	+	-	-	**5**
42 (1)	55	+	+	+	+	-	-	-	-	+	-	-	+	-	-	**6**
Total	**62**	**8**	**9**	**9**	**19**	**14**	**7**	**10**	**0**	**4**	**10**	**8**	**14**	**1**	**0**	

6.4.9. PAVF profiles among SK isolates

PAVF gene profiles identified among SK isolates were analyzed to create a virulence gene profile (Table 28). Twenty five profiles were identified among the 61SK isolates including those which had none of the virulence factors. Most frequently found three 1, 12, 4 and 3 (*n = 30*) profiles accounted for 49.1 % of the SK isolates, each with 0, 1or 2 virulence factor. Fourteen profiles were found only once. None of the SK isolates had >5 virulence factor genes.

Table 28: PAVF Profiles of Impetigo isolates

Impetigo Profile No	No of isolates	speC	mf2	speI	speH	ssa	speL	speM	mf4	mf3	sla	speK	speA	sda	sdn	No of PAVF
1 (10)	4, 8, 28, 29, 31, 63, st62, 85, st1389, st11014	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**0**
3 (4)	44/61, 81, 110, stKNB2	-	-	-	+	-	-	-	-	-	-	-	-	-	-	**1**
4 (7)	25, 39, 57, 60, 105, 118, st1731	-	-	-	-	+	-	-	-	-	-	-	-	-	-	**1**
6 (3)	3, 11, 109	-	-	-	-	-	-	-	-	-	-	-	+	-	-	**1**
7 (3)	42, 58(2)	+	+	-	-	-	-	-	-	-	-	-	-	-	-	**2**
8 (2)	65, 69	-	-	+	+	-	-	-	-	-	-	-	-	-	-	**2**
10 (2)	18, 75	-	-	-	+	-	-	-	-	-	-	-	+	-	-	**2**
12 (9)	49, 54, 76, 77, 80, 89, 100, 108, st3211	-	-	-	-	-	+	+	-	-	-	-	-	-	-	**2**
14 (1)	9	-	-	-	-	-	-	-	-	-	+	+	-	-	-	**2**
15 (1)	63	-	-	-	-	-	-	-	-	+	-	-	+	-	-	**2**
17 (1)	92	+	+	-	-	-	-	-	-	-	-	-	+	-	-	**3**
18 (1)	71	-	-	+	+	-	-	-	-	-	-	-	+	-	-	**3**
19 (1)	25	-	-	+	+	+	-	-	-	-	-	-	-	-	-	**3**
20 (1)	87	-	-	+	+	-	-	-	-	+	-	-	-	-	-	**3**
21 (1)	49	-	-	-	+	-	+	+	-	-	-	-	-	-	-	**3**
22 (2)	70, 86	-	-	-	-	+	-	-	-	-	+	+	-	-	-	**3**
25 (1)	82	+	+	+	+	-	-	-	-	-	-	-	-	-	-	**4**
26 (1)	93	+	+	-	-	-	+	+	-	-	-	-	-	-	-	**4**
27 (1)	81	+	+	-	+	-	-	-	-	+	-	-	-	-	-	**4**
28 (3)	15(2), 74	+	+	-	+	-	-	-	-	-	-	-	+	-	-	**4**
31 (2)	53(2)	-	-	+	+	-	-	+	-	-	+	-	-	-	-	**4**
33 (1)	56	-	-	+	+	-	-	-	-	-	+	+	-	-	-	**4**
35 (1)	122	-	-	-	+	-	-	-	-	-	+	+	+	-	-	**4**
39 (1)	st2147	-	-	+	+	+	-	-	-	-	+	+	-	-	-	**5**
40 (1)	st854	-	-	-	+	+	-	-	-	-	+	+	+	-	-	**5**
Total	**61**	**10**	**10**	**10**	**23**	**12**	**11**	**13**	**0**	**3**	**9**	**7**	**13**	**0**	**0**	

6.4.10. PAVF profiles among Invasive isolates

PAVF genes identified from invasive isolates were analyzed to create a virulence gene profile (Table 29). Twenty profiles were identified among 49 invasive isolates including those which had none of the virulence factors. Most frequently found four profiles namely, 1, 17, 4 and 8 ($n = 28$) accounted for 57.1 % of the invasive isolates, each with 0, 1, 2 or 3 virulence factor genes. Thirteen profiles were found only once including one isolate with 6 virulence factor genes.

Table 29: PAVF Profiles of Invasive isolates.

Invasive Profile No	*emm* types	speC	mf2	speI	speH	ssa	speL	speM	mf4	mf3	sla	speK	speA	sda	sdn	No. of PAVF
1 (15)	8, 28, 49, 63(2), 85(2), 104,110, 124(2), st1731, st6735(2) , stKNB9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**0**
2 (1)	58	-	+	-	-	-	-	-	-	-	-	-	-	-	-	**1**
3 (2)	44/61, 110	-	-	-	+	-	-	-	-	-	-	-	-	-	-	**1**
4 (4)	60, 105(2), 118	-	-	-	-	+	-	-	-	-	-	-	-	-	-	**1**
6 (1)	18	-	-	-	-	-	-	-	-	-	-	-	+	-	-	**1**
7 (1)	110	+	+	-	-	-	-	-	-	-	-	-	-	-	-	**2**
8 (4)	85, 102(2), 110	-	-	+	+	-	-	-	-	-	-	-	-	-	-	**2**
12 (1)	stKNB6	-	-	-	-	-	+	+	-	-	-	-	-	-	-	**2**
14 (1)	81	-	-	-	-	-	-	-	-	-	+	+	-	-	-	**2**
17 (5)	15, 74(3), 112	+	+	-	-	-	-	-	-	-	-	-	+	-	-	**3**
19 (1)	25	-	-	+	+	+	-	-	-	-	-	-	-	-	-	**3**
22 (1)	86	-	-	-	-	+	-	-	-	-	+	+	-	-	-	**3**
23 (3)	st854(3)	-	-	-	-	+	-	-	-	-	+	+	+	-	-	**4**
24 (1)	119	-	-	-	-	+	+	+	-	-	-	-	+	-	-	**4**
26 (3)	43, 100(2)	+	+	-	-	-	+	+	-	-	-	-	-	-	-	**4**
27 (1)	55	+	+	-	+	-	-	-	-	+	-	-	-	-	-	**4**
31 (1)	53	-	-	+	+	-	-	+	-	-	+	-	-	-	-	**4**
33 (1)	56	-	-	+	+	-	-	-	-	-	+	+	-	-	-	**4**
41 (1)	86	+	+	-	-	+	-	-	-	-	+	+	-	-	-	**5**
43 (1)	93	+	+	+	+	-	+	+	-	-	-	-	-	-	-	**6**
Total	**49**	**12**	**13**	**8**	**11**	**11**	**6**	**7**	**0**	**1**	**8**	**7**	**10**	**0**	**0**	

6.4.11. Comparison of PAVF profiles of community and invasive isolates

A comparison of prevalence of various PAVF genes showed that the highest prevalence of isolates with no PAVF was found among invasive isolates (Table 30). Further, prevalence of three or more virulence genes was also higher among invasive isolates.

Table 30: Comparative prevalence of PAVF genes among PT, NT, SK and Invasive isolates

Number of PAVF genes	% PT Isolates	% NT Isolates	% SK Isolates	% Invasive Isolates
0	21.7	19.4	16.4	30.6
1	2.2	27.4	22.9	16.3
2	26.1	32.2	29.5	14.3
3	10.9	3.2	11.5	14.3
4	15.2	9.7	16.4	20.4
5	4.3	6.5	3.3	2.0
6	0	1.6	0	2.0
7	2.2	0	0	0
Total	**46**	**62**	**61**	**49**

6.4.12. Distribution of PAVF among different groups of isolates

The distribution of the 14 PAVF's in four groups of GAS isolates is compared in Fig. 23. The commonest PAVF, *speH,* was the most predominant among PT, NT and SK isolates while *speC* and *mf2* were the most predominant among IV isolates. The factor *sda* was seen only in one single NT isolate.

6.5 CHROMOSOMALLY CODED VIRULENCE FACTORS (CCVF).

6.5.1. General Prevalence of CCVF.

A total of 140 GAS isolates including 46 PT isolates, 56 SK isolates, 38 NT isolates were tested for the presence of 22 CCVF genes (Fig 24). Seven toxin genes were identified among 138 – 140 isolates while two toxin genes namely, *speJ* and *smeZ* were seen only in 2 and 53 isolates respectively. The *sic* gene was identified in 15 isolates. The latter included *emm* types 1 (*n =4*), 3 (*n =2*), 12 (*n =1*), 31 (*n =1*), 55 (*n =2*), 57(*n =2*), 39 (*n =2*) all of which belonged to family pattern, A-C and type 28 (*n = 1*) which belonged to pattern E.

Among fibronectin binding proteins, *fba* was seen in 128 isolates and *sciB* in 136 isolates (Fig 24).

6.5.2. Comparison of CCVF among PT, NT & SK isolates

Fig. 24 gives the comparative distribution of 10 chromocomally mediated toxin genes and 12 chromosomally mediated fibronectin binding proteins (FBP). Among the 10 toxin genes,, speJ and smeZ were present in 5 % to 22 % of isolates. The *sic* gene was present in 15 (10.7 %) of the 140 GAS isolates tested. Of the 12 FBP tested for, two namely, *fba* and *sciB* were predominant among the three groups of isolates (Fig. 25). The gene *fbp54* which was present only in one of the 46 PT isolates and two of the 56 SK isolates was absent in all 36 NT isolates. Of the three *prtf* genes, *prtf1* was absent in all three groups of GAS isolates while *prtf2* was predominant among SK isolates.

7. DISCUSSION

Knowledge about the epidemiology and pathogenesis of microbial infections in any community are important for designing appropriate intervention strategies. This is especially true of GAS infections in countries like India because of the serious complications that they can cause. Since school age children and young adults are the most affected population for GAS infection, epidemiological studies are generally conducted among school going children [3, 7]. For this reason the present study was aimed at the molecular characterization of GAS isolates associated with human infections, those recovered from children aged 7-11 years and attending a rural school situated 10 Km south west of Vellore town were selected for the same.

The overall aim of this study was to characterize diverse GAS isolates recovered from children with throat and skin infections as well those which colonize throat of normal healthy children by molecular methods. The objective was to identify strain variations among diverse GAS isolates circulating and causing infections in southern India, which is highly endemic for GAS infections. This will also characterize the isolates in such a way that one will be able to determine the types predominantly associated with throat and skin infections. Knowledge of the *emm* types that colonize the throat of normal healthy individuals will help to correlate the association of colonizing types with throat and skin infecting types. Further, information on specific types present in our community will help to determine if the 26-valent M type based vaccine [248] will be suitable for the Indian population.

Most reported literature on GAS infections are based on cross sectional studies in which the epidemiological pattern of GAS infection and transmission can not be adequately represented [103]. Such studies include strains collected only from

individuals with disease or without appropriate control strains or else, no control strains were collected in a representative fashion [249]. Studies on the characterization of Indian GAS isolates by *emm* typing are few and far in between 11 - 94; and only very few reports are based on properly designed epidemiological studies [39, 244, 250-252]. The present study was a school based cohort study and the sample size was calculated as 707; it was possible to successfully characterize 698 GAS isolates for meeting the overall objective. To the best of knowledge, this is the first of its kind in India and is one of the largest series of GAS isolates subjected to *emm* typing in the world.

The second aim to evaluate the existing system of population genetics and the development of a phylogenetic background to study the pathogenesis and tissue tropism. This was based on the *emm* family pattern as reported from studies in the western countries [25]; however, studies from tropical developing countries show a different pattern [30]. Such studies which have not been undertaken in India so far will reveal the true association of *emm* family pattern and tissue tropism in an endemic situation. The present study also looked at the relationship between *emm* types and family patterns through a phylogenetic tree; this revealed a remarkable clustering of certain *emm* types with specific family patterns. This information gave certain insight into the evolution of *emm* types in an endemic community.

Finally, screening for PAVF and CCVF among south Indian GAS isolates showed extreme heterogeneity of virulence genes among them. Subtle variations through acquisition and/or deletion of virulence genes among the isolates resulted in the development of newer subtypes even among the same *emm* type. Distribution of a wide variety of these virulence genes shows the spread of potential pathogenic strains

in this population. This study also showed that PAVF and CCVF profiles could be used to study the magnitude of strain variations, genetic heterogeneity and population genetics of GAS isolates in an endemic situation.

7.1.1. Selection of GAS isolates

Distribution of 698 GAS isolates selected for this study reflected the general epidemiological features of GAS infections in this south Indian community (Table 1). GAS isolates colonizing throat of normal healthy individuals (n=306) was far higher than those associated with pharyngitis (PT = 116) or impetigo (SK = 276). This is all the more interesting because throat colonization was studied on a monthly basis while pharyngitis and impetigo were looked for, on a weekly basis. These results show that high throat colonization rate indirectly reflects the endemicity of infection in the community.

7.1.2. Diversity of *emm* types among south Indian GAS isolates

Identification of 77 *emm* types (along with one non typable isolate) among 698 isolates reflects the high heterogeneity of strains circulating in this community (Fig 26). To the best of knowledge, this is the largest series, of prospectively collected and community based GAS isolates in the world that were subjected to *emm* typing. In a study reported from Ethiopia, Tewodros and Kronvall identified 78 different *emm* /st types among 217 GAS isolates recovered from children showing much higher heterogeneity among Ethiopian isolates as compared to the south Indian isolates [16]. All Indian studies reported so far comprised of much smaller number of isolates and therefore were not sufficiently representative of strains circulating in the communities [39, 250-253]. Similarly, in a small study conducted in China, Jing et al. reported 21 *emm* types among 86 GAS isolates [254]. In contrast, a large surveillance study

conducted in pediatric primary care offices at diverse sites in USA (Fig. 27), during 2000-2002 have shown only 38 *emm* types from a collection of 1975 isolates [15]. Another study conducted in the air force base at San Antonio, Texas, reported 35 *emm* types among 675 GAS isolates [255] while in a surveillance study of invasive GAS disease, only 41 *emm* types were found among 1,061 isolates [256]. Similar findings were reported from Canada, Mexico, Great Britain and other developed countries [26, 257-260]. The difference in the diversity of *emm* types was clearly demonstrated in a comparative study of GAS isolates from Belgium (Industrialized country) and Brazil (developing country) (Fig. 28). This study showed that 20 *emm* types were identified among 200 Belgian GAS isolates, while 48 *emm* types were identified from 128 Brazilian isolates [261]. Thus it is very clear that the *emm* type distribution in regions endemic to GAS infections is much more heterogeneous than in temperate western countries, thus confirming the contrasting epidemiology of these infections in developed and developing countries.

The epidemiological differences of GAS infections reflected by the genetic diversity of GAS is associated with living conditions, hygiene and the effective medical care [25]. In the developed countries where the GAS infection have been effectively controlled, the increase in incidence of GAS infections is due to epidemic emergence of certain virulent clones of GAS [29, 94, 262]. Effective control in the chain of transmission of GAS strains in such population possibly resulted in eradicating most of the different types. The few types that survived were able to expand clonally by evolution and emerge into highly virulent clones. In countries like India where a substantial susceptible population exists, diverse types are able to successfully survive and circulate in the community. This also gives sufficient opportunity for the organisms to transfer genetic elements and gives rise to newer genotypes.

The theory of evolutionary history, origin and migration of human beings hypothesize that Africa is the place of human origin, and the migration took place through South Asia to other parts of the world [263]. The human ancestry can be identified by allele frequency derived haplotype frame works, suggesting that differently expressed phenotype evolved relatively early in the human history [264]. It is interesting that the observed global GAS epidemiology is consistent with what is observed in the human population genetics. The *emm* types in Ethiopia are the most heterogeneous among all reports, followed by the present Indian study and the reports from Nepal, and then the rest of the world [16, 22, 39]. This gradual decline in heterogeneity among GAS genotypes is in line with the theories of human origin and migration and probably has a strong genetic background as well [16, 263].

7.1.3. Differences in the prevalence of *emm* types among south Indian GAS isolates

The distribution of *emm* types identified in the present is quite different from most other reports including those from the northern part of India. Eleven most predominant *emm* types, namely 63, 81, 28, 49, 122, 44/61, 82, st854, 42, 118 and st1731 has not been reported as the predominating types in other countries, except *emm* 28. In USA, types 1, 2, 3, 4, 6, 12, 18 and 28 were the most predominant types [15]. Other studies have also confirmed the predominance of these types [36, 255]; although the order of frequency may be different. The other more frequently encountered types in the USA, such as 89, 77, 75, 22, 44/61, 5 and 11 were rarely encountered in the present series, with the possible exception of type 44/61.

Reported literature confirms that countries differ in the distribution of predominant *emm* types [16, 22, 250, 265]. Studies from Mexico have shown that the type

distribution in that country reflects those encountered in the US [258] which is probably associated with the migration of people across the two countries. Also, reports from France, Denmark and other European countries show strong similarities in the distribution of predominant types with the American types [26, 266-268]. Japanese studies also confirm such similarities [260] thereby indicating that industrialized countries have a very similar epidemiology of GAS infections. Chinese, Ethiopian and Nepalese studies indicate that *emm* type 1 is a common predominant type seen in these countries [254] while in the present study type 1 accounted for only 2.6 % of the 698 isolates studied and occupied the 13th position out of a total of 78. Further analysis of *emm* type data of this study with Ethiopia and US data has shown that, 37 and 54 types respectively were not seen in those countries [15, 16], indicating that many *emm* types are unique for this region of the world.

7.1.4. Distribution of predominant *emm* types among south Indian GAS isolates.

In the present study, the most frequently occurring 11 *emm* types accounted for 39 % of the 698 isolates while 17 types accounted for 50 % of the isolates. This indicates the absence of any one type being very predominant. Further, the most predominant type, i.e. type 63, accounted for only 4.6 % of all isolates. This result indicates that in endemic country like India, no particular type can dominate and the exact reason is not known.

In a study conducted in US on the isolates collected during 1988-1990, six types collectively accounted for 60 % of the 866 isolates obtained from throat cultures of patients with uncomplicated pharyngitis [36]. The *emm* types shown from another surveillance of GAS pharyngitis isolates, obtained in between 2000-2002 in USA, had 3 predominant *emm* types, accounting for 49.2 % of the total 1975 studied [15]. A

small study conducted in a semi-closed community of about 500 children, 92 % of the 111 GAS isolates were M1 during the first half of the study, and during the latter half, 84% of the 126 GAS isolates were M6 type where M1 was 13 % [269]. In Spain the 7 most predominant *emm* types accounted for 61.9 % of the 126 *emm* types, same as that of USA, Canada and Europe [270]. In China 6 *emm* types accounted for 74.6% of the 86 strains tested [254]. These results go on to show that in most non-endemic regions, a few types tend to predominant significantly, which however can change from year to year. Even in the Ethiopian study, 12 different *emm* types made up 40.1 % of the 217 GAS isolates obtained from different sources [16]. This is very similar to the present study where 11 most frequently seen types accounted for 39% of 698 isolates. This confirms that in endemic regions, many *emm* types circulate in the community with a relative lack of predominant types among them.

In a report from North India, 15 *emm* types accounted for 69.4 % of the 59 GAS isolates [39]. This seems to be a higher level of predominance than what is observed from southern India where 15 *emm* types accounted for only 49.3 % of the 698 isolates. However, the sample size is not comparable and the study design differs between the studies.

The epidemiological difference in the predominance of certain *emm* types were also shown in a study conducted from the isolates of Belgium (Industrial Country) and Brazil (developing country). Six most predominant *emm* types accounted for 75 % of the 200 Belgian GAS isolates, whereas the most predominant 6 *emm* types accounted for only 36 % of the 128 Brazilian GAS isolates. This study is a good representation for the epidemiological differences in the predominance of *emm* types in a developed versus developing countries [261].

The predominance of particular genotypes found in this study is lesser than many other studies, but similar to that of the Ethiopian data (Fig. 29) [16]. The relative lack of predominance of any *emm* type shows that there may not be any clone with enhanced virulence, and if given a chance any strain can cause infection according to the favorable environmental conditions and the genetic makeup of the strain. The other possibility is that the so called highly virulent clones would not have reached these places to dominate among the GAS population. For example, the *emm* type 1 (*emm* 1-2.2 and *emm* 1.4) identified from this study were considered as different clones than what was found most commonly in western countries [250], that is described in the CDC website http://www.cdc.gov /ncidod/biotech/strep/types_*emm* 103-124.htm.

One other possible reason for lack of predominance of *emm* types in this population may be due to, the antibodies raised against the epitopes from other than hyper variable region and the circulation of multitudes of *emm* types [271]. The antibodies against the conserved region epitopes or the other immunogenic surface proteins might contribute to the prevalence of any type in a community. This means that the exposure of *emm* types that are closely related other *emm* types will result in the control of such *emm* types in the same individual. Therefore the *emm* types other than closely related once can only cause infection when it is exposed, which will be always there in endemic regions like India with poor hygienic conditions and crowed population. This might be also a reason for high diversity of *emm* types in this community.

7.1.5. Validated types, sequence types and new types

In this study 602 (86.2 %) of the total 698 isolates belongs to 61 validated *emm* types and 83 isolates (11.9 %) belongs to 10 provisional sequence types excluding 12 isolates (1.7 %) of 6 new types. Similar findings have been observed in few studies reported from southern and northern India as well Ethiopia [16, 250-252]. The number of provisional types identified in this study is relatively higher than many other reports. These results are somewhat similar to the Ethiopian study where validated types accounted for 76.5 % and sequence types for 11.9 %; interestingly, 10 new types accounting for 11.5 % of 217 isolates [16]. In a study done in USA during 2000-2002 on 1975 GAS isolates, almost all GAS *emm* types except 3 new types were within the 117 validated types recognized by the international panel [15]. In Canada, non-typable isolates accounted for 15.4 % of the GAS isolates before the update in the validated *emm* types in CDC database [257]. In Mexico only one type was sequence type among 27 *emm* types identified in the study [258]. In UK, only 4 provisional types were identified even among the characterization of M nontypable isolates. Similar results were reported from Spain, Italy, China and Japan where only one no sequence types were reported [254, 270, 272, 273]. In short, most of the reported literature shows significant differences in the prevalence of (not validated) sequence types and new types between non endemic and endemic regions [18]. Not validated sequence types are relatively new types that have been encountered only in certain geographical areas and therefore have not been assigned distinct *emm* type status. They are more predominantly seen in endemic regions and probably restricted by geographic and temporal barriers.

There is nearly always a 1:1 relationship between 5' *emm* sequence types and M serotypes, and the terms "M types" and "*emm* types" refer to serotypes and 5' *emm* sequence types, respectively. There are 83 GAS M serotypes unequivocally acknowledged by the authors to be both serologically unique and encoded by unique *emm* gene sequences, designated within the order of M1 to M93. Although serologic characterization of M proteins provides extremely valuable information, it is no longer practical for timely and much-needed extension of the Lancefield GAS classification system. Therefore, it is expedient to recognize potential new M protein serotypes by virtue of new *emm* sequence types alone. A working committee reported on an exchange of 13 strains of GAS reportedly having novel *emm* sequence types, 8 of which had *emm* sequences that passed all the criteria for new types and were designated types *emm 94* to *emm 102.* A latter report on an exchange of strains between the 6 GAS reference laboratories resulted in the addition 22 sequence types designated *emm 103* to *emm 124.* All other sequences reported to the curator of the CDC data base as new types were given a provisional name starting with st which stands for sequence type. So the time at which the study was conducted should be considered for analyzing the type difference because the sequence types in the studies that were conducted before including the newly validated types may show all sequence types with different designation. The correct data can be obtained only after updating the type designation. More than 70 sequence types in the CDC database that have been documented for GAS. There is an increasing numbers of sequence types added to the database [17]. This observation indicate that there are still more undocumented *emm* types with genetic diversity exist in the places were not much studies were done.

Over the years, there has been an increasing usage of the CDC *emm*data base while the rate of deposition of new sequences have decreased indicating that much of the existing genetic diversity within GAS has already been documented [22]. More studies especially from endemic regions will be needed to identify the extent of genetic diversity in the GAS population.

7.1.6. New *emm* types among south Indian GAS isolates

A new type of GAS is defined by the *emm* sequence obtained with primer 1 or *emm* seq2 on the basis of sharing less than 92 % sequence identity over the first 90 bases encoding the deduced processed M protein of the *emm* type reference sequence in the database. If the sequence of any strain does not show 92 % or more identity with the existing *emm* type sequence it shows that the strain is different in its epitope of the M protein. There were 6 new types, stKNB1 to stKNB6, accounting for 1.7 % of all isolates from this study, while Only 3 new types were identified among 1975 isolates, in a study from US [15].

In Thailand 13 new types were reported from 53 nontypable isolates by M typing method before the extension of from *emm* 103 to *emm* 124 in the CDC database [32]. In Brazil too 13 new types were reported from randomly selected isolates [274]. It is in contrary to reports from Israel, China and Japan where there was no new *emm* types identified [254]. In a study from North India, there was one new type identified among 59 isolates consisting of 33 different types [39]. Thus development of new types is a very random event and seems to have no association with any particular geographic region.

The development of new types leads to antigenic variations, which may occur due to the immune pressure towards a particular pathogen during the course of infection. A change in the epitope of the antigen makes the developed immunity to become ineffective, allowing the pathogen to evade the immune system and establish infection. The identified sequence changes were predominantly due to deletions or insertions combined with point mutations, either transitions or transversions in the hyper variable region. A non-synonymous mutation has caused the formation of a new subtype stKNB2.1 from another subtype stKNB2.2. One of the new subtypes of a new type stKNB4 identified in this study was reported (ftp://ftp.cdc.gov/pub/ infectious_diseases/biotech/ts*emm*/stKNB4.1.sds) from US. This shows that the new types identified in this study has relationship with strains from other parts of the world [22].

Though the new types were represented by very low frequency of isolates in the community, development of such types are advantageous to the organism. Genetic variations in the *emm* gene may happen rapidly and randomly in many numbers; these such changes may not affect the survival of the strain unless it leads to change in the basic conformation of the protein that affect the virulence properties [275]. Even though they get propagated, only a minority of mutant become fixed due to positive selection. Organisms are generally so well adapted to the environment that many non-synonymous changes are deleterious and, therefore, quickly removed from the population by negative selection [45]. However the mutations are advantageous for the survival of the organism in the case of population with immunity against the existing types. This mechanism is the primary way in which these human specific pathogens are maintained in the population over time.

7.1.7. Nontypable south Indian GAS isolates

In this study only one strain was nontypable using the specified primer set defined for the amplification of *emm* gene. This shows that even though mutant strains may develop naturally, it may not survive due to the negative selection. The nontypable strain obtained from this study would have been isolated by chance. One of the studies from north India has shown 3 nontypable isolates from the total of 94 isolates [250]. In USA only one nontypable strain was identified among 1975 pharyngitis isolates studied by *emm* typing [15]. That was predicted to be a nonfunctional M protein. This data shows that the strains that we isolate may not be always the real cause of the disease because in many cases more than one type may present in the throat either due to introduction of another strain separately or due to the development of new strains by mutation [275]. The strain that lack M protein might have risen due to error in the replication of the genome during multiplication. It has been shown that the strain that lack M protein can not with stand the immune system of the host [150]. But if this mutated strain is present along with other wild type strain then there is chance for the survival due to the antiphagocytic environment provided by the neighboring cells. In such case, the mutated strain would have also been picked up and during pure culture of the organism the colony formed by the mutated strain would have been chosen for characterization.

7.1.8. Subtypes among *emm* types of south Indian GAS isolates

In the present study 63 of the 77 types were single subtypes. Such large a number of *emm* types with a single subtype shows that the immunity developed against one subtype is generally effective against other subtypes; hence the newly formed subtypes or the freshly introduced subtypes were constantly eliminated by the

immune system and the expansion of those subtypes were controlled [276, 277]. This is not so for all subtypes; there are reports saying that the immunity developed against one subtype may not be effective in all cases [278]. This is confirmed by our finding that 14 of the 77 types had >1 subtype. An interesting observation has been that a new type, stKNB2 (n = 6) had three subtypes. Type stKNB2 is new type that has been identified for the first time in literature in the present study. To have three variants during such a short time clearly shows that the original strains underwent mutation very fast, presumably to evade host related factors.

The factor/s that leads to the formation of new subtypes is not known. It is also difficult to identify which strain came first and the exact mechanism by which it became different [154]. The first reported type is given the designation as the type reference strain (with subtype designation as 0) while subsequent variants are designated as subtypes 1, 2, etc. Forty of the 77 *emm* types identified from this study were same as that of the original reference strain in the CDC data base. Identification of the subtype may be important epidemiologically to show the transmission and geographic differences among *emm* type variants.

Thirty five of the 95 subtypes in the present study were recognized subtypes. [17]. This is the first study showing the highest number of subtypes globally. It is possible that either a large proportion of the variants have developed in the later years or that the subtypes existed only in this region of the world. Sub types in five *emm* types (*emm* 42.3, 71.1, 88.3, 122.1, st1389.2) were identified for the first time, in this study. In the Ethiopian study, 32 of the 86 subtypes identified were new [16]. Development of new subtypes therefore seems to be a feature of endemic regions which is designed by the original strain to overcome host factors.

It is interesting that the number of subtypes per *emm* type found in this study is very less when compared to western countries. There were 11 *emm* types with 2 subtypes and 3 *emm* types with 3 subtypes. This is similar to reports from North India, Ethiopia and Nepal [39, 250, 251]. In geographical areas, where a few *emm* types predominate, such types have many subtypes. For example, reports from US show that type 12 had 32 subtypes while *emm* 89 had 14 subtypes [15]. Thus, in such conditions, where only certain *emm* types predominate in the community, expansion of clones takes place by the development of new subtypes rather than new *emm* types. Four hundred and seventy one (67.5 %) of the 698 GAS isolates had no or only one subtypes were identified in this study. These represented 63 *emm* types. The *emm* types with subtypes were comparatively lesser than with those which have subtypes. Three types with 3 subtypes accounted for 5 % of the isolates. It is also important to note that the *emm* types with more than one sub types were among the most predominant *emm* types except the new types that were developed in this community and expanding clonally, but in less numbers. This also means that the *emm* types with subtypes got the opportunity to expand clonally by antigenic variation. The less number of isolates in the new types also means that they are recently developed and expanding vertically by antigenic variation. All the different subtypes were isolated from different students, meaning that they might have introduced in the community separately from other places and it may be by chance surviving in the community.

7.1.9. *emm* types of south Indian GAS isolates associated with pharyngitis (PT)

Forty seven *emm* types were identified among 116 PT isolates subjected to *emm* typing in the present study. In other words, 116 episodes of GAS pharyngitis were associated with 47 different genotypes showing high diversity among them. In a large

study conducted in the USA, only 35 *emm* types were identified among 1975 GAS isolates [15]. A much higher degree of diversity was seen among Ethiopian isolates with 33 different *emm* types among only 57 isolates [16] . This is comparable with data from north India which shows 33 different *emm* types from 59 isolates associated with GAS pharyngitis [39].

The most predominant type 28, accounted for 7.8 % while the most prevalent 11 types accounted for 44.8 % of the isolates. In USA, the most predominant type, *emm* 1 represented 19 % of the 1975 GAS isolates while ten most predominant *emm* types' accounted for 86-90 % of the isolates [15]. Other reports showed even higher percentages [256, 279]. The most predominant 13 *emm* types in Ethiopia isolated from pharyngitis accounted for 65 % of the 57 isolates [16]. In North India 15 *emm* types together accounted for 70 % of the pharyngitis isolates; in comparison, 19 types accounted for 65 % of the 116 pharyngitis isolates in the present study. These data once again confirm the high degree of heterogeneity among pharyngitis associated GAS isolates.

The *emm* types that were predominantly found in US and other developed countries were usually the same but may vary in the order of its predominance over time and place. The most predominant *emm* types 1, 3, 5, 6, 14, 18, 19 and 24, were commonly reported from throat infections [4]. In recent reports, the *emm* type 2, 4, 11, 12, 22, 28, 44/61, 75, 77 and 89 were also shown to be common [15, 255]. In Ethiopia, *emm* 1 and 12 were the most predominant types [16]. Similarily, *emm* 28 and *emm* 1 were the most predominant types among pharyngitis isolates [49, 280, 281] in the present study.

Among the 47 *emm* types obtained from pharyngitis isolates, 37 of them were validated types, 8 were provisional and only one was a new type. Among the 37 validated *emm* types, only 15 types were found in 35 *emm* types identified in US. Twenty *emm* types from US and 31 *emm* types of this study were unique to its own places. No provisional types were identified from US, but there were 8 provisional types which by itself show the existence of very different *emm* types in this community. But 3 new types were identified from US among the very large number of isolates studied, while only one new type was found in the pharyngitis isolates of this study. In both the places the development of new types are rare, but it indicate that the new types are still evolving, because this represents a sample of what is happening in the nature at large. In Ethiopia 4 new types were identified among 57 isolates from pharyngitis cases [16]. The chance to get more new types will become less if the same place will be studied again. Because, when the studies are conducted in a place for the first time, certain newly formed *emm* types prevailing in that area will be more and as the studies are repeated, the chance of getting new types decreases. More over not all the new type in a community need to be pathogenic or virulent enough to cause infection because even if new types are formed in a community only with those with positive mutations will survive and the other will be get extinguished due to negative selection [282].

There were only 4 *emm* types with 2 subtypes and 42 of the 47 *emm* types were found to be as single subtypes excluding one nontypable isolate. In Ethiopia only one *emm* type from pharyngitis has two subtypes, whereas in US as there was high predominance certain *emm* types, the subtypes identified in each *emm* types were high. As it was described earlier the *emm* 12 had 32 subtypes, *emm* 89 had 14 variants, *and emm* 5 exhibited 5 subtypes in the first year and much more in the

second year. Like wise *emm* 6 exhibited 8 subtypes in the first year and 9 subtypes in year 2. This data shows that the clonal expansion takes place by creating new subtypes or types as the particular *emm* type establishes itself in a community.

Monthly distribution of *emm* types of PT isolates during the two years of study showed that 27 of the 47 types were found only in the first 7 surveys while another 11 types were found only in the latter 8 surveys. Remaining 9 types were found in both the halves, but only during 2nd - 8th surveys. The rapid and complete replacement of *emm* types with other types in this population seemed to be the rule rather than exception. This may be due to elimination of particular serotypes following effective penicillin prophylaxis. It may also be due to replacement of types through type specific natural immunity. A third possible reason is the introduction of new types into the school population by children who live in different villages, but attend the same school.

7.1.10. *emm* types of south Indian GAS isolates associated with impetigo (SK)

GAS infections of the skin had in the past received relatively little attention due to the less severe and manageable complications. The resurgence of invasive GAS infections in the developed countries raised the importance of GAS causing a variety of skin and soft tissue infections, some of which are severe and even life-threatening [71]. Some of the studies indicate that bacteremia due to GAS is associated more frequently with Skin infections than with respiratory infections [69]. The term skin infections are used for all kind of pyoderma including impetigo. The epidemiology of skin infection also varies in different places and it appears more often in warmer or tropical climates and less common in industrialized western countries [25, 69].

Certain report from few countries also suspects the involvement of skin infection in the pathogenesis of ARF [34].

The data related to skin infections are relatively less in the recent past. In Northern Territory of Australia, where impetigo is far more prevalent than pharyngeal infection there are no discrete seasonal peaks in incidence of disease [30]. GAS infections in these aboriginal communities are up to 70 % in children having pyoderma [128]. Very few studies are available related to the *emm* types associated with GAS impetigo.

Like GAS pharyngitis, *emm* types associated with GAS impetigo also shows a high degree of heterogeneity. In Ethiopia where GAS impetigo is rampant, 32 *emm* types were identified among 47 impetigo associated GAS isolates [16]. Similarly, in Nepal, 27 *emm* types were associated with 53 isolates [22]. In the present study, 62 types were identified among 276 SK isolates. This shows that the *emm* types associated with impetigo in southern India is as diverse as those seen in GAS pharyngitis.

The *emm* type 122, the most predominant type identified in the present study accounted for only 5.2 % of the 276 isolates. Four most predominant types, 122, 81, 63 and 44/61 together accounted for 21 % of all SK isolates. To the best of our knowledge, our data on 276 GAS isolates is one of the few that describes the molecular epidemiology of GAS impetigo. The epidemiological data of more than a decade old studies states that the *emm* types associated with skin were 2, 49, 57, 59, 60 and 61 [4]. But in our study, *emm* 122, 81, 63, 44/61, 82, st854, 85, 53, 49, 105 and 109, were the predominant *emm* types isolated from impetigo. Among these, types 49 and 61 were found to be associated with impetigo somewhere else as well [4]. This data shows that the *emm* types prevalent in the study community were different from other places except very few *emm* types.

There were 12 *emm* types with 2 subtypes and 50 of the 62 *emm* types were found to be as single subtypes. The 12 *emm* types with 2 subtypes accounted for 30.4% of the 276 isolates. In Ethiopia, none of the *emm* types had subtypes out of 42 *emm* types consisting of 62 isolates from skin. In Nepal 27 different *emm* types were identified among 53 isolates from skin and none of the *emm* type had more than one subtype. The isolates with more than one subtype are more in skin than in pharyngitis or normal throat isolates (Table 11). This shows that GAS isolates causing skin infections are more common and are highly variable than those that infects throat.

There is a general opinion that the skin strains generally are with higher numbers above M50. In this study only 16 of the 62 *emm* types were below *emm* 50 and 37 of the 53 validated *emm* types were above *emm* 50. In this study, 7 sequence (st) types and 2 new types were identified and all the isolates were typable. The two new types were also isolated from throat along with all other new types, meaning that the new types would have developed in the mucosal environment of throat rather than in the skin.

The difference in the number of *emm* types studied is not comparable because none of the other studies were done with sufficient number of isolates. Comparing the available data on the sequence types and subtypes from skin isolates [16, 22] with our study shows that the highest number of *emm* types with more than one subtype is present in certain *emm* types of this study. More over 2 of the 3 new subtypes were identified from skin isolates is present only in skin suggesting that the clonal expansion takes place even in skin infections as the particular *emm* type establishes itself in a community.

The pattern of infection in skin seems to be much more complicated than it was known before the establishment of molecular techniques. There is no single pattern observed in all the students. The pattern of infection varies from student to student. Once the impetigo is formed, it takes minimum one week to two weeks to be controlled by the immune system by producing antibodies against the M protein [154]. Among the 140 students who were followed up for the whole two years of the study period, 88 students had skin infections. Among the 88 students who had skin infection with GAS, 38 students had infections with only one *emm* type. Another 38 students had infection with 2 or 3 types, while 9 of the students had infection with 4 types. Only 2 students had infection with 5 types and one student had infection with 8 types. This shows that the repeated infection with GAS is common in this community. Such a complicated multiple infection of GAS in skin has not been shown so clearly. One of the report from Australia has shown multiple infections and multiple strains in the same candidates [283]. The immunity formed against one type may not be effective against repeated infection because few *emm* types were found to infect the same candidate again in the skin infections. The relative risk of impetigo is 1.7 with respect to candidates, and 2.4 with respect to multiplicity of infection when compared to pharyngitis. But the GAS infection with more than one type in the same student also indicates that the immunity formed is type specific.

Unique patterns of transmission of *emm* types from SK isolates were seen in cases of GAS impetigo. One child had 8 episodes of GAS impetigo with 8 *emm* types over a period of 13 months while another child had multiple episodes with four different *emm* types (Fig. 11).

Monthly distribution of *emm* types of SK isolates in 2 years for 15 consecutive surveys showed that 22 of the 62 types were found only in one survey and 26 of the 62 types were found only in the first 7 surveys while 12 other types were found only in the latter 8 surveys. Remaining 24 types were found in both the halves yet none of the type was found in all the surveys. The rapid and complete replacement of *emm* types with another type is the rule rather than exception while few types are maintained for long in this community.

7.1.11. *emm* types of south Indian GAS isolates colonizing normal throat (NT)

The definition for throat carriers is those who harbor the organism on throat without any symptom during the time of sampling. It is important to note that approximately half of group A streptococci isolated from upper respiratory tracts who has symptomatic pharyngitis are carriers with non-streptococcal infection [23].

There were very limited studies for comparison of GAS isolates from asymptomatic carriers with symptomatic isolates. These kind of comparative data are necessary to accurately evaluate the general pathogenic ability of strains to cause disease. The study of the asymptomatic isolates acts as control to evaluate whether the types present in pharyngitis or skin is really different or the same in relation to its pathogenic ability. Carriage data also helps to compare the genotypes to know the epidemiological trend and pattern of GAS infection in the community.

In the present study 67 *emm* types were identified among 306 asymptomatic GAS isolates. In the Ethiopian study, 53 types were identified from 90 isolates from throat carriage showing much higher diversity among the Ethiopian isolates [16]. Type 28 (4.9 %) was the most predominant type followed by *emm* 118, 49, and 82 which

together accounted for 16.7 % of the 306 isolates. This once again confirms the *emm* types from asymptomatic carriage are also heterogeneous with relative lack of any predominant types associated with normal colonization.

The pattern of GAS colonization was also highly varied among normal children. Of the 140 school children who were regularly swabbed every month for 15 consecutive months, 101 (72.1 %) children had GAS colonization of the throat at least once. Ninety three children (66.4 %) were colonized with one or two types while 8 (5.7 %) were colonized with three or four types. Varied patterns of colonization were also seen among children during the survey period. One child was colonized in the throat with three different *emm* types wherein one type, st854.2 persisted on six occasions during a period of 14 months. Another child was colonized with four different types during a period of 8 months while in another child colonization was seen with only one *emm* type during the entire two year period. Therefore the risk of GAS carriage is very complicated and varies from person to person.

All the new types identified in the study were found in normal throat which supports the hypothesis about the development of new *emm* types in the mucosal layer of the pharynx especially during the carrier state (Table 13).

Month wise distribution of 67 *emm* types for 15 consecutive months during 2002 – 2004 has shown that 27 types were found only in the first 7 surveys while 14 other types were found in the latter 8 surveys; 26 types were found in both halves of the study. No type was found in all the surveys. Type 82 was the most widely distributed type and was found in 8 surveys. Type 85 was present in the first 2 surveys and then reemerged in the last survey. Few types appeared and reemerged with intervals during many surveys. The change in the *emm* types has been found in pharyngitis and even in

skin infection over the period of time. The three possible reasons for the change in the *emm* types may be i) The eradication of one serotype may be replaced due to the influence of effective treatment, ii) in an natural condition if the *emm* type shift takes place, it may be due serotype specific natural immunity or even against heterologous types. iii) it may be due to environmental and other confounders that attributes to season or the chance of transmission due to break in get together like vacations etc. The rapid change in the *emm* types is the reason for the low prevalence of *emm* types in this community but the reason for the change can be assumed partly as the role of immunity and transmission but can not be pointed out unless specially studied.

The students enrolled in the study were screened for asymptomatic pharyngeal carriage of GAS on a monthly basis for 15 times in the whole of the study period. 140 students were screened for the full study period. One of the student (VAP 51) had infection with *emm* type 28 which caused symptomatic infection two times with long period of carriage asymptomatically (Fig. 10). This may be due to the host related factors which have not been investigated because that was not in the plan of the study. One of the student (VAP 110) had six weeks of GAS pharyngitis with 4 different types with intermittent asymptomatic colonization of GAS with different types (Fig. 10). This result very clearly demonstrates both the type specific immunity and the immunity across *emm* types. On the other hand these examples indicate the multiplicity of GAS colonization in throat and the risk of GAS infection in this community.

7.1.12. *emm* types of skin isolates versus throat isolates

When the same species can infect two site of human body, it is important to know whether the types causing the infections are of the same genotype or not.

Conventional thinking is that certain types are specific for throat infection and certain other types specific for skin infection. In the present study out of 69 *emm* types identified among throat and skin isolates, 40 types were found to be common for both. Since skin infections are more prevalent in this community, generally the types found in skin infections can be expected to dominate in the community. However, 24 of the 40 types were more frequently associated with pharyngitis than skin infection; only 16 types were more frequently seen among skin isolates. This shows that certain types although found in both the sites, have a stronger predilection towards pharyngitis than impetigo while certain other types have a skin predilection.

Similar findings have also been reported for north Indian GAS isolates although the sample size in that study was much smaller [33]. Thus it is very clear that the epidemiology of GAS infections in areas endemic for these infections is significantly different and more complex than non-endemic regions. Our results show the predominance of *emm* 28 in pharyngitis and *emm 1*22 in impetigo. Probably, there is some specific association between genotypes and sites of infection. However, due to the high degree of heterogeneity among the genotypes and endemicity of GAS infections, it will be very difficult to distinguish between throat specific and skin specific GAS strains in our population. This is further complicated by the distinctly different climatic conditions in tropical countries which influences the spread of GAS strains in the population. Rural school children living in neighboring villages attend the same school which facilitates the transmission of different *emm* types that are unique to each of the villages. We have observed in another study, the distribution of *emm* types in two neighboring villages is significantly different, although certain types tend to be present in both villages (RR: Personal Communication).

7.1.13. *emm* types of pharyngitis isolates versus colonizing isolates

In regions endemic to GAS infection, laboratory confirmation of colonizing GAS isolates and their differentiation from pharyngitis associated isolates is a difficult task. In this study we defined colonizing isolates as those isolated from children with no signs and symptoms of sore throat at the time of the visit. Since no strict criteria were followed for this definition, it is possible that some of the GAS isolates were those which caused infection in the recent past [76]. Our study showed most of the *emm* types identified among PT isolates could also be identified among normal colonizers. We conclude that since throat colonizing types reflected both throat and skin infecting types to a large extent, such types could be studied to characterize the types that circulate in a community.

7.1.14. *emm* types of pharyngitis, impetigo and colonizing isolates

The fact that 37 of the 77 *emm* types identified among 698 GAS isolates were seen among three categories shows that much overlap exists among types associated with throat and skin infection. Type 63, the most predominant type identified in the present study, was among the predominance types among all three categories of isolates. Thus it is very difficult to define the association of this type with any infection; probably it can cause both throat and skin infections. Type 49, which is considered to be an impetigo associated type, was also identified among pharyngitis associated isolates. In general, both the pharyngitis associated and impetigo associated *emm* types were identified among throat colonizing types; thus identifying the *emm* types associated with normal colonization of the throat can give information on the types that generally cause throat and skin infection in a community.

7.1.15. Transmission of GAS strains from skin to throat

It is known that GAS are usually transmitted from person to person through respiratory droplets or by close contact [30]. Even though this study was not aimed at finding out the transmission patterns of GAS, we identified 30 candidates having the same *emm* type associated with GAS pharyngitis and impetigo. In this cohort study, the data obtained shows that no child had throat infection or colonization in the throat prior to the development of skin lesions. But, skin infection leading to throat colonization with the same type was found in 14 (46.7 %) children, skin infection leading to pharyngitis in 12 (40 %), skin infection and pharyngitis at the same time in 2 cases (6.7 %), and skin infection and throat colonization at the same time in 2 cases (6.7 %). This shows that skin infection plays a major role in the transmission of GAS infection in this community where GAS impetigo is far more prevalent than GAS pharyngitis. Similar findings have been reported earlier [41]. The transmissibility of GAS related to respiratory tract infections in Air force Base in Wyoming demonstrated the importance of crowding. The distance between beds or children in case of classrooms is important in the transmission and therefore the incidence of GAS pharyngitis. We have not seen any child with GAS pharyngitis with one *emm* type followed by infection of the skin by the same type. Therefore there are reasons to believe that skin acts as a source of throat infection rather than a throat infection acting as a source for skin infection. Till now there was not even a single report about the transmission of GAS from throat to Skin, but there were few reports about the skin infections leading to throat infections with the same type. An example given in a review shows that, M type 49 reappeared in the Red Lake population, where the epidemic strain first became predominant in skin lesions and subsequently in the upper respiratory tract. Therefore, it seems that the primary source of skin infection is

most probably another skin lesion or the respiratory tract. This concept was documented in few studies [41]. In Australia, the major transmission of GAS in the community has been attributed to skin sores [25]. Data from the present study also indicate that in certain individuals self transmission also occurs from skin to throat.

7.1.16. Monthly distribution of GAS *emm* types in a south Indian community

The dynamic transmission of diverse *emm* types circulating in a community was determined by plotting their monthly distribution. Since normal colonization pattern was studied on a monthly basis, for the sake of comparison, it was decided to plot the distribution of PT and SK associated *emm* types also on a monthly basis. The overall picture among the three categories was very striking. Among all three categories, certain types predominated during the first eight months of survey. These types were replaced by others during the remaining 7 months of the study period. Indeed, there were some types that overlapped during the two time periods. This pattern was similar among *emm* types of pharyngitis associated, impetigo associated and colonizing GAS isolates.

One reason for such a distinct pattern could be the effect of penicillin prophylaxis. In this study, all children with bacteriologically proven GAS pharyngitis and impetigo were given Benzathine penicillin within 72 hours of diagnosis. The effect of the prophylactic treatment is also shown by the reduction in the number of GAS isolates during the latter surveys. The second reason could be the effect of type specific immunity prevailing in this population. Since both these factors were either not analyzed or studied, it is not possible for one to confirm the reason(s) for this phenomenon.

7.1.17. *emm* types of south Indian GAS isolates and vaccine:

The epidemiology identified in this region is different than what is reported from other countries and the vaccine formulated based on the surveillance report from US may not need to be reformulated if it is planned for the global usage. From the results of this study, only 15.6 % and 23.3 % of isolates identified from total and pharyngitis will be covered (Table 31), if the same vaccine will be used which was anticipated to protect 90 % infections in US [240]. If a vaccine to be formulated based on this study, vaccine carrying at least 37 different *emm* types would be required to cover approximately 80 % of the isolates represented by the *emm* types found commonly from all the disease category. Even though such a vaccine will be useful to minimize the GAS burden, the types that were not covered by the vaccine would dominate on the one hand and the immunological pressure may cause the existing type to mutate and allow the vaccine escape mutants to dominate on the other hand. Such a dynamic change taken place over the period of time was already shown above, which might have taken place due to the effect of treatment. Hence the fluctuations in the *emm* type distributions would need to be monitored continuously. Yet this data is from a single study is one particular region, but a national wide survey is required for effective prevention.

Table 31: Evaluation of the 26 valent vaccine with the *emm* types of this study

***Emm* types**	**Total**	**PH**	**NT**	**SK**
1	0	0	0	0
1-2	14	6	7	1
2	0	0	0	0
3	10	3	2	5
5	0	0	0	0
6	0	0	0	0
11	3	0	2	1
12	2	0	2	0
12	0	0	0	0
14	0	0	0	0
18	1	0	0	1
19	0	0	0	0
22	3	1	2	0
24	0	0	0	0
28	29	9	15	5
29	0	0	0	0
33	0	0	0	0
43	0	0	0	0
59	0	0	0	0
75	15	4	6	5
76	1	0	0	1
77	17	2	8	7
89	11	2	4	5
92	3	0	1	2
101	0	0	0	0
114	0	0	0	0
Total	109 (15.6 %)	27 (23.3 %)	49 (16 %)	33 (12 %)

7.2. *emm* family pattern

The results of the *emm* family pattern for 67 different *emm* types shows that the pattern E (53 %) was the most predominant pattern represented in this GAS population. Pattern E is reported to be generalists which means that types belonging to this pattern can cause both throat and skin infection [41]. Pattern E was followed by pattern D (30 %) which is a skin specialist. Least prevalence was seen among family pattern A-C (9 %) which is throat specialists. This shows that types that are strictly throat infecting are less often seen in this GAS population. This data is in agreement with other reports from tropical countries where pattern D is much more common and A-C are much less common among GAS strains, than in temperate regions [22, 30]. In Australia, 43% of the GAS isolates in the community belonged to pattern D [30]. In Nepal and Ethiopia, 28.3 % and 28.1 % of the isolates were pattern D respectively [16, 22]. In contrast, only 6.9 % were predicted to be pattern D and A-C to account for 42.5 % of the invasive isolates in US [256]. The A-C pattern strains in Ethiopia, Australia and Nepal were 17.9 %, 16 % and 10.8 % respectively. Comparatively, Australia showed the highest prevalence of pattern D isolates [30]; in Nepal pattern E constituted 60.8 % of their isolates [22]. This confirms the general observation that in tropical countries, where GAS impetigo is generally much higher that pharyngitis, the generalists, namely pattern E predominate among the GAS population. Indeed certain Australian studies have shown a complete absence of even a single case of GAS pharyngitis among the indigenous people in the Northern territory where incidence of GAS impetigo as well as acute rheumatic fever/rheumatic heart disease is among the highest in the world [284]. The high prevalence of post-streptococcal sequelae in tropical countries and increasing incidence of invasive GAS disease in the temperate countries is probably associated with the contrasting differences in the family

patterns. This is also supported by the fact that *emm* types seen in tropical countries are vastly different from those identified in the temperate climates.

Among 7 of the 74 *emm* types studied (*emm* 65, 87, 100, 122, stKNB4, stKNB5 and one nontypable strain), the *emm* family pattern could not be determined because the PCR product did not show any pattern corresponding to any of the three pattern. The problem in the designation of 4 of the 82 isolates was also reported in the article where this scheme of *emm* family pattern was published [30]. This may be due to the change of gene fragments in the *vir* locus of these *emm* types as the recombination is very common in this locus [56, 285] or else, it may even be due to the alteration in the primer binding region.

The first report on the specificity of GAS *emm* types for tissue tropism, it was hypothesized that family pattern A-C was disproportionately associated with nasopharynx, whereas family pattern D was most often isolated from impetigo lesions and pattern E with both sites. Such a hypothesis was based on a study from temperate region which showed 95 % of skin isolates and 75 % of throat isolates displayed pattern D and A-C respectively [41]. The reports from the tropical countries were contrary to what was reported [22, 30]. The results of the present study shows that, 12 (30 %) of the 40 types associated with pharyngitis belonged to pattern D which is reportedly skin specific (Table 20). Likewise, five of the six A-C patterns *emm* types were also found to be isolated from skin infection (data not shown). Our data therefore, also shows that *emm* family pattern may not be a satisfactory genotypic marker for skin or throat specificity. Similar reports were also published from Nepal, where 10 of the 13 isolates of pattern A-C were recovered from skin sites and the pattern D were almost nearly distributed evenly between skin and throat isolates [22].

In Australia, 19 % of the pattern D strains were isolated from throat and 13 % of the GAS recovered from Impetigo were pattern A-C [30]. In Ethiopia, the A-C pattern was strongly associated with throat whereas the frequency of pattern D strains did not differ among skin and throat isolates [16]. Therefore the association between the *emm* family pattern and the tissue specificity of GAS isolates remains controversial. The reason for such a distinct difference in family pattern distribution may be related to large number of *emm* types and their heterogeneity seen among the south Indian GAS population.

7.2.1. Sof

It was known that approximately 40-60 % of GAS isolates were capable to opacify sera due to the expression of *sof* (serum opacity factor) gene [42]. M types of GAS have been divided into OF positive and OF negative groups, and this division has been reported to parallel differences in the architecture of *emm* like genes in the vir regulon, which is termed as *emm* family pattern. In our study, 59.45 % of the 74 *emm* types tested were *sof* gene positive. We have also confirmed the presence *sof* gene in 38 of the 39 *emm* types tested from *emm* family pattern E except *emm* type 15, which was also shown previously in other reports [111, 206]. Only 3 of the 28 *emm* types from pattern D and *emm* type 12 belonging to pattern A-C had this gene, which was also consistent with reported literature [42, 206]. Taken together, these observations have led to the suggestion that OF+ and OF- M types represent two distinct evolutionary lineages of group A streptococci.

Application of the OF results of the individual *emm* types to all the isolates of the corresponding types, we have identified that 58.5 % of the 665 isolates for which, the

sof from this study could be correlated, had the gene. This report is in concordance with previous reports.

Fibronectin activity of serum opacity factor [204] now raises the issue that this might be the factor responsible for the binding of the *emm* family pattern E strains to any of the tissue site. However, further specific studies are required to find out the implications of this factor in binding with different tissues.

7.3. Phylogenetics

Development of a phylogenetic tree helps to determine the evolutionary lineage of diverse genotypes in any bacterial species. This is especially useful for an organism like GAS which shows much diversity in countries like India. Phylogenetic tree helps one to understand how this heterogeneity is formed and what relation exists between genotypes that circulate and cause infection in a population. Against this background, a phylogenetic tree was developed based on the whole *emm* gene sequences of 76 of the 77 *emm* types identified in this study (Fig. 17).

A comprehensive analysis of the tree shows that 53 of the 77 *emm* types are clustered together into one of the two main clusters (upper cluster). These *emm* types predominantly belong to family pattern E which has been described as a generalist, meaning they can cause both throat and skin infection. This confirms our results of *emm* typing which showed that 37 of the 77 types identified could be identified among all three groups of GAS isolates. Obviously, high degree of intragenic recombination occurs in these *emm* types resulting in the development of new genotypes and consequent high diversity among these types.

The second or the lower was smaller and included 23 *emm* types divided into 5 sub-clusters labeled 12 to 16. All types of sub clusters 12 to 14 belonged to sub family pattern D except *emm* 57 which belonged to the pattern A-C. It is important to note that *emm* 57 did not cause pharyngitis in any of the candidate and is more associated with skin infection or throat colonization. Therefore it seems that *emm* 57, though it belongs to subfamily pattern A-C, might overlap in its specificity for skin. This result indicates that the sub family pattern based on *emm* genotypes is not an exact marker for tissue preference.

It is very interesting to note that 38 of the 39 *emm* types that belonged to family pattern E were *sof* gene positive (*sof* $^{+}$). Since opacity producing GAS isolates belong to Class II M proteins most of the *emm* types (n = 38) that belong to E may also be non rheumatogenic [4, 177]. Therefore by this logic, it may be inferred that most of the GAS genotypes circulating in this community may not have the ability to initiate a rheumatogenic process in children of this community.

One important finding from this phylogenic data is that the pattern D is intermediate to patterns E and A-C. The *emm* types of clusters 9, 10, 12, 13 and 14 all belong to pattern D and are predominantly *sof* $^{-ve}$. Such observation was also confirmed by the distance matrix plotted between the clusters (Fig 18) which also shows that one of the patterns, either D or A-C has evolved from one of the other. Though the exact direction is not known, according to the application of additive acquisition of gene theory, the pattern D could have evolved from patterns A- C and pattern E would have evolved from pattern D. Supporting this hypothesis is the order of family pattern D in the phylogenetic tree which is aligned inbetween A-C and E patterns.

It is noteworthy that the six newly identified *emm* types, namely stKNB1 through stKNB6 are situated at different sites and clusters of the phylogenetic tree. Types stKNB1, 2 and 3 are situated in cluster 1 at the top of the end of the tree (family pattern E), stKNB4 in cluster 11 (pattern not known), stKNB5 in cluster 9 (pattern D) and stKNB6 in cluster 5 (pattern E). Thus geneologically these types are distinct. Moreover, types stKNB1, 2 and 3 belong to pattern E and are *sof* $^{+}$. On the other hand, stKNB4 was *sof* $^{-}$ while its pattern was not known. Type stKNB5 belonged to cluster 9 and pattern D was *sof* $^{+}$. Thus four of the six newly identified types were defined as generalists, while stKNB5 alone was skin specialist. This confirms our findings that most of the genotypes of south Indian GAS isolates are oriented towards skin tropism, a finding supported by a high incidence of GAS impetigo in this community.

If the evolution of one type from the other can be dipicted in a sequencial order then the theory of additive aqusition of genes over time works real when it is compared with the evolution based on the vir regulon (the *emm* sub family pattern). The evolution can not always be represented in a regular tree form because the evolution of any organism does not follow a particular order; rather it takes place in a random fashion according to natural selection. The exact path of evolution is unpredictable as there is no method or program to depict the correct representation of evolution [130]. The correct representation requires all known *emm* types including extinct ones, because incomplete lineage sorting can cause serious difficulties for phylogenetic inference [286]. But studying all existing *emm* types itself gives valuable information regarding population genetics of the GAS as shown in the present study.

Analysis of distance between 16 genetic clusters identified in the phylogenetic tree was done using the MEGA software (Fig. 18). The matrix values depicted the distance between the clusters thus describing the similarity and diversity among the various *emm* types circulating in our population. Thus the distance between cluster 1 and 2 was 0.04 showing a 4 % mismatch and 96 % similarity among the *emm* types included in these two clusters. Similarly the distance between cluster 1 and 16 was 0.31 showing a difference of 31 % and similarity of 67 %.

The matrix which depicts the range of diversity also confirms the evolution in the context of their family patterns, namely E, D and A – C (Fig. 18). The *emm* types from the pattern E were closely related (the distance between clusters ranges from 4% to11 % variation) and establish rapidly in the community, accounting for 47.8% of the total GAS isolates. While the pattern D and A-C shows 15 - 35 % and 27 % variations respectively among the clusters, which means that they were evolved earlier than family pattern E. According to the order of distance between the clusters, pattern E is closer to pattern D than A-C, while pattern D is equidistant from both E and A-C patterns. This means that pattern D evolved from pattern A-C while pattern E have developed later from pattern D. This observation explains the ancestry of A-C pattern and branched into a separate lineage to form pattern E through pattern D.

Approximately the first 300 bp of the *emm* gene sequence of very closely related new types obtained by using the primer 1 and 2 were analyzed for the changes leading to antigenic variations (Fig. 20). They were clonally related and clustered together as shown by the phylogenetic analysis. They were evolved by varied ranges of simple mutations to additions or deletions of set of nucleotides, but the frame shift was not identified in any of the isolate.

The new *emm* types were all found in the normal throat while only two were associated with impetigo which is consistent with the idea that the antigenic variation occurs due to the immunologic pressure exerted in the mucosa of the pharynx [209, 275]. The clustering of new *emm* types together shows that they were developed from a common ancestor and expand clonally into different *emm* types.

7.4. Distribution of PAVF among south Indian GAS isolates

Whole genome sequencing of epidemiologically important GAS isolates and their genetic studies have shown high plasticity of this pathogen [140]. It is well known that phages have played a major role in the genomic variations and the development of new pathogenic clones by lysogenic conversion [47]. The scope of this part of study is to look into the involvement of PAVF in the GAS isolates of uncomplicated infections and throat colonization among south Indian rural school children as well as invasive isolates obtained from the patients from a tertiary care hospital.

The presence of virulence factors in a GAS isolate has direct implication in its virulence nature, as manifested by the extent of disease severity. The presence of PAVF in 78.3 % of 218 GAS isolates is higher than what can be expected from any geographical area. Analysis also shows the inverse relation between the number of isolates and the number of virulence factors (Table 26, 31). Number of isolates with many virulence factors was very few while isolates with no PAVF was 21.7 %. In other words, 95 (43.6 %) of the 218 isolates had only one or no PAVF at all, while 149 (68.3 %) had two or less number of PAVF in them. This indicates that there are many strains circulating in the community with minimal virulence genes. Forty five (20.8 %) had four or more PAVF in them. Only one isolate had seven PAVF in it while there was no isolate with eight or more PAVF. The existence of very few strains

with large number of virulence factors may mean that most of them do not cause severe GAS diseases in India [287].

Presence of individual PAVF in these isolates varied from 0.5 % (*sda*) to 32.1 % (*speH*) which was the highest. Genes, *mf4* and *sdn* were not identified in any of the 218 isolates. (Fig. 21). Highest prevalence was found with *speH* (32.1 %), followed by *ssa* (21.6 %) and speA (21.1 %). Factors *mf4* and *sdn* were not found in any of the isolates showing that some phages or some bacterial clones were not yet transmitted to this population.

A comparison of PAVF profiles among our strains with already sequenced strains of the same *emm* types showed distinct difference in the PAVF content (Table 32). All sequenced strains associated with severe invasive diseases harbor 5-7 virulence factors [47] while the same *emm* type identified in this study had only one or two PAVF. One of the isolates of *emm* 18.12 from skin had 2 PAVF while the same type isolated from blood from a bacteremia patient had only one PAVF (Table 32). This may indicate that many of the PAVF responsible for severe infections were not yet identified among our strains. The most predominant type in the present study, *emm* 28 had no PAVF at all while only one PAVF was identified in *emm* 1-2.2. This pattern of PAVF distribution was seen among isolates from uncomplicated infections, colonizing strains as well as those causing invasive GAS disease. This irrespective site of lesions or nature of disease from south Indian GAS isolates may have significantly different phage mediated virulence factors among them.

Table 32: Comparison of PAVF in sequenced strain and strains of the present study

GAS strain	PAVF					No. of PAVF
M1 (SF370)	speC, mf2	speI, speH	mf3			5
Emm 1-2.2				speA		1
Emm 1.4	spec, mf2					2
M3 (MGAS315)	ssa	mf4	speK, sla	speA3	sdn	6
Emm 3.22				speA		1
M18 (MGAS8232)	spec, mf2	speL, speM	spd3	speA1	sda	7
Emm 18.12 (skin)		speH		speA		2
Emm 18.12 (Blood)				speA		1

Note: The red colour refers to the reported whole genome sequence of the respective GAS strain
The black colour refers to the data generated in the present study

In general, lesser the number of virulence factors, lesser is the disease causing ability and rate of transmission. This has been previously reported in serotypes M3 which has more virulence factors and hence found more often than M28 [49]. If that is true, it correlates well with our findings. The strains with higher number of virulence factors and enhanced disease causing ability are not frequently encountered in India. Already sequenced strains of predominantly encountered *emm* types in western countries show a PAVF pattern that is different from the same *emm* type found in the present study (Table 32). Moreover, virulent *emm* types identified in other countries are not commonly seen in India. Types like *emm* 1, 3, 6, 12, and 18 are very rare and even if are present they possess less number of virulence factors.

The GAS isolates included in the study represented 44 different profiles based upon different combination of PAVF (Table. 24). Similar observations have been reported

in other studies with isolates from geographically wide spread area [49, 170]. The number of profiles identified in this community is relatively less considering that number of profiles that can be made with the combination of 14 PAVF with each other is far greater than what is observed in this study. The relatively less number of profiles is indeed reflected by the low prevalence of PAVF in this GAS population.

The present study did not show any direct correlation between PAVF profiles and their sites of recovery (Table 25). Among the 44 profiles identified, 10 were shared by all four groups while 5, 7, 3 and 5 profiles were found to be unique for PT, NT, SK and IV isolates respectively. Although few differences were identified, there was no real association of any of these profiles with their sites of isolation.

There were no profiles unique to PT and IV isolates; but (Table 25) there were 3 profiles unique to SK and IV isolates suggesting that more often, skin infections may be associated with invasive diseases. This is of special significance to this population where the incidence of GAS impetigo is much higher than throat infection.

The generalized distribution pattern of PAVF and their relatively low prevalence among all groups of isolates indicate that PAVF may not play an important role in disease patterns in this population. Phages do play an important role in transferring a variety of virulence factors among this GAS population, however lack of any PAVF in 30.6 % of the invasive isolates (Table 25) show that these factors may not play an important role in the differential pathogenesis of GAS infections in this population. Indeed, the present study did not look at the expression of these genes; therefore it is difficult to prove the precise role of these virulence genes in causing invasive GAS disease.

7.4.1. Strain specific Virulence gene profile

Additive acquisition of new prophage-encoded virulence genes over time by already successful GAS strain might create an unusually virulent sub-clone with novel virulence capacity or profile. This is shown very well in *emm* 49 in the present study by the movement of phages and the strain specific virulence genes (Table 33). All *emm* 49 isolates studied are different in their virulence factor content and the site of isolation. Whether they have any association to specific sites is not known. However, this shows that PAVF can be used to study the development of new strains with new virulence genes within the same *emm* type. Such studies will give lot of insights in the distribution of clones, especially in case of epidemic investigation.

Table 33: Strain specific PAVF profiles in *emm* 49.

Emm type	No of isolates	Site of isolation	Profile number	Virulence gene
49.4	1	IV	1	none
	2	Skin, Normal Throat	12	SpeL, SpeM
49	1	Skin	21	SpeL, SpeM, speH
	1	Pharyngitis throat	32	SpeL, SpeM, speH, SpeI
	1	Normal Throat	38	SpeL, SpeM, speH, SpeI, mf2

IV = Invasive isolate
SK = Impetigo isolate
NT = Colonizing throat isolate
PH = Pharyngeal throat isolate

7.5. Distribution of chromosomally coded virulence factors (CCVF) among south Indian GAS isolates

The CCVF should generally be present in all GAS isolates because 10 % differences found in the genomic sequence of GAS isolates is associated with mostly the prophage region [47]. The present study showed that eight of the 21 genes studied were present in 98.6 % to 100 % of isolates (Fig. 24) of which six were toxins. Prevalence of the 12 fibronectin binding proteins varied from, 2.1 % (*fbp54*) to 91.4% (*fba*) and 97.1 % (*sciB*); gene *prtf2* was not found among any of the isolates.

The *sic* gene was identified in 15 (10.7 %) isolates. This included *emm* types 1 (n =4), 3 (n =2), 12 (n =1), 31 (n =1), 55 (n =2), 57(n =2), 39 (n =2) all of which belonged to family pattern, A-C and type 28 (n = 1) which belonged to pattern E. Thus the distribution of this gene is predominantly restricted to A-C family pattern, which is a throat specialist; therefore, probably it might be a primary or secondary determinant for the throat specificity. The *sic* gene is originally identified from *emm* 1 which was related with epidemics [209]. The *sic* gene identified from *emm* 12 and 55 were already reported as distantly related streptococcal inhibitor of complement from North India and Australia [211, 288]. In Japan, 72.7 % of the isolates of GAS harbored this gene [289]; however, when all the isolates of the *emm* types that contained this gene was calculated, it was only 13.2 % of the total isolates including the *emm* 28.

The distribution of CCVF among the GAS isolates does not match any of the clusters identified in the phylogenetic analysis or to *emm* family pattern except a significant association of *fba* with all the isolates of family patterns E and D. Interestingly, all isolates which did not have this gene belonged to the A-C pattern. Thus gene *fba* appears to be a marker for skin and general specialist GAS isolates.

Among the three *prtf* genes, *prtf1* was absent in all groups of GAS isolates studied. It has been shown earlier that this gene is present only in *emm* 6 [214]. It is noteworthy that *prtf2* was predominant among SK isolates (Fig. 25). But the absence of this gene in majority of the skin isolates raises the question of its specific involvement in the tissue specificity. Rather, it's a random acquisition of the genes through horizontal gene transfer [290].

In conclusion our studies on PAVF and CCVF showed much heterogeneity in the acquisition of these genes among south Indian GAS isolates. It did not reveal any significant differences in its distribution among the isolates from various disease groups. Indeed a recent study on microarray profiling of 226 virulence genes did not show any increased propensity of any of these genes to cause invasive diseases, underscoring the pathogenic complexity and the involvement of multiple factors [291]. Detailed study on the expression of the products of these genes with respect to various host conditions will have to be studied to prove the true nature and role of these factors in the pathogenesis of GAS infections among south Indian children.

8. Summary and conclusion

8.1. Summary

1. A total of 698 GAS isolated from throat cultures of children with pharyngitis (PT, $n = 116$), skin cultures of children with impetigo (SK, $n = 276$) and throat cultures from asymptomatic children (NT, $n = 306$) were included for the study. These isolates were selected from a total 769 GAS isolates that were subjected to *emm* typing. Same *emm* types of GAS isolated from same child in consecutive weeks were considered as the same strain and were therefore not included in this analysis. (Table 1).

2. Seventy seven *emm* types were identified among the 698 GAS isolates. This results show a high heterogeneity among GAS isolates circulating in and around Vellore (Table 2). To the best of our knowledge, this study has the largest series of GAS isolates subjected to *emm* typing from one community in a prospective manner.

3. The four most common types namely, *emm* 63 (4.6 %), *emm* 81(4.3 %), 28 (4.15 %) and 49 (3.9 %) accounted for 16.9 % of all isolates (Table 15, Fig. 26). Eleven types accounted for 272 isolates (38.96 %). This shows that many types predominate among this GAS population. Further, in an endemic country like India, no single or few types can be shown to predominate.

4. In this study, sixteen sequence types (st) identified among 698 GAS isolates accounted for 95 (13.6 %) isolates. Six of them namely, stKNB1 through stKNB6 were hitherto unrecognized new types (Table 2). The number of provisional types and new types identified in this study is much higher than many other reported literatures.

5. This shows that newer *emm* types tend to evolve through genetic mutation among strains that circulate and cause infections in this endemic area. Possibly, a highly

susceptible population enhances the rate of mutation which in turn results in the development of new strains. The only non typable strain isolated in this study suggests that such strains are incidental and they may not be associated with pathogenesis.

6. Forty six *emm* types were identified among 116 PT GAS isolates shows a high heterogeneity among them. This probably results from a high rate of transmission of GAS among children who come from neighboring villages, but attend the same school. Types 28, 1 and 49 accounted for 17.2 % of the 116 isolates show predominance of certain types. Interestingly, type 1 which is a globally recognized invasive type accounted for only 2.6 % of 698 GAS isolates. Moreover, subtype 1.2-2 accounted for 14 of the 18 isolates of *emm* 1. (Fig. 5, Table 2).

7. Sixty two types were identified among 276 SK GAS isolates in the children of the same cohort again shows high heterogeneity among impetigo associated types (Fig 6). Type 122 was the most predominant (5.79 %) and seven types namely, 122, 81, 63, 44/61, 82, st854 and 85 accounted for 32.6 % of the 276 isolates. This shows that certain types do tend to predominate despite the high heterogeneity.

8. Sixty seven types were identified among 306 GAS isolates associated with normal throat colonization. The distribution pattern reflected the types that were predominant among PT and SK types. Types 81, 82 and st854 found in NT were associated with SK isolates while types 49, 63, 118, 42, st1731 and 77 were associated with both PT and SK isolates. This shows that colonization of normal throat with GAS can occur with strains associated with both pharyngitis and impetigo (Fig. 7). Moreover, the risk of GAS carriage and persistence of infecting strains varied from person to person which demonstrates the type specific immunity and immunity across *emm* types.

9. In this cohort of children, skin infection leading to throat infection or throat colonization was identified in 30 of the 140 students (Table 16). In contrast, throat infection or throat colonization prior to the development of skin lesion was not observed in any child. This shows that skin infection plays a major role in the transmission of GAS in this community.

10. Comparison of predominant *emm* types among PT, SK and NT isolates showed that three types namely 49, 63 and 44/61 were common to PT and SK isolates, while types 28, 1, 49, 63, 42, st1731 and st2147 were seen both among PT and NT isolates. Similarly, types 81, 63, 82, st854 and 49 were seen among SK and NT isolates. This again shows that impetigo associated types can colonize normal throat and may also cause pharyngitis or vice versa. (Fig. 8 & Table 3).

11. A total of 40 *emm* types common to PT and SK isolates were identified (Table 14). The tissue preference index was more for PT isolates in 24 types and in 16 *emm* types for SK isolates (Table 14). This comparison shows that some types do have predilection towards pharyngitis while others have a skin predilection.

12. Monthly distribution of *emm* types among PT isolates showed a tri-model clustering. Twenty seven types were restricted to first 8 months of survey and 11 types were restricted to the last 7 months while 9 types were seen scattered among different months of surveys. This probably is associated with the intensive treatment given to all symptomatic children as a result of which newer types appeared at later stages of survey (Fig 14). Similar dynamics of transmission of GAS strains were seen among SK and NT isolates. Since *emm* types associated with SK and NT isolates was far more heterogeneous, the survey-wise distribution of types was less pronounced (Figs. 15 & 16).

13. Seventy four GAS isolates representing 74 *emm* types were selected for determining the *emm* family pattern (Table 19). These included 40 from PT isolates, 29 from NT isolates and 5 from SK isolates. Family patterns could be established in 67 of the 74 types while it could not be determined in seven types. The latter included types 65, 87, 100, 122 as well as two new types, stKNB4 & stKNB5 and the lone nontypable isolate. Thirty nine (52.7 %) of the 74 *emm* types belonged to the pattern E while, 22 (29.7 %) of the 74 types belonged to pattern D. Thus, 67 (90.5 %) of the 74 types belonged to either of these types. This pattern distribution was similar among the three groups of isolates. Further, 12 (30 %) of the 40 types associated with pharyngitis belonged to pattern D which is reportedly skin specific and five of the six A-C patterns *emm* types were also found to be isolated from skin infection. Therefore, the *emm* family pattern may not be a satisfactory genotypic marker for skin or throat specificity.

14. In our study, 59.45 % of the 74 *emm* types tested were *sof* gene positive. Further 38 of the 39 types that belonged to family pattern E were *sof* gene positive while only 3 of the 22 types that belonged to pattern D were *sof* gene positive. Similarly only one of the six types that belonged to pattern A-C were *sof* gene positive. As it is known that all *sof* gene positive types belong to Class II M proteins which are believed to be non rheumatogenic. Thus 39 of the 74 types which were *sof* gene positive, belonged to M protein class II (Table 19).

15. A phylogenetic tree was developed based on whole *emm* gene sequences of 76 representative *emm* types (Fig. 17). This showed two distinct clusters with 53 and 23 *emm* types respectively. The patterns of clustering of the types show that there are probably two lineages among the 76 *emm* types. The types seen in the upper cluster

are more congruent and therefore seem to have evolved from a common ancestor belonging to *emm* family pattern D, developed into E with very little genetic changes. In contrast, the lower cluster consists of more divergent types belonging to sub clusters of *emm* family pattern D and A-C which probably reflect more complicated genetic changes that might have occurred during evolution (Fig. 17). The sequential order of *emm* types arranged in the phylogenetic tree and the matrix analysis have shown that the pattern D is located in between pattern E and A-C, showing that pattern D has probably evolved from pattern A-C while pattern E has developed later from pattern D.

16. Analysis of the conserved region sequences of 76 *emm* types has shown that ten types including 18.12 indicate an intermediary position to Class II and I *emm* types (Fig 19). Only these two classes are known so far, but this result shows an intermediary group which has a combination of both the existing classes. This confirms the development of types of pattern E from those of pattern D.

17. The first 300 bp of the *emm* gene sequence of very closely related new types, stKNB1, 2, 3, 4, 5 and 6 were analyzed for the changes leading to antigenic variations (Fig. 20). They were clonally related and clustered together as shown by the phylogenetic analysis.

18. The presence of PAVF in 78.3 % of 218 GAS isolates had at least one PAVF while 149 (68.3 %) had two or less number of PAVF in them which indicates that there are many strains circulating in the community with minimal virulence genes. Forty five (20.8 %) had four or more PAVF in them. Only one isolate had seven PAVF in it while there was no isolate with eight or more PAVF. The existence of very few strains with large number of virulence genes mean that most of isolates are not capable of

expressing virulence factors to cause complicated diseases. Presence of individual PAVF in these isolates varied from 0.5 % (*sda*) to 32.1 % (*speH*) which also supports the very low level of PAVF in these isolates (Fig. 21). Factors *mf4* and *sdn* were not found in any of the isolates showing that some phages or some bacterial clones were not yet transmitted to this population.

19. All sequenced strains associated with severe invasive diseases harbor 5-7 virulence factors while the same *emm* type identified in the cohort of children had only one or two PAVF (Table 32). Also, no difference in the PAVF distribution was seen among isolates from uncomplicated infections, colonizing strains as well as those causing invasive GAS disease. Therefore it is difficult to prove the precise role of these virulence genes in causing GAS disease.

20. Distribution of chromosomally mediated toxin and fibronectin binding protein genes were studied in 140 GAS isolates. The most of the commonly known toxins like *slo, ska, scp, speB, speF* and *speG* are present in almost all the isolates except *speZ* and *speJ* which is present only in 37.4% and 16.5 % respectively (Fig. 24). Chromosomally encoded extra cellular products and fibronectin binding proteins are considered to be more stable and might have slowly altered in tune with natural selection and adaptation. Among the surface proteins, *scl* is present in all the isolates and a significant association of *fba* was identified with all the isolates of family patterns E and D. Thus gene *fba* appears to be a marker for skin and general specialist GAS isolates. The *sic* gene was identified in 15 (10.7 %) isolates, where all of which belonged to family pattern A-C except type 28 (n = 1) which belonged to pattern E. All the other factors are randomly present which raises the question as to whether they are associated with mobile genetic elements or phages.

8.2. CONCLUSION

A highly sensitive molecular technique, *emm* typing was successfully standardized to identify a very high degree of heterogeneity among GAS isolates circulating and causing infections among south Indian rural school children. The findings of this study have shown a distinctly different epidemiology from those reported from western countries and Australia. However it is very similar to those reported from Africa, the indigenous population of Australia and Nepal. This study shows high diversity of genotypes among the GAS population that is encountered in different sites of the human body. Further, a comparison of *emm* types contained in the 26-valent hyper-variable region of M proteins with the types identified in this study shows that such a vaccine will not be effective in this population. It is also likely that the high heterogeneity of types encountered in this community will not permit the use of a multivalent M protein based vaccine.

The distinction of throat and skin specific GAS becomes difficult due to the colonization of pharyngitis and impetigo associated types in the normal children combined with high heterogeneity and endemicity of GAS infections. Self transmission of skin isolates to throat plays a major role in the transmission of GAS in this community.

Studies on month-wise distribution of types show dynamic changes in their distribution patterns. This is probably related to introduction of different strains by children attending the same school, but come from different villages in and around the school.

There was no significant association between *emm* family pattern and the three groups of isolates, namely, PT, SK and NT isolates. However, with 53 % of the 74 *emm* types belonging to pattern E, a high percentage of the GAS population in this community are generalists; i.e. they can infect throat or skin or both.

The sequential order of *emm* types arranged in the phylogenetic tree, matrix analysis and the conserved region have confirmed that pattern D is located in between pattern E and A-C and that pattern D has evolved from pattern A-C while pattern E has developed later from pattern D.

The distribution of PAVF among GAS isolates from uncomplicated infections, colonizing strains as well as those causing invasive GAS disease did not show significance difference among them. However a relative lack of these factors were seen among these isolates probably indicating a lack of ability to express virulence factors that may be necessary for the development of disease complications. Therefore it is difficult to prove the precise role of these virulence genes in causing GAS disease.

9. Recommendations

The present study shows high diversity of genotypes among the GAS population circulating in this community. Phylogenetic analysis of these isolates show constant mutational changes resulting in the development of newer strains. This explains the endemicity of GAS infections in this population. Differences in the *emm* types identified in this population shows significant differences from those reported from the west. Therefore a multivalent vaccine based on hyper variable region of M types is not recommended for this population.

In the light of these observations, further studies need to be carried out to find alternate strategies for the control and prevention, if possible, of GAS infections and their complications. Limited study done on the conserved region sequences in the present study indicates that this region of the M protein may be an ideal candidate for a GAS vaccine. Indeed workers in Australia and some other Asian countries have shown the efficacy of such a vaccine. If studies can be conducted to determine the similarities in the conserved region sequences, then it is possible to go in the direction of a pan-Asian conserved region based vaccine for GAS infections.

10. Bibliography

1. Fauci AS. Infectious diseases: considerations for the 21st century. Clin Infect Dis **2001** Mar 1;32(5):675-85.
2. Stevens DL. The flesh-eating bacterium: what's next? J Infect Dis **1999** Mar;179 Suppl 2:S366-74.
3. Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. Lancet Infect Dis **2005** Nov;5(11):685-94.
4. Cunningham MW. Pathogenesis of group A streptococcal infections. Clin Microbiol Rev **2000** Jul;13(3):470-511.
5. Krause RM. A half-century of streptococcal research: then & now. Indian J Med Res **2002** Jun;115:215-41.
6. Tart AH, Walker MJ, Musser JM. New understanding of the group A Streptococcus pathogenesis cycle. Trends Microbiol **2007** Jul;15(7):318-25.
7. Shet A, Kaplan E. Addressing the burden of group A streptococcal disease in India. Indian J Pediatr **2004** Jan;71(1):41-8.
8. Musser JM, Krause, R.M.,. The revival of group A Streptococcal diseases, with commentary on staphylococcal toxic shock syndrome. In Emerging infections. New York: Academic Press **1998**:185-218.
9. 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.
10. 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.
11. Sela S, Barzilai A. Why do we fail with penicillin in the treatment of group A streptococcus infections? Ann Med **1999** Oct;31(5):303-7.
12. Kaplan EL, Chhatwal GS, Rohde M. Reduced ability of penicillin to eradicate ingested group A streptococci from epithelial cells: clinical and pathogenetic implications. Clin Infect Dis **2006** Dec 1;43(11):1398-406.
13. Reid SD, Hoe NP, Smoot LM, Musser JM. Group A Streptococcus: allelic variation, population genetics, and host-pathogen interactions. J Clin Invest **2001** Feb;107(4):393-9.
14. Kehoe MA, Kapur V, Whatmore AM, Musser JM. Horizontal gene transfer among group A streptococci: implications for pathogenesis and epidemiology. Trends Microbiol **1996** Nov;4(11):436-43.
15. Shulman ST, Tanz RR, Kabat W, et al. Group A streptococcal pharyngitis serotype surveillance in North America, 2000-2002. Clin Infect Dis **2004** Aug 1;39(3):325-32.
16. 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.
17. Facklam RF, Martin DR, Lovgren M, et al. Extension of the Lancefield classification for group A streptococci by addition of 22 new M protein gene sequence types from clinical isolates: *emm*103 to *emm*124. Clin Infect Dis **2002** Jan 1;34(1):28-38.
18. Tanna A, Emery M, Dhami C, Arnold E, Efstratiou A. Molecular characterization of clinical isolates of M non-typable group A streptococci from invasive disease cases. J Med Microbiol **2006** Oct;55(Pt 10):1419-23.

19. Facklam R, Beall B, Efstratiou A, et al. *emm* typing and validation of provisional M types for group A streptococci. Emerg Infect Dis **1999** Mar-Apr;5(2):247-53.
20. 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.
21. Podbielski A, Melzer B, Lutticken R. Application of the polymerase chain reaction to study the M protein(-like) gene family in beta-hemolytic streptococci. Med Microbiol Immunol **1991**;180(4):213-27.
22. Sakota V, Fry AM, Lietman TM, Facklam RR, Li Z, Beall B. Genetically diverse group A streptococci from children in far-western Nepal share high genetic relatedness with isolates from other countries. J Clin Microbiol **2006** Jun;44(6):2160-6.
23. Kaplan EL. The group A streptococcal upper respiratory tract carrier state: an enigma. J Pediatr **1980** Sep;97(3):337-45.
24. Anthony BF, Kaplan EL, Wannamaker LW, Chapman SS. The dynamics of streptococcal infections in a defined population of children: serotypes associated with skin and respiratory infections. Am J Epidemiol **1976** Dec;104(6):652-66.
25. Carapetis JR, Currie BJ, Kaplan EL. Epidemiology and prevention of group A streptococcal infections: acute respiratory tract infections, skin infections, and their sequelae at the close of the twentieth century. Clin Infect Dis **1999** Feb;28(2):205-10.
26. Colman G, Tanna A, Efstratiou A, Gaworzewska ET. The serotypes of Streptococcus pyogenes present in Britain during 1980-1990 and their association with disease. J Med Microbiol **1993** Sep;39(3):165-78.
27. Stollerman GH. Rheumatic fever. Lancet **1997** Mar 29;349(9056):935-42.
28. 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.
29. Musser JM, Kapur V, Szeto J, Pan X, Swanson DS, Martin DR. Genetic diversity and relationships among Streptococcus pyogenes strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease. Infect Immun **1995** Mar;63(3):994-1003.
30. 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.
31. Erdem G, Ford JM, Kanenaka RY, Abe L, Yamaga K, Effler PV. Molecular epidemiologic comparison of 2 unusual clusters of group a streptococcal necrotizing fasciitis in Hawaii. Clin Infect Dis **2005** Jun 15;40(12):1851-4.
32. Pruksakorn S, Sittisombut N, Phornphutkul C, Pruksachatkunakorn C, Good MF, Brandt E. Epidemiological analysis of non-M-typeable group A Streptococcus isolates from a Thai population in northern Thailand. J Clin Microbiol **2000** Mar;38(3):1250-4.
33. Dierksen KP, Inglis M, Tagg JR. High pharyngeal carriage rates of Streptococcus pyogenes in Dunedin school children with a low incidence of rheumatic fever. N Z Med J **2000** Nov 24;113(1122):496-9.

34. McDonald M, Brown A, Edwards T, et al. Apparent contrasting rates of pharyngitis and pyoderma in regions where rheumatic heart disease is highly prevalent. Heart Lung Circ **2007** Aug;16(4):254-9.
35. Carapetis JR, Currie BJ. Group A streptococcus, pyoderma, and rheumatic fever. Lancet **1996** May 4;347(9010):1271-2.
36. Johnson DR, Stevens DL, Kaplan EL. Epidemiologic analysis of group A streptococcal serotypes associated with severe systemic infections, rheumatic fever, or uncomplicated pharyngitis. J Infect Dis **1992** Aug;166(2):374-82.
37. Ho PL, Johnson DR, Yue AW, et al. Epidemiologic analysis of invasive and noninvasive group a streptococcal isolates in Hong Kong. J Clin Microbiol **2003** Mar;41(3):937-42.
38. Abdissa A, Asrat D, Kronvall G, et al. High diversity of group A streptococcal *emm* types among healthy schoolchildren in Ethiopia. Clin Infect Dis **2006** May 15;42(10):1362-7.
39. Dey N, McMillan, D.J., Yarwood, P.J., Joshi, R.M., Kumar,R., Good, M.F., Sriprakash, K.S., Vohra,H.,. High Diversity of Group A Streptococcal *emm* types in an Indian Community: The Need to Tailor Multivalent Vaccines. Clinical Infectious Diseases **2005**;40:46-51.
40. Abuhammour W, Hasan RA, Unuvar E. Group A beta-hemolytic streptococcal bacteremia. Indian J Pediatr **2004** Oct;71(10):915-9.
41. Bessen DE, Sotir CM, Readdy TL, Hollingshead SK. Genetic correlates of throat and skin isolates of group A streptococci. J Infect Dis **1996** Apr;173(4):896-900.
42. Beall B, Gherardi G, Lovgren M, Facklam RR, Forwick BA, Tyrrell GJ. *emm* and sof gene sequence variation in relation to serological typing of opacity-factor-positive group A streptococci. Microbiology **2000** May;146 (Pt 5):1195-209.
43. Kalia A, Bessen DE. Natural selection and evolution of streptococcal virulence genes involved in tissue-specific adaptations. J Bacteriol **2004** Jan;186(1):110-21.
44. Struelens MJ. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. Clin Microbiol Infect **1996** Aug;2(1):2-11.
45. Van Belkum A, Struelens M, de Visser A, Verbrugh H, Tibayrenc M. Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology. Clin Microbiol Rev **2001** Jul;14(3):547-60.
46. Boyd EF, Brussow H. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. Trends Microbiol **2002** Nov;10(11):521-9.
47. Banks DJ, Beres SB, Musser JM. The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. Trends Microbiol **2002** Nov;10(11):515-21.
48. Banks DJ, Porcella SF, Barbian KD, et al. Progress toward characterization of the group A Streptococcus metagenome: complete genome sequence of a macrolide-resistant serotype M6 strain. J Infect Dis **2004** Aug 15;190(4):727-38.
49. Green NM, Beres SB, Graviss EA, et al. Genetic diversity among type *emm*28 group A Streptococcus strains causing invasive infections and pharyngitis. J Clin Microbiol **2005** Aug;43(8):4083-91.

50. Okada N, Pentland AP, Falk P, Caparon MG. M protein and protein F act as important determinants of cell-specific tropism of Streptococcus pyogenes in skin tissue. J Clin Invest **1994** Sep;94(3):965-77.
51. Hanski E, Caparon M. Protein F, a fibronectin-binding protein, is an adhesin of the group A streptococcus Streptococcus pyogenes. Proc Natl Acad Sci U S A **1992** Jul 1;89(13):6172-6.
52. Molinari G, Talay SR, Valentin-Weigand P, Rohde M, Chhatwal GS. The fibronectin-binding protein of Streptococcus pyogenes, SfbI, is involved in the internalization of group A streptococci by epithelial cells. Infect Immun **1997** Apr;65(4):1357-63.
53. Rocha CL, Fischetti VA. Identification and characterization of a novel fibronectin-binding protein on the surface of group A streptococci. Infect Immun **1999** Jun;67(6):2720-8.
54. Musser JM, DeLeo FR. Toward a genome-wide systems biology analysis of host-pathogen interactions in group A Streptococcus. Am J Pathol **2005** Dec;167(6):1461-72.
55. Simpson WJ, Musser JM, Cleary PP. Evidence consistent with horizontal transfer of the gene (*emm*12) encoding serotype M12 protein between group A and group G pathogenic streptococci. Infect Immun **1992** May;60(5):1890-3.
56. Hollingshead SK, Arnold J, Readdy TL, Bessen DE. Molecular evolution of a multigene family in group A streptococci. Mol Biol Evol **1994** Mar;11(2):208-19.
57. Reda KB, Kapur V, Mollick JA, Lamphear JG, Musser JM, Rich RR. Molecular characterization and phylogenetic distribution of the streptococcal superantigen gene (ssa) from Streptococcus pyogenes. Infect Immun **1994** May;62(5):1867-74.
58. Johnson DR, E. Kaplan, J. Sramek, R. Bicova, J. Havlicek, H. Havlickova,J. Motlova, and P. Kriz. Laboratory methods for the diagnosis of group A streptococcal infections,WHO Manual **1996**.
59. Koneman E. Text book of diagnostic microbiology, 5 th edition. **1997**.
60. Kim SJ. Bacteriologic characteristics and serotypings of Streptococcus pyogenes isolated from throats of school children. Yonsei Med J **2000** Feb;41(1):56-60.
61. Kocka FE, Chittom AL, Sanders L, et al. Nutritionally variant Streptococcus pyogenes from a periorbital abscess. J Clin Microbiol **1987** Apr;25(4):736-7.
62. Winn WJ, Allen S., Janda W., Koneman E., Procop G., Schreckenberger P., Woods G.,. Color Atlas and Textbook of Diagnostic Microbiology. Book, Lippincott Williams & Wilkins **2006**;5th edition:674.
63. Efstratiou A. Group A streptococci in the 1990s. J Antimicrob Chemother **2000** Feb;45 Suppl:3-12.
64. Hoffmann S. The throat carrier rate of group A and other beta hemolytic streptococci among patients in general practice. Acta Pathol Microbiol Immunol Scand [B] **1985** Oct;93(5):347-51.
65. Kaplan EL. Recent epidemiology of group A streptococcal infections in North America and abroad: an overview. Pediatrics **1996** Jun;97(6 Pt 2):945-8.
66. Rathore MH, Barton LL, Kaplan EL. Suppurative group A beta-hemolytic streptococcal infections in children. Pediatrics **1992** Apr;89(4 Pt 2):743-6.
67. Heggie AD, Jacobs MR, Linz PE, Han DP, Kaplan EL, Boxerbaum B. Prevalence and characteristics of pharyngeal group A beta-hemolytic

streptococci in US Navy recruits receiving benzathine penicillin prophylaxis. J Infect Dis **1992** Nov;166(5):1006-13.

68. Bisno AL, Craven DE, McCabe WR. M proteins of group G streptococci isolated from bacteremic human infections. Infect Immun **1987** Mar;55(3):753-7.
69. Wannamaker LW. Differences between streptococcal infections of the throat and of the skin (second of two parts). N Engl J Med **1970** Jan 8;282(2):78-85.
70. Johnson DR, Kaplan EL. A review of the correlation of T-agglutination patterns and M-protein typing and opacity factor production in the identification of group A streptococci. J Med Microbiol **1993** May;38(5):311-5.
71. Bisno AL, Stevens DL. Streptococcal infections of skin and soft tissues. N Engl J Med **1996** Jan 25;334(4):240-5.
72. Dudding BA, Burnett JW, Chapman SS, et al. The role of normal skin in the spread of streptococcal pyoderma. J Hyg (Camb) **1970**;68:19-28.
73. Streeton CL, Hanna JN, Messer RD, Merianos A. An epidemic of acute post-streptococcal glomerulonephritis among aboriginal children. J Paediatr Child Health **1995** Jun;31(3):245-8.
74. Eriksson BK, Andersson J, Holm SE, Norgren M. Epidemiological and clinical aspects of invasive group A streptococcal infections and the streptococcal toxic shock syndrome. Clin Infect Dis **1998** Dec;27(6):1428-36.
75. Eriksson BK, Andersson J, Holm SE, Norgren M. Invasive group A streptococcal infections: T1M1 isolates expressing pyrogenic exotoxins A and B in combination with selective lack of toxin-neutralizing antibodies are associated with increased risk of streptococcal toxic shock syndrome. J Infect Dis **1999** Aug;180(2):410-8.
76. Guzman CA, Talay SR, Molinari G, Medina E, Chhatwal GS. Protective immune response against Streptococcus pyogenes in mice after intranasal vaccination with the fibronectin-binding protein SfbI. J Infect Dis **1999** Apr;179(4):901-6.
77. Bisno AL, Pearce IA, Wall HP, Moody MD, Stollerman GH. Contrasting epidemiology of acute rheumatic fever and acute glomerulonephritis. N Engl J Med **1970** Sep 10;283(11):561-5.
78. Carapetis JR, Currie BJ, Good MF. Towards understanding the pathogenesis of rheumatic fever. Scand J Rheumatol **1996**;25(3):127-31; discussion 32-3.
79. Stollerman GH. Rheumatogenic streptococci and autoimmunity. Clin Immunol Immunopathol **1991** Nov;61(2 Pt 1):131-42.
80. Veasy LG, Hill HR. Immunologic and clinical correlations in rheumatic fever and rheumatic heart disease. Pediatr Infect Dis J **1997** Apr;16(4):400-7.
81. Smoot JC, Barbian KD, Van Gompel JJ, et al. Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks. Proc Natl Acad Sci U S A **2002** Apr 2;99(7):4668-73.
82. Pichichero ME. The rising incidence of penicillin treatment failures in group A streptococcal tonsillopharyngitis: an emerging role for the cephalosporins? Pediatr Infect Dis J **1991** Oct;10(10 Suppl):S50-5.
83. Seppala H, Nissinen A, Jarvinen H, et al. Resistance to erythromycin in group A streptococci. N Engl J Med **1992** Jan 30;326(5):292-7.

84. Davies HD, McGeer A, Schwartz B, et al. Invasive group A streptococcal infections in Ontario, Canada. Ontario Group A Streptococcal Study Group. N Engl J Med **1996** Aug 22;335(8):547-54.
85. Stevens DL, Gibbons AE, Bergstrom R, Winn V. The Eagle effect revisited: efficacy of clindamycin, erythromycin, and penicillin in the treatment of streptococcal myositis. J Infect Dis **1988** Jul;158(1):23-8.
86. Krause RM. Prevention of streptococcal sequelae by penecillin prophylaxis: A reassessment. J Infect Diseases **1975**;131(5):592.
87. Koshi G, Benjamin V, Cherian G. Rheumatic fever and rheumatic heart disease in rural South Indian children. Bull World Health Organ **1981**;59(4):599-603.
88. Jose vJ, Gomathi M.,. Declining Prevalence of Rheumatic Heart Disease in Rural School children in India: 2001-2002. Indian Heart J **2003**;55:158-60.
89. Menon T, Shanmugasundaram S, Kumar MP, Kumar CP. Group A streptococcal infections of the pharynx in a rural population in south India. Indian J Med Res **2004** May;119 Suppl:171-3.
90. Brahmadathan KN, Gladstone P. Microbiological diagnosis of streptococcal pharyngitis: lacunae and their implications. Indian J Med Microbiol **2006** Apr;24(2):92-6.
91. Kaplan EL, Johnson DR, Rehder CD. Recent changes in group A streptococcal serotypes from uncomplicated pharyngitis: a reflection of the changing epidemiology of severe group A infections? J Infect Dis **1994** Nov;170(5):1346-7.
92. Stevens DL, Salmi DB, McIndoo ER, Bryant AE. Molecular epidemiology of nga and NAD glycohydrolase/ADP-ribosyltransferase activity among Streptococcus pyogenes causing streptococcal toxic shock syndrome. J Infect Dis **2000** Oct;182(4):1117-28.
93. Kaplan EL. Pathogenesis of acute rheumatic fever and rheumatic heart disease: evasive after half a century of clinical, epidemiological, and laboratory investigation. Heart **2005** Jan;91(1):3-4.
94. Cleary PP, Kaplan EL, Handley JP, et al. Clonal basis for resurgence of serious Streptococcus pyogenes disease in the 1980s. Lancet **1992** Feb 29;339(8792):518-21.
95. Nelson K, Schlievert PM, Selander RK, Musser JM. Characterization and clonal distribution of four alleles of the speA gene encoding pyrogenic exotoxin A (scarlet fever toxin) in Streptococcus pyogenes. J Exp Med **1991** Nov 1;174(5):1271-4.
96. Kapur V, Nelson K, Schlievert PM, Selander RK, Musser JM. Molecular population genetic evidence of horizontal spread of two alleles of the pyrogenic exotoxin C gene (speC) among pathogenic clones of Streptococcus pyogenes. Infect Immun **1992** Sep;60(9):3513-7.
97. Musser JM, Nelson K, Selander RK, et al. Temporal variation in bacterial disease frequency: molecular population genetic analysis of scarlet fever epidemics in Ottawa and in eastern Germany. J Infect Dis **1993** Mar;167(3):759-62.
98. Musser JM, Kapur V, Kanjilal S, et al. Geographic and temporal distribution and molecular characterization of two highly pathogenic clones of Streptococcus pyogenes expressing allelic variants of pyrogenic exotoxin A (Scarlet fever toxin). J Infect Dis **1993** Feb;167(2):337-46.

99. Sumby P, Barbian KD, Gardner DJ, et al. Extracellular deoxyribonuclease made by group A Streptococcus assists pathogenesis by enhancing evasion of the innate immune response. Proc Natl Acad Sci U S A **2005** Feb 1;102(5):1679-84.
100. Sumby P, Porcella SF, Madrigal AG, et al. Evolutionary origin and emergence of a highly successful clone of serotype M1 group a Streptococcus involved multiple horizontal gene transfer events. J Infect Dis **2005** Sep 1;192(5):771-82.
101. Higgins PM. Streptococcal pharyngitis in general practice. 1. Some unusual features of the epidemiology. Epidemiol Infect **1992** Oct;109(2):181-9.
102. Levine OS, Van Beneden CA, Jernigan DB. A new old opportunity for preventing serious group A streptococcal infections. Clin Infect Dis **2005** Aug 1;41(3):343-4.
103. Riley LW. Molecular epidemiology of infectious diseases: Principles and practices. Book, ASM Press **2004**.
104. Struelens M. Consensus Guidelines for appropriate use and evaluation of microbial typing systems. . Clinical microbiol infect **1996**;2:2-11.
105. Tenover F.C. Interpreating chromosal DNA restriction patterns produced by pulse field gel electrophoresis: criteria for bacterial strain typing. J Clin Microbiol **1994**;33:2233-9.
106. Van Belkum. sM, et. al. Role of genomic typing in taxonomy, Evolutionary Genetics, and Microbial Epidemiology. Clin Micriobiol Reviews **2001**;14(3):547-60.
107. Gardiner D, Hartas J, Currie B, Mathews JD, Kemp DJ, Sriprakash KS. Vir typing: a long-PCR typing method for group A streptococci. PCR Methods Appl **1995** Apr;4(5):288-93.
108. Rota PA. Characterisation of a novel coronavirus associated with Severe acute respirstory syndrome. Science **2003**;300:1394-9.
109. Mead PB, Winn WC. Vaginal-rectal colonization with group A streptococci in late pregnancy. Infect Dis Obstet Gynecol **2000**;8(5-6):217-9.
110. Reid SD, Green NM, Sylva GL, et al. Postgenomic analysis of four novel antigens of group a streptococcus: growth phase-dependent gene transcription and human serologic response. J Bacteriol **2002** Nov;184(22):6316-24.
111. Whatmore AM, Kapur V, Sullivan DJ, Musser JM, Kehoe MA. Non-congruent relationships between variation in *emm* gene sequences and the population genetic structure of group A streptococci. Mol Microbiol **1994** Nov;14(4):619-31.
112. Kratovac Z, Manoharan A, Luo F, Lizano S, Bessen DE. Population genetics and linkage analysis of loci within the FCT region of Streptococcus pyogenes. J Bacteriol **2007** Feb;189(4):1299-310.
113. Bessen DE, Manoharan A, Luo F, Wertz JE, Robinson DA. Evolution of transcription regulatory genes is linked to niche specialization in the bacterial pathogen Streptococcus pyogenes. J Bacteriol **2005** Jun;187(12):4163-72.
114. Evans MJ, Hamer JM, Gason LM, Irvine CH. Factors affecting uterine clearance of inoculated materials in mares. J Reprod Fertil Suppl **1987**;35:327-34.
115. Read SE, Fischetti VA, Utermohlen V, Falk RE, Zabriskie JB. Cellular reactivity studies to streptococcal antigens. Migration inhibition studies in patients with streptococcal infections and rheumatic fever. J Clin Invest **1974** Aug;54(2):439-50.

116. van de Rijn I, Fischetti VA. Immunochemical analysis of intact M protein secreted from cell wall-less streptococci. Infect Immun **1981** Apr;32(1):86-91.
117. Gardiner DL, Goodfellow AM, Martin DR, Sriprakash KS. Group A streptococcal Vir types are M-protein gene (*emm*) sequence type specific. J Clin Microbiol **1998** Apr;36(4):902-7.
118. Kaufhold A, Podbielski A, Johnson DR, Kaplan EL, Lutticken R. M protein gene typing of Streptococcus pyogenes by nonradioactively labeled oligonucleotide probes. J Clin Microbiol **1992** Sep;30(9):2391-7.
119. 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.
120. Martin JM, Wald ER, Green M. Field inversion gel electrophoresis as a typing system for group A streptococcus. J Infect Dis **1998** Feb;177(2):504-7.
121. Mascini EM, Jansze M, Schouls LM, Fluit AC, Verhoef J, van Dijk H. Invasive and noninvasive group A streptococcal isolates with different speA alleles in The Netherlands: genetic relatedness and production of pyrogenic exotoxins A and B. J Clin Microbiol **1999** Nov;37(11):3469-74.
122. Single LA, Martin DR. Clonal differences within M-types of the group A Streptococcus revealed by pulsed field gel electrophoresis. FEMS Microbiol Lett **1992** Feb 1;70(1):85-9.
123. Cleary PP, Kaplan EL, Livdahl C, Skjold S. DNA fingerprints of Streptococcus pyogenes are M type specific. J Infect Dis **1988** Dec;158(6):1317-23.
124. Johnson DR, Kaplan EL. Microtechnique for serum opacity factor characterization of group A streptococci adaptable to the use of human sera. J Clin Microbiol **1988** Oct;26(10):2025-30.
125. Bruneau S, de Montclos H, Drouet E, Denoyel GA. rRNA gene restriction patterns of Streptococcus pyogenes: epidemiological applications and relation to serotypes. J Clin Microbiol **1994** Dec;32(12):2953-8.
126. Seppala H, Vuopio-Varkila J, Osterblad M, et al. Evaluation of methods for epidemiologic typing of group A streptococci. J Infect Dis **1994** Mar;169(3):519-25.
127. Sriprakash KS, Gardiner DL. Lack of polymorphism within the rRNA operons of group A streptococci. Mol Gen Genet **1997** Jun;255(1):125-30.
128. Gardiner DL, Sriprakash KS. Molecular epidemiology of impetiginous group A streptococcal infections in aboriginal communities of northern Australia. J Clin Microbiol **1996** Jun;34(6):1448-52.
129. Facklam R, Beall B. Anomalies in *emm* typing of group A streptococci. Adv Exp Med Biol **1997**;418:335-7.
130. Salemi M, Vandamme A.M.,. The Phylogebetic Handbook, A practical approach to DNA and Proteion Phylogeny. Cambridge University Press **2003**.
131. Kimura M. The neutral theory of molecular evolution cambridge. Cambridge University Press **1983**.
132. Reid SD, Green NM, Buss JK, Lei B, Musser JM. Multilocus analysis of extracellular putative virulence proteins made by group A Streptococcus: population genetics, human serologic response, and gene transcription. Proc Natl Acad Sci U S A **2001** Jun 19;98(13):7552-7.
133. Bessen DE, Fiorentino TR, Hollingshead SK. Molecular markers for throat and skin isolates of group A streptococci. Adv Exp Med Biol **1997**;418:537-43.

134. 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.
135. Hollingshead SK, Readdy TL, Yung DL, Bessen DE. Structural heterogeneity of the *emm* gene cluster in group A streptococci. Mol Microbiol **1993** May;8(4):707-17.
136. Hollingshead SK, Bessen DE. Evolution of the *emm* gene family: virulence gene clusters in group A streptococci. Dev Biol Stand **1995**;85:163-8.
137. Spratt BG. Multilocus sequence typing: molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the internet. Curr Opin Microbiol **1999**;2:312-6.
138. Maiden MCJ, Bygraves J.A., Feil E., et al.,. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci **1998**;95:3140-5.
139. McGregor KF, Spratt BG, Kalia A, et al. Multilocus sequence typing of Streptococcus pyogenes representing most known *emm* types and distinctions among subpopulation genetic structures. J Bacteriol **2004** Jul;186(13):4285-94.
140. Ferretti JJ, McShan WM, Ajdic D, et al. Complete genome sequence of an M1 strain of Streptococcus pyogenes. Proc Natl Acad Sci U S A **2001** Apr 10;98(8):4658-63.
141. Smoot JC, Korgenski EK, Daly JA, Veasy LG, Musser JM. Molecular analysis of group A Streptococcus type *emm*18 isolates temporally associated with acute rheumatic fever outbreaks in Salt Lake City, Utah. J Clin Microbiol **2002** May;40(5):1805-10.
142. Shelburne SA, 3rd, Sumby P, Sitkiewicz I, Granville C, DeLeo FR, Musser JM. Central role of a bacterial two-component gene regulatory system of previously unknown function in pathogen persistence in human saliva. Proc Natl Acad Sci U S A **2005** Nov 1;102(44):16037-42.
143. Ye RW, Wang T., Bedzyk L., Croker K.M. Applications of DNA microarrays in Microbial systems. J moicrobiol Methods **2001**;47:257-72.
144. Fluckiger U, Jones KF, Fischetti VA. Immunoglobulins to group A streptococcal surface molecules decrease adherence to and invasion of human pharyngeal cells. Infect Immun **1998** Mar;66(3):974-9.
145. Fontan PA, Pancholi V, Nociari MM, Fischetti VA. Antibodies to streptococcal surface enolase react with human alpha-enolase: implications in poststreptococcal sequelae. J Infect Dis **2000** Dec;182(6):1712-21.
146. Courtney HS, Hasty DL, Dale JB. Molecular mechanisms of adhesion, colonization, and invasion of group A streptococci. Ann Med **2002**;34(2):77-87.
147. Hasty DL, Ofek I, Courtney HS, Doyle RJ. Multiple adhesins of streptococci. Infect Immun **1992** Jun;60(6):2147-52.
148. Ellen RP, Gibbons RJ. M protein-associated adherence of Streptococcus pyogenes to epithelial surfaces: prerequisite for virulence. Infect Immun **1972** May;5(5):826-30.
149. Pancholi V, Fischetti VA. alpha-enolase, a novel strong plasmin(ogen) binding protein on the surface of pathogenic streptococci. J Biol Chem **1998** Jun 5;273(23):14503-15.

150. Hollingshead SK, Simecka JW, Michalek SM. Role of M protein in pharyngeal colonization by group A streptococci in rats. Infect Immun **1993** Jun;61(6):2277-83.
151. Beachey EH, Ofek I. Epithelial cell binding of group A streptococci by lipoteichoic acid on fimbriae denuded of M protein. J Exp Med **1976** Apr 1;143(4):759-71.
152. Ofek I, Simpson WA, Beachey EH. Formation of molecular complexes between a structurally defined M protein and acylated or deacylated lipoteichoic acid of Streptococcus pyogenes. J Bacteriol **1982** Feb;149(2):426-33.
153. Simpson WA, Beachey EH. Adherence of group A streptococci to fibronectin on oral epithelial cells. Infect Immun **1983** Jan;39(1):275-9.
154. Fischetti VA. Streptococcal M protein. Sci Am **1991** Jun;264(6):58-65.
155. Moses AE, Wessels MR, Zalcman K, et al. Relative contributions of hyaluronic acid capsule and M protein to virulence in a mucoid strain of the group A Streptococcus. Infect Immun **1997** Jan;65(1):64-71.
156. 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.
157. Hoe NP, Ireland RM, DeLeo FR, et al. Insight into the molecular basis of pathogen abundance: group A Streptococcus inhibitor of complement inhibits bacterial adherence and internalization into human cells. Proc Natl Acad Sci U S A **2002** May 28;99(11):7646-51.
158. Iafrate A, Feuk., L., Rivera M.,N., et. al., . Detection of large scale variation in the human genome. Nat Genet **2004**;36:949-51.
159. LaPenta D, Rubens C, Chi E, Cleary PP. Group A streptococci efficiently invade human respiratory epithelial cells. Proc Natl Acad Sci U S A **1994** Dec 6;91(25):12115-9.
160. Purushothaman SS, Wang B, Cleary PP. M1 protein triggers a phosphoinositide cascade for group A Streptococcus invasion of epithelial cells. Infect Immun **2003** Oct;71(10):5823-30.
161. Osterlund A, Engstrand L. An intracellular sanctuary for Streptococcus pyogenes in human tonsillar epithelium--studies of asymptomatic carriers and in vitro cultured biopsies. Acta Otolaryngol **1997** Nov;117(6):883-8.
162. Shikhman AR, Greenspan NS, Cunningham MW. A subset of mouse monoclonal antibodies cross-reactive with cytoskeletal proteins and group A streptococcal M proteins recognizes N-acetyl-beta-D-glucosamine. J Immunol **1993** Oct 1;151(7):3902-13.
163. Cunningham MW, Antone SM, Smart M, Liu R, Kosanke S. Molecular analysis of human cardiac myosin-cross-reactive B- and T-cell epitopes of the group A streptococcal M5 protein. Infect Immun **1997** Sep;65(9):3913-23.
164. Kil KS, Cunningham MW, Barnett LA. Cloning and sequence analysis of a gene encoding a 67-kilodalton myosin-cross-reactive antigen of Streptococcus pyogenes reveals its similarity with class II major histocompatibility antigens. Infect Immun **1994** Jun;62(6):2440-9.
165. Dinkla K, Rohde M, Jansen WT, Kaplan EL, Chhatwal GS, Talay SR. Rheumatic fever-associated Streptococcus pyogenes isolates aggregate collagen. J Clin Invest **2003** Jun;111(12):1905-12.
166. Kotb M. Bacterial pyrogenic exotoxins as superantigens. Clin Micriobiol Reviews **1995**;8:411-26.

167. Musser JM. Streptococcal superantigen, mitogenic factor, and pyrogenic exotoxin B expressed by Streptococcus pyogenes. Structure and function. Prep Biochem Biotechnol **1997** May-Aug;27(2-3):143-72.
168. Beres SB, Sylva G.L., Barbian K.D., et al. . Genome sequence of serotype M3 strain of group A streptococcus: phage-encoded toxins , the high virulence phenotype, and clone emergance. Proc Natl Acad Sci **2002**;99:10078-83.
169. Canchaya C, Desiere F, McShan WM, Ferretti JJ, Parkhill J, Brussow H. Genome analysis of an inducible prophage and prophage remnants integrated in the Streptococcus pyogenes strain SF370. Virology **2002** Oct 25;302(2):245-58.
170. Beres SB, Sylva GL, Sturdevant DE, et al. Genome-wide molecular dissection of serotype M3 group A Streptococcus strains causing two epidemics of invasive infections. Proc Natl Acad Sci U S A **2004** Aug 10;101(32):11833-8.
171. Hollingshead SK, Fischetti VA, Scott JR. Complete nucleotide sequence of type 6 M protein of the group A Streptococcus. Repetitive structure and membrane anchor. J Biol Chem **1986** Feb 5;261(4):1677-86.
172. Fischetti VA. Streptococcal M protein: molecular design and biological behavior. Clin Microbiol Rev **1989**;2:285-314.
173. Bessen D, Fischetti VA. Influence of intranasal immunization with synthetic peptides corresponding to conserved epitopes of M protein on mucosal colonization by group A streptococci. Infect Immun **1988** Oct;56(10):2666-72.
174. Fischetti VA, Jones KF, Scott JR. Size variation of the M protein in group A streptococci. J Exp Med **1985** Jun 1;161(6):1384-401.
175. Caparon MG, Stephens DS, Olsen A, Scott JR. Role of M protein in adherence of group A streptococci. Infect Immun **1991** May;59(5):1811-7.
176. Kreikemeyer B, McIver KS, Podbielski A. Virulence factor regulation and regulatory networks in Streptococcus pyogenes and their impact on pathogen-host interactions. Trends Microbiol **2003** May;11(5):224-32.
177. Bessen D, Jones KF, Fischetti VA. Evidence for two distinct classes of streptococcal M protein and their relationship to rheumatic fever. J Exp Med **1989** Jan 1;169(1):269-83.
178. Bessen DE, Fischetti VA. Differentiation between two biologically distinct classes of group A streptococci by limited substitutions of amino acids within the shared region of M protein-like molecules. J Exp Med **1990** Dec 1;172(6):1757-64.
179. Brandt ER, Currie B, Mammo L, Pruksakorn S, Good MF. Can class I epitope of M protein be a diagnostic marker for rheumatic fever in populations endemic for group A streptococci? Lancet **1998** Jun 20;351(9119):1860.
180. Bessen DE, Veasy LG, Hill HR, Augustine NH, Fischetti VA. Serologic evidence for a class I group A streptococcal infection among rheumatic fever patients. J Infect Dis **1995** Dec;172(6):1608-11.
181. Perez-Casal J, Caparon MG, Scott JR. Introduction of the *emm*6 gene into an *emm*-deleted strain of Streptococcus pyogenes restores its ability to resist phagocytosis. Res Microbiol **1992** Jul-Aug;143(6):549-58.
182. Perez-Casal J, Okada N, Caparon MG, Scott JR. Role of the conserved C-repeat region of the M protein of Streptococcus pyogenes. Mol Microbiol **1995** Mar;15(5):907-16.
183. Christner R, Li Z, Raeder R, Podbielski A, Boyle MD. Identification of key gene products required for acquisition of plasmin-like enzymatic activity by group A streptococci. J Infect Dis **1997** May;175(5):1115-20.

184. Dale JB, Beachey EH. Epitopes of streptococcal M proteins shared with cardiac myosin. J Exp Med **1985** Aug 1;162(2):583-91.
185. Dale JB, Chiang EY, Lederer JW. Recombinant tetravalent group A streptococcal M protein vaccine. J Immunol **1993** Aug 15;151(4):2188-94.
186. Schrager HM, Alberti S, Cywes C, Dougherty GJ, Wessels MR. Hyaluronic acid capsule modulates M protein-mediated adherence and acts as a ligand for attachment of group A Streptococcus to CD44 on human keratinocytes. J Clin Invest **1998** Apr 15;101(8):1708-16.
187. Dale JB, Washburn RG, Marques MB, Wessels MR. Hyaluronate capsule and surface M protein in resistance to opsonization of group A streptococci. Infect Immun **1996** May;64(5):1495-501.
188. Husmann LK, Yung DL, Hollingshead SK, Scott JR. Role of putative virulence factors of Streptococcus pyogenes in mouse models of long-term throat colonization and pneumonia. Infect Immun **1997** Apr;65(4):1422-30.
189. Wessels MR, Bronze MS. Critical role of the group A streptococcal capsule in pharyngeal colonization and infection in mice. Proc Natl Acad Sci U S A **1994** Dec 6;91(25):12238-42.
190. 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.
191. Schrager HM, Rheinwald JG, Wessels MR. Hyaluronic acid capsule and the role of streptococcal entry into keratinocytes in invasive skin infection. J Clin Invest **1996** Nov 1;98(9):1954-8.
192. Kehoe M, Timmis KN. Cloning and expression in Escherichia coli of the streptolysin O determinant from Streptococcus pyogenes: characterization of the cloned streptolysin O determinant and demonstration of the absence of substantial homology with determinants of other thiol-activated toxins. Infect Immun **1984** Mar;43(3):804-10.
193. Limbago B, Penumalli V, Weinrick B, Scott JR. Role of streptolysin O in a mouse model of invasive group A streptococcal disease. Infect Immun **2000** Nov;68(11):6384-90.
194. Gupta RK, Gupta S. Cardiotoxicity of streptolysin O of group A and group C haemolytic streptococci. Indian J Pathol Microbiol **1986** Jul;29(3):249-54.
195. Kaplan MH. Rheumatic fever, rheumatic heart disease, and the streptococcal connection: the role of streptococcal antigens cross-reactive with heart tissue. Rev Infect Dis **1979** Nov-Dec;1(6):988-86.
196. Ringdahl U, Svensson M, Wistedt AC, et al. Molecular co-operation between protein PAM and streptokinase for plasmin acquisition by Streptococcus pyogenes. J Biol Chem **1998** Mar 13;273(11):6424-30.
197. Ohkuni H, Todome Y, Suzuki H, et al. Immunochemical studies and complete amino acid sequence of the streptokinase from Streptococcus pyogenes (group A) M type 12 strain A374. Infect Immun **1992** Jan;60(1):278-83.
198. Lottenberg R, DesJardin LE, Wang H, Boyle MD. Streptokinase-producing streptococci grown in human plasma acquire unregulated cell-associated plasmin activity. J Infect Dis **1992** Aug;166(2):436-40.
199. Chen CC, Cleary PP. Complete nucleotide sequence of the streptococcal C5a peptidase gene of Streptococcus pyogenes. J Biol Chem **1990** Feb 25;265(6):3161-7.

200. Shet A, Kaplan EL, Johnson DR, Cleary PP. Immune response to group A streptococcal C5a peptidase in children: implications for vaccine development. J Infect Dis **2003** Sep 15;188(6):809-17.
201. O'Connor SP, Cleary PP. Localization of the streptococcal C5a peptidase to the surface of group A streptococci. Infect Immun **1986** Aug;53(2):432-4.
202. Ji Y, McLandsborough L, Kondagunta A, Cleary PP. C5a peptidase alters clearance and trafficking of group A streptococci by infected mice. Infect Immun **1996** Feb;64(2):503-10.
203. Kapur V, Maffei JT, Greer RS, Li LL, Adams GJ, Musser JM. Vaccination with streptococcal extracellular cysteine protease (interleukin-1 beta convertase) protects mice against challenge with heterologous group A streptococci. Microb Pathog **1994** Jun;16(6):443-50.
204. Rakonjac JV, Robbins JC, Fischetti VA. DNA sequence of the serum opacity factor of group A streptococci: identification of a fibronectin-binding repeat domain. Infect Immun **1995** Feb;63(2):622-31.
205. Courtney HS, Hasty DL, Li Y, Chiang HC, Thacker JL, Dale JB. Serum opacity factor is a major fibronectin-binding protein and a virulence determinant of M type 2 Streptococcus pyogenes. Mol Microbiol **1999** Apr;32(1):89-98.
206. Johnson DR, Kaplan EL, VanGheem A, Facklam RR, Beall B. Characterization of group A streptococci (Streptococcus pyogenes): correlation of M-protein and *emm*-gene type with T-protein agglutination pattern and serum opacity factor. J Med Microbiol **2006** Feb;55(Pt 2):157-64.
207. Akesson P, Sjoholm AG, Bjorck L. Protein SIC, a novel extracellular protein of Streptococcus pyogenes interfering with complement function. J Biol Chem **1996** Jan 12;271(2):1081-8.
208. Frick IM, Akesson P, Rasmussen M, Schmidtchen A, Bjorck L. SIC, a secreted protein of Streptococcus pyogenes that inactivates antibacterial peptides. J Biol Chem **2003** May 9;278(19):16561-6.
209. Hoe NP, Nakashima K, Lukomski S, et al. Rapid selection of complement-inhibiting protein variants in group A Streptococcus epidemic waves. Nat Med **1999** Aug;5(8):924-9.
210. Fernie-King BA, Seilly DJ, Davies A, Lachmann PJ. Streptococcal inhibitor of complement inhibits two additional components of the mucosal innate immune system: secretory leukocyte proteinase inhibitor and lysozyme. Infect Immun **2002** Sep;70(9):4908-16.
211. Hartas J, Sriprakash KS. Streptococcus pyogenes strains containing *emm*12 and *emm*55 possess a novel gene coding for distantly related SIC protein. Microb Pathog **1999** Jan;26(1):25-33.
212. Bessen DE, Kalia A. Genomic localization of a T serotype locus to a recombinatorial zone encoding extracellular matrix-binding proteins in Streptococcus pyogenes. Infect Immun **2002** Mar;70(3):1159-67.
213. Kreikemeyer B, Nakata M, Oehmcke S, Gschwendtner C, Normann J, Podbielski A. Streptococcus pyogenes collagen type I-binding Cpa surface protein. Expression profile, binding characteristics, biological functions, and potential clinical impact. J Biol Chem **2005** Sep 30;280(39):33228-39.
214. 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.

215. Lukomski S, Nakashima K, Abdi I, et al. Identification and characterization of the scl gene encoding a group A Streptococcus extracellular protein virulence factor with similarity to human collagen. Infect Immun **2000** Dec;68(12):6542-53.
216. Rasmussen M, Eden A, Bjorck L. SclA, a novel collagen-like surface protein of Streptococcus pyogenes. Infect Immun **2000** Nov;68(11):6370-7.
217. Rasmussen M, Bjorck L. Unique regulation of SclB - a novel collagen-like surface protein of Streptococcus pyogenes. Mol Microbiol **2001** Jun;40(6):1427-38.
218. Natanson S, Sela S, Moses AE, Musser JM, Caparon MG, Hanski E. Distribution of fibronectin-binding proteins among group A streptococci of different M types. J Infect Dis **1995** Apr;171(4):871-8.
219. Fogg GC, Caparon MG. Constitutive expression of fibronectin binding in Streptococcus pyogenes as a result of anaerobic activation of rofA. J Bacteriol **1997** Oct;179(19):6172-80.
220. Jaffe J, Natanson-Yaron S, Caparon MG, Hanski E. Protein F2, a novel fibronectin-binding protein from Streptococcus pyogenes, possesses two binding domains. Mol Microbiol **1996** Jul;21(2):373-84.
221. Goodfellow AM, Hibble M, Talay SR, et al. Distribution and antigenicity of fibronectin binding proteins (SfbI and SfbII) of Streptococcus pyogenes clinical isolates from the northern territory, Australia. J Clin Microbiol **2000** Jan;38(1):389-92.
222. Katerov V, Andreev A, Schalen C, Totolian AA. Protein F, a fibronectin-binding protein of Streptococcus pyogenes, also binds human fibrinogen: isolation of the protein and mapping of the binding region. Microbiology **1998** Jan;144 (Pt 1):119-26.
223. Terao Y, Kawabata S, Kunitomo E, Murakami J, Nakagawa I, Hamada S. Fba, a novel fibronectin-binding protein from Streptococcus pyogenes, promotes bacterial entry into epithelial cells, and the fba gene is positively transcribed under the Mga regulator. Mol Microbiol **2001** Oct;42(1):75-86.
224. Ramachandran V, McArthur JD, Behm CE, et al. Two distinct genotypes of prtF2, encoding a fibronectin binding protein, and evolution of the gene family in Streptococcus pyogenes. J Bacteriol **2004** Nov;186(22):7601-9.
225. Chhatwal GS. Anchorless adhesins and invasins of Gram-positive bacteria: a new class of virulence factors. Trends Microbiol **2002**;10:205-8.
226. Courtney HS, Dale JB, Hasty DI. Differential effects of the streptococcal fibronectin-binding protein, FBP54, on adhesion of group A streptococci to human buccal cells and HEp-2 tissue culture cells. Infect Immun **1996** Jul;64(7):2415-9.
227. Delvecchio A, Currie BJ, McArthur JD, Walker MJ, Sriprakash KS. Streptococcus pyogenes prtFII, but not sfbI, sfbII or fbp54, is represented more frequently among invasive-disease isolates of tropical Australia. Epidemiol Infect **2002** Jun;128(3):391-6.
228. Kawabata S, Kunitomo E, Terao Y, et al. Systemic and mucosal immunizations with fibronectin-binding protein FBP54 induce protective immune responses against Streptococcus pyogenes challenge in mice. Infect Immun **2001** Feb;69(2):924-30.
229. Talay SR, Valentin-Weigand P, Timmis KN, Chhatwal GS. Domain structure and conserved epitopes of Sfb protein, the fibronectin-binding adhesin of Streptococcus pyogenes. Mol Microbiol **1994** Aug;13(3):531-9.

230. Kreikemeyer B, Talay SR, Chhatwal GS. Characterization of a novel fibronectin-binding surface protein in group A streptococci. Mol Microbiol **1995** Jul;17(1):137-45.
231. Lukomski S, Burns EH, Jr., Wyde PR, et al. Genetic inactivation of an extracellular cysteine protease (SpeB) expressed by Streptococcus pyogenes decreases resistance to phagocytosis and dissemination to organs. Infect Immun **1998** Feb;66(2):771-6.
232. Norrby-Teglund A, Newton D, Kotb M, Holm SE, Norgren M. Superantigenic properties of the group A streptococcal exotoxin SpeF (MF). Infect Immun **1994** Dec;62(12):5227-33.
233. Proft T, Moffatt SL, Berkahn CJ, Fraser JD. Identification and characterization of novel superantigens from Streptococcus pyogenes. J Exp Med **1999** Jan 4;189(1):89-102.
234. Sachse S, Seidel P, Gerlach D, et al. Superantigen-like gene(s) in human pathogenic Streptococcus dysgalactiae, subsp equisimilis: genomic localisation of the gene encoding streptococcal pyrogenic exotoxin G (speG(dys)). FEMS Immunol Med Microbiol **2002** Oct 11;34(2):159-67.
235. Igwe EI, Shewmaker PL, Facklam RR, Farley MM, van Beneden C, Beall B. Identification of superantigen genes speM, ssa, and smeZ in invasive strains of beta-hemolytic group C and G streptococci recovered from humans. FEMS Microbiol Lett **2003** Dec 12;229(2):259-64.
236. Broudy TB, Fischetti VA. In vivo lysogenic conversion of Tox(-) Streptococcus pyogenes to Tox(+) with Lysogenic Streptococci or free phage. Infect Immun **2003** Jul;71(7):3782-6.
237. Banks DJ, Lei B, Musser JM. Prophage induction and expression of prophage-encoded virulence factors in group A Streptococcus serotype M3 strain MGAS315. Infect Immun **2003** Dec;71(12):7079-86.
238. Sitkiewicz I, Nagiec M.J., Sumby p., Butler S.D., Bently C.C., Musser J.M.,. Emergance of a bacterial clone with enhanced virulence by acquisition of a phage encoding a secreted phospholipase A2. Proc Natl Acad Sci **2006**;103:16009-14.
239. Graham MR, Smoot L.M., Lei B., Musser J.M.,. Towards a genome scale understanding of group A Streptococcus pathogenesis. Curr Opin Microbiol **2001**;4:65-70.
240. 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.
241. 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.
242. Musser JM. The next chapter in reverse vaccinology. Nat Biotechnol **2006** Feb;24(2):157-8.
243. Koshi G, Thangavelu, C. P., Brahmadathan, K. N.,. The reliability and rapidity of the coagglutination technic and its comparison with precipitin technics in the grouping of Streptococci. American Journal of Clinical Pathology **1979**;71(6):709-12.
244. Jose JJ, Brahmadathan KN. Evaluation of simplified DNA extraction methods for *emm* typing of group A streptococci. Indian J Med Microbiol **2006** Apr;24(2):127-30.

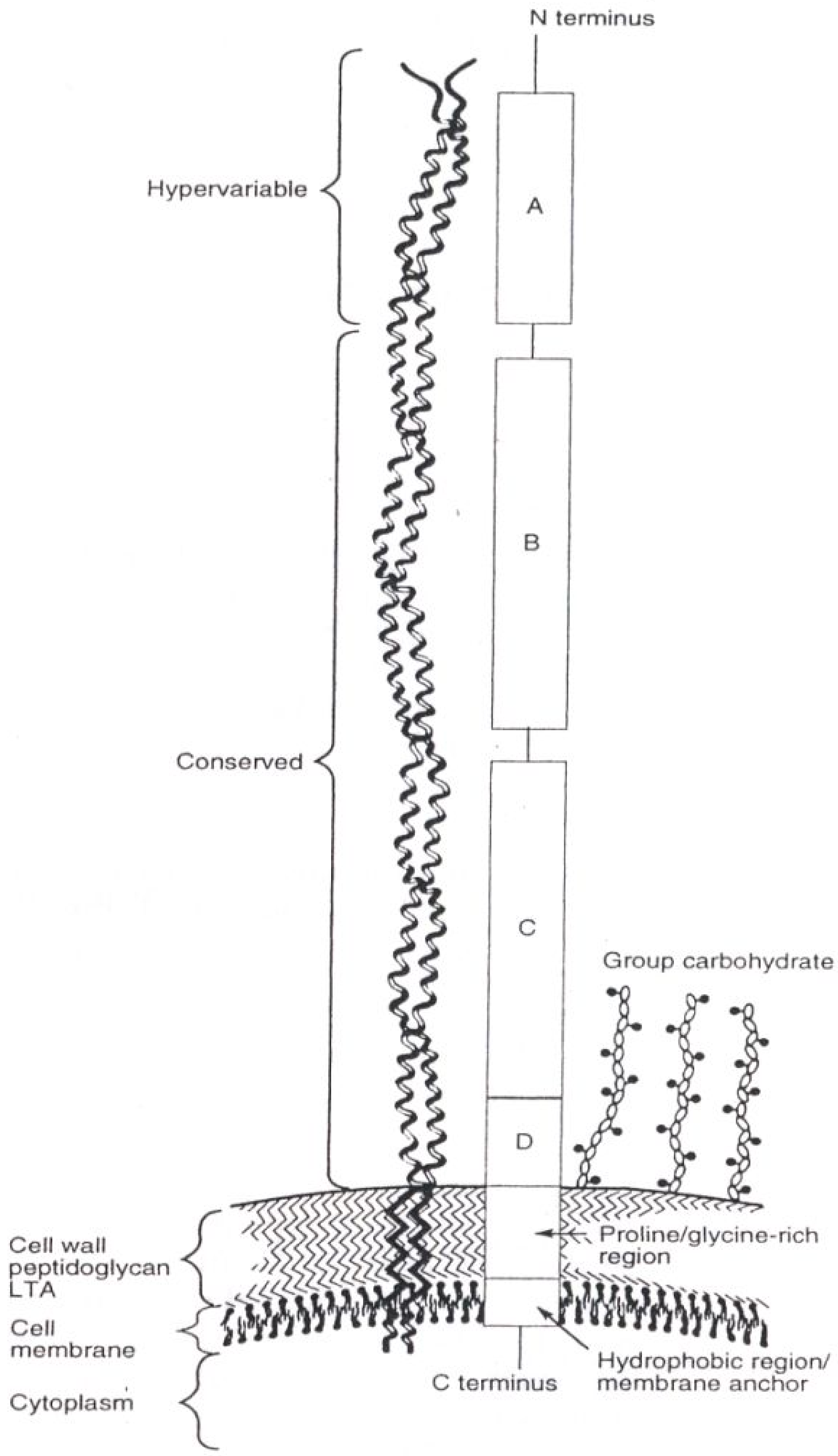
245. Bessen DE, Izzo MW, Fiorentino TR, Caringal RM, Hollingshead SK, Beall B. Genetic linkage of exotoxin alleles and *emm* gene markers for tissue tropism in group A streptococci. J Infect Dis **1999** Mar;179(3):627-36.
246. Tamura K DJ, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution **2007**;24:1596-9.
247. Matsumoto M, Hoe NP, Liu M, et al. Intrahost sequence variation in the streptococcal inhibitor of complement gene in patients with human pharyngitis. J Infect Dis **2003** Feb 15;187(4):604-12.
248. McNeil SA, Halperin SA, Langley JM, et al. Safety and immunogenicity of 26-valent group a streptococcus vaccine in healthy adult volunteers. Clin Infect Dis **2005** Oct 15;41(8):1114-22.
249. Rogers S, Commons R, Danchin MH, et al. Strain prevalence, rather than innate virulence potential, is the major factor responsible for an increase in serious group A streptococcus infections. J Infect Dis **2007** Jun 1;195(11):1625-33.
250. Sagar V, Kumar R, Ganguly NK, Chakraborti A. Comparative analysis of *emm* type pattern of Group A Streptococcus throat and skin isolates from India and their association with closely related SIC, a streptococcal virulence factor. BMC Microbiol **2008**;8:150.
251. Sagar V, Bakshi DK, Nandi S, Ganguly NK, Kumar R, Chakraborti A. Molecular heterogeneity among north Indian isolates of Group A Streptococcus. Lett Appl Microbiol **2004**;39(1):84-8.
252. Menon T, Whatmore AM, Srivani S, Kumar MP, Anbumani N, Rajaji S. *EMM* types of streptococcus pyogenes in Chennai. Indian J Med Microbiol **2001** 19 (3):161-2.
253. Sindhulina C, Geethalakshmi S, Thenmozhivalli PR, Jose JM, Brahmadathan KN. Bacteriological and molecular studies of group A streptococcal pharyngitis in a south Indian hospital. Indian J Med Microbiol **2008** Apr-Jun;26(2):197-8.
254. Jing HB, Ning BA, Hao HJ, et al. Epidemiological analysis of group A streptococci recovered from patients in China. J Med Microbiol **2006** Aug;55(Pt 8):1101-7.
255. Hoe NP, Fullerton KE, Liu M, et al. Molecular genetic analysis of 675 group A streptococcus isolates collected in a carrier study at Lackland Air Force Base, San Antonio, Texas. J Infect Dis **2003** Sep 15;188(6):818-27.
256. Li Z, Sakota V, Jackson D, Franklin AR, Beall B. Array of M protein gene subtypes in 1064 recent invasive group A streptococcus isolates recovered from the active bacterial core surveillance. J Infect Dis **2003** Nov 15;188(10):1587-92.
257. Tyrrell GJ, Lovgren M, Forwick B, Hoe NP, Musser JM, Talbot JA. M types of group a streptococcal isolates submitted to the National Centre for Streptococcus (Canada) from 1993 to 1999. J Clin Microbiol **2002** Dec;40(12):4466-71.
258. 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.

259. Moses AE, Hidalgo-Grass C, Dan-Goor M, et al. *emm* typing of M nontypeable invasive group A streptococcal isolates in Israel. J Clin Microbiol **2003** Oct;41(10):4655-9.
260. Tanaka D, Gyobu Y, Kodama H, et al. *emm* Typing of group A streptococcus clinical isolates: identification of dominant types for throat and skin isolates. Microbiol Immunol **2002**;46(7):419-23.
261. Smeesters PR VA, Campos D., Aguiar E. D., Deyi V.Y.M., Melderen L.V.,. Differences between Belgian and Brazilian Group A Streptococcus Epidemiologic Landscape. PLoS One **2006**;10(1):1371.
262. Musser JM, Gray BM, Schlievert PM, Pichichero ME. Streptococcus pyogenes pharyngitis: characterization of strains by multilocus enzyme genotype, M and T protein serotype, and pyrogenic exotoxin gene probing. J Clin Microbiol **1992** Mar;30(3):600-3.
263. Zhang F, Su B, Zhang YP, Jin L. Genetic studies of human diversity in East Asia. Philos Trans R Soc Lond B Biol Sci **2007** Jun 29;362(1482):987-95.
264. Ouyang C, Smith DD, Krontiris TG. Evolutionary signatures of common human cis-regulatory haplotypes. PLoS ONE **2008**;3(10):e3362.
265. Smeesters PR, Vergison A., Campos D., Aguiar E. D., Deyi V.Y.M., Melderen L.V.,. Differences between Belgian and Brazilian Group A Streptococcus Epidemiologic Landscape. PLoS One **2006**;10(1):1371.
266. Loubinoux J, Florent M, Merad B, Collobert G, Bouvet A. Epidemiological markers of group A streptococcal infections in France. Indian J Med Res **2004** May;119 Suppl:152-4.
267. Luca-Harari B, Ekelund K, van der Linden M, Staum-Kaltoft M, Hammerum AM, Jasir A. Clinical and epidemiological aspects of invasive Streptococcus pyogenes infections in Denmark during 2003 and 2004. J Clin Microbiol **2008** Jan;46(1):79-86.
268. Creti R, Imperi M, Baldassarri L, et al. *emm* Types, virulence factors, and antibiotic resistance of invasive Streptococcus pyogenes isolates from Italy: What has changed in 11 years? J Clin Microbiol **2007** Jul;45(7):2249-56.
269. Kaplan EL, Wotton JT, Johnson DR. Dynamic epidemiology of group A streptococcal serotypes associated with pharyngitis. Lancet **2001** Oct 20;358(9290):1334-7.
270. 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.
271. Brandt ER, Yarwood PJ, McMillan DJ, et al. Antibody levels to the class I and II epitopes of the M protein and myosin are related to group A streptococcal exposure in endemic populations. Int Immunol **2001** Oct;13(10):1335-43.
272. Lorino G, Gherardi G, Angeletti S, et al. Molecular characterisation and clonal analysis of group A streptococci causing pharyngitis among paediatric patients in Palermo, Italy. Clin Microbiol Infect **2006** Feb;12(2):189-92.
273. 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.
274. Teixeira LM, Barros RR, Castro AC, et al. Genetic and phenotypic features of Streptococcus pyogenes strains isolated in Brazil that harbor new *emm* sequences. J Clin Microbiol **2001** Sep;39(9):3290-5.
275. Fischetti VA. Streptococcal M protein: molecular design and biological behavior. Clin Microbiol Rev **1989** Jul;2(3):285-314.

276. Dale RC. Post-streptococcal autoimmune disorders of the central nervous system. Dev Med Child Neurol **2005** Nov;47(11):785-91.
277. Bessen D, Fischetti VA. Synthetic peptide vaccine against mucosal colonization by group A streptococci. I. Protection against a heterologous M serotype with shared C repeat region epitopes. J Immunol **1990** Aug 15;145(4):1251-6.
278. Eriksson BK, Villasenor-Sierra A, Norgren M, Stevens DL. Opsonization of T1M1 group A Streptococcus: dynamics of antibody production and strain specificity. Clin Infect Dis **2001** Jan 15;32(2):E24-30.
279. O'Brien KL, Beall B, Barrett NL, et al. Epidemiology of invasive group a streptococcus disease in the United States, 1995-1999. Clin Infect Dis **2002** Aug 1;35(3):268-76.
280. Aziz RK, Edwards RA, Taylor WW, Low DE, McGeer A, Kotb M. Mosaic prophages with horizontally acquired genes account for the emergence and diversification of the globally disseminated M1T1 clone of Streptococcus pyogenes. J Bacteriol **2005** May;187(10):3311-8.
281. Cleary PP, LaPenta D, Vessela R, Lam H, Cue D. A globally disseminated M1 subclone of group A streptococci differs from other subclones by 70 kilobases of prophage DNA and capacity for high-frequency intracellular invasion. Infect Immun **1998** Nov;66(11):5592-7.
282. Condoner FM, Daros J. A., Sole R.V., Elena S.F., . The fittest versus the flattest: Experimenmtal confirmation of the quasispecies effect with subviral pathogens. PLoS Pathog **2006**;2(12):e136.
283. Carapetis J, Gardiner D, Currie B, Mathews JD. Multiple strains of Streptococcus pyogenes in skin sores of aboriginal Australians. J Clin Microbiol **1995** Jun;33(6):1471-2.
284. McDonald ML, Towers, R.J., Andrews, R.M., Benger, N., Currie, B.J., Carapetis, J.R.,. Low rates of Streptococcal Pharyngitis and high rates of pyoderma in Australian aboriginal communities where acute rheumatic fever is hyperendemic. Clin Infect Dis **2006**;15(43 (6)):683 - 9.
285. Bessen DE, Hollingshead SK. Horizontal transfer and mosaic-like *emm* gene structures in group A streptococci. Dev Biol Stand **1995**;85:169-73.
286. Maddison WP, Knowles L. L.,. Inferring phylogeny despite incomplete linage sorting. Syst Biol **2006**;55(1):21-30.
287. Mathur P, Kapil A, Das B, Dhawan B, Dwivedi SN. Invasive beta-haemolytic streptococcal infections in a tertiary care hospital in northern India. J Med Microbiol **2002** Sep;51(9):791-2.
288. Sagar V, Kumar R, Ganguly NK, Menon T, Chakraborti A. DRS is far less divergent than streptococcal inhibitor of complement of group A streptococcus. J Bacteriol **2007** Apr;189(7):2933-5.
289. Ma X, Kikuta H, Ishiguro N, et al. Association of the prtF1 gene (encoding fibronectin-binding protein F1) and the sic gene (encoding the streptococcal inhibitor of complement) with *emm* types of group A streptococci isolated from Japanese children with pharyngitis. J Clin Microbiol **2002** Oct;40(10):3835-7.
290. Towers RJ, Fagan PK, Talay SR, et al. Evolution of sfbI encoding streptococcal fibronectin-binding protein I: horizontal genetic transfer and gene mosaic structure. J Clin Microbiol **2003** Dec;41(12):5398-406.
291. McMillan DJ, Beiko,R. G, Geffers, R. , Jan Buer, Schouls, L. M, Vlaminckx, B. J. M. , Wannet, W. J. B. , Sriprakash, K. S. , and Chhatwal, G. S. Genes for

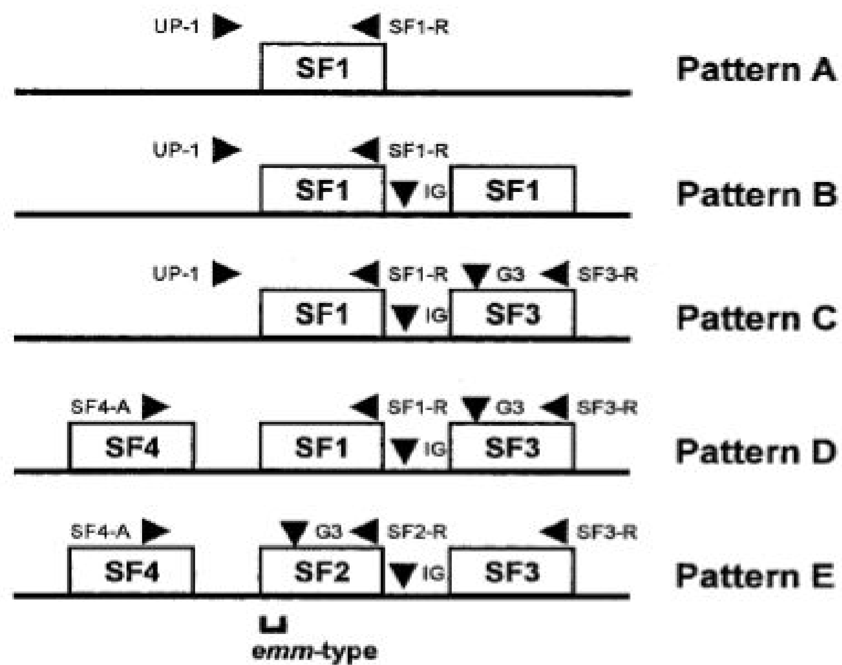
the majority of Group A Streptococcal virulence factors and extracellular surface proteins do not confer an increased propensity to cause invasive disease. Clinical Infectious Diseases **2006**;43:884-91.

Fig. 1: Model of the Streptococcal M protein.



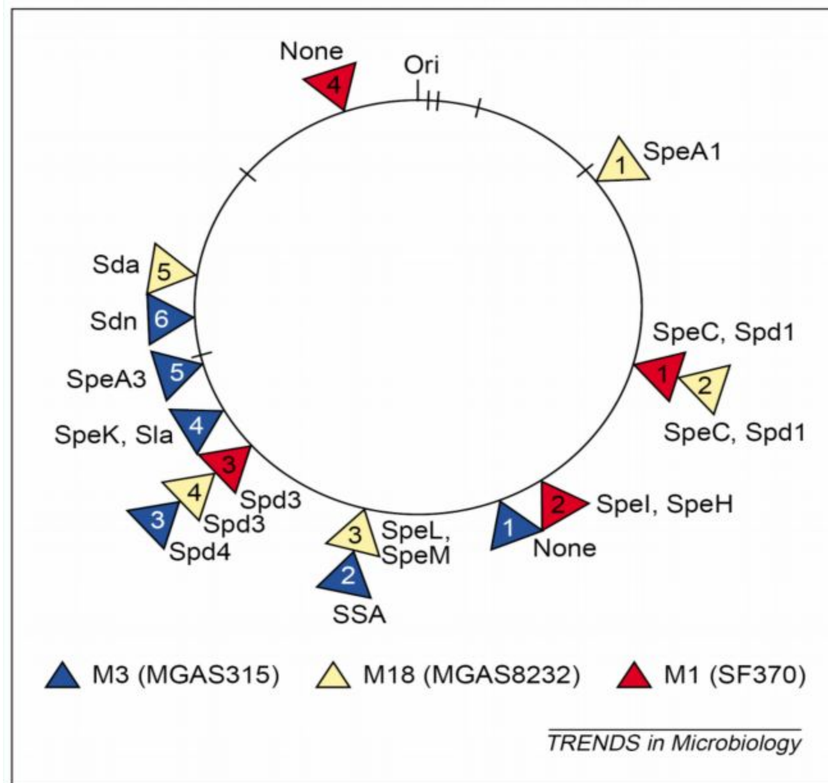
Johnson D.R., et al.[58].

Fig 2: The *emm* chromosomal patterns. Content and relative arrangement of *emm* subfamily (SF) genes (SF1–SF4), on the basis of phylogenetic divergence at the 3′ ends of *emm* and *emm*-like genes, is shown. These 5 *emm* patterns (A–E) account for all *emm* gene arrangements among the group A streptococci (GAS) isolates under study in this report. Intergenic distances are typically 0.2–0.25 kb. The 5′ portion of the central *emm* gene that contains the determinants for *emm* sequence type (i.e., *emm* type) is indicated. Approximate positions for oligonucleotide primer sites, used for polymerase chain reaction–based mapping are also shown.



Bessen et al. [30].

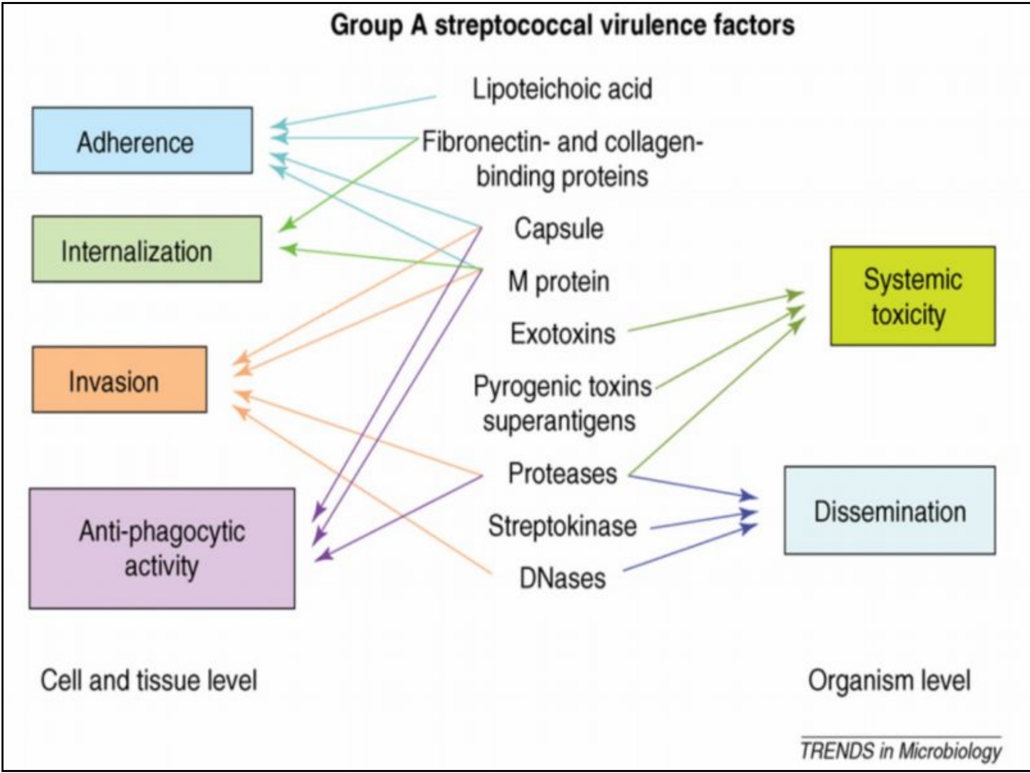
Fig. 3: Schematic representation of various phage associated virulence factors identified from the genome sequence of GAS serotypes M1, M3, M18.



Note: The location of prophages in the genomes of group A Streptococcus serotype M3 (strain MGAS315), M18 (strain MGAS8232) and M1 (strain SF370). Black lines represent rRNA gene locations. Abbreviations: Spe, streptococcal pyrogenic exotoxin; Spd, Streptococcus pyogenes DNase; Sda, streptodornase α; Sdn, streptodornase; Sla, streptococcal phospholipase A2; SSA, streptococcal superantigen. Only proven or putative secreted virulence factors encoded by genes located at the phage ends are listed

Banks D.J. et al. [47].

Fig. 4: GAS virulence factors interact with the host at many levels. GAS has an arsenal of virulence factors at its disposal that enable it to colonize and thrive successfully in the host. At the cell and tissue level, these factors contribute to the pathogenicity of GAS by mediating adherence to host cells, by promoting internalization and invasion, and by evading phagocytosis. At the organism level, these factors are involved in facilitating dissemination throughout the host and can induce systemic toxicity. Importantly, many of the known streptococcal virulence factors function at several stages of infection.



Tart A.H., et al., [6].

Fig. 5: Distribution of 47 *emm* types of PT GAS isolates (*n = 116*).

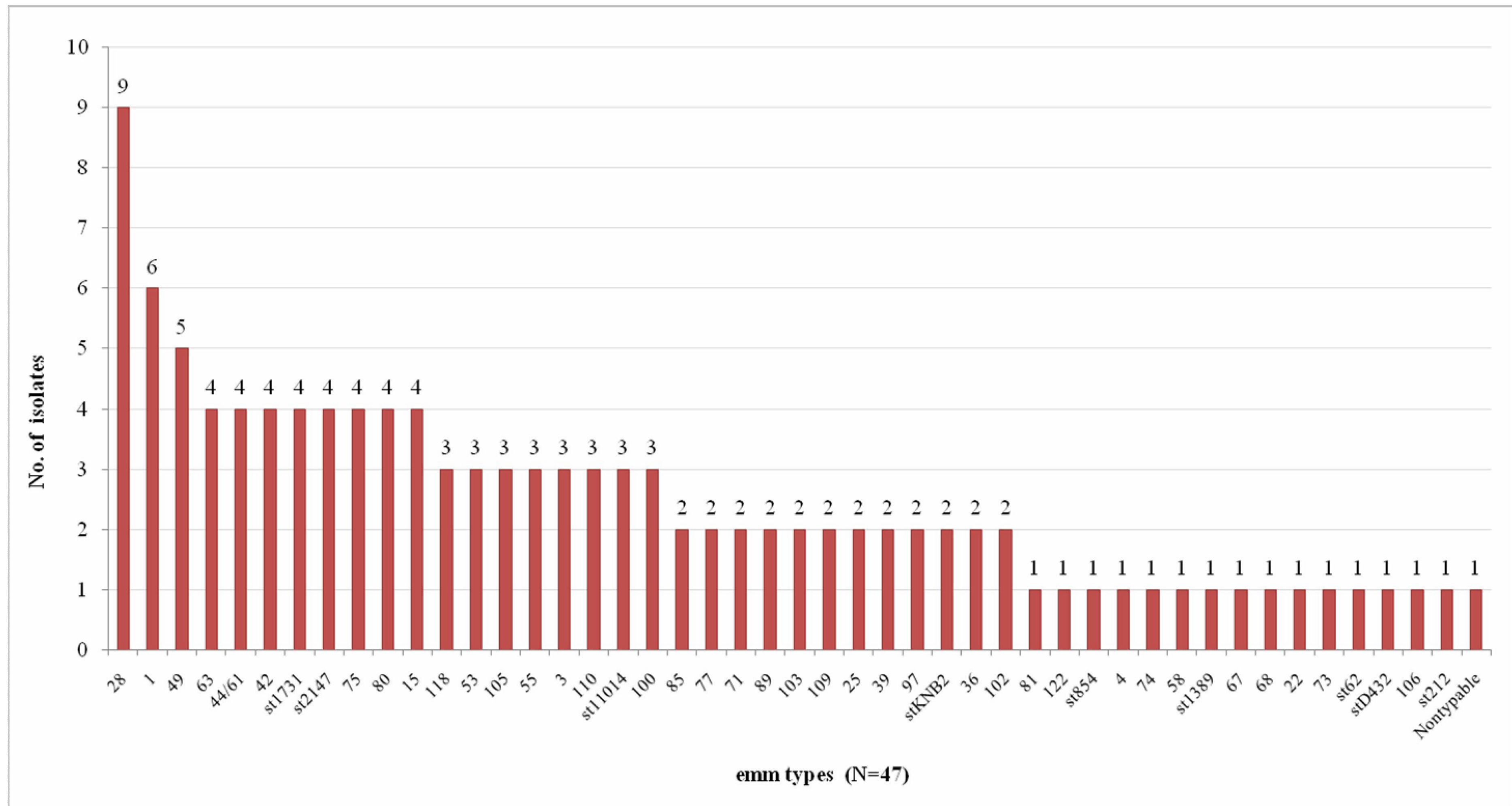


Fig. 6: Distribution of 62 *emm* types among SK GAS isolates (*n* = *276*)

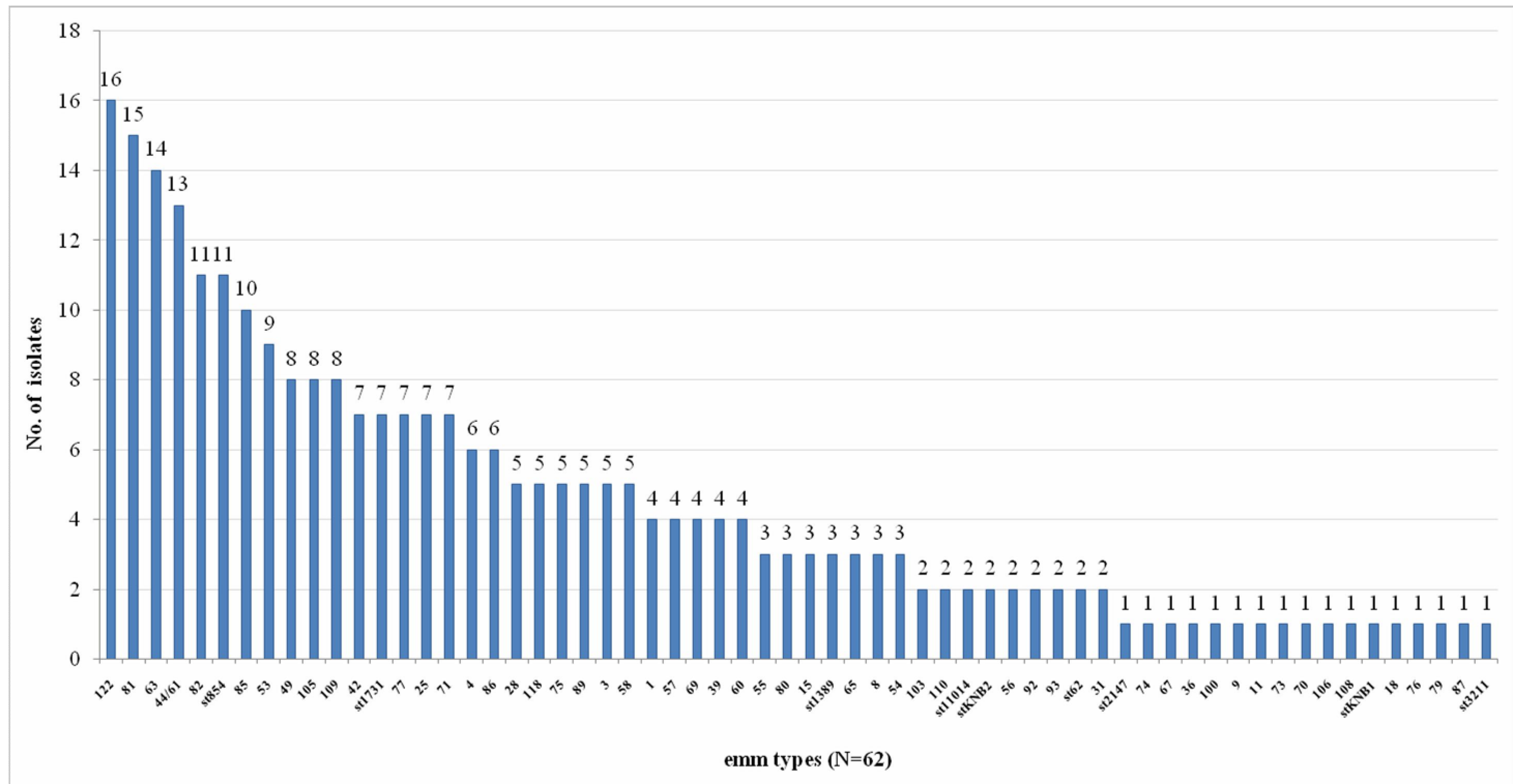


Fig. 7: Distribution of 67 *emm* types of normal throat (NT) GAS isolates (n = *306*)

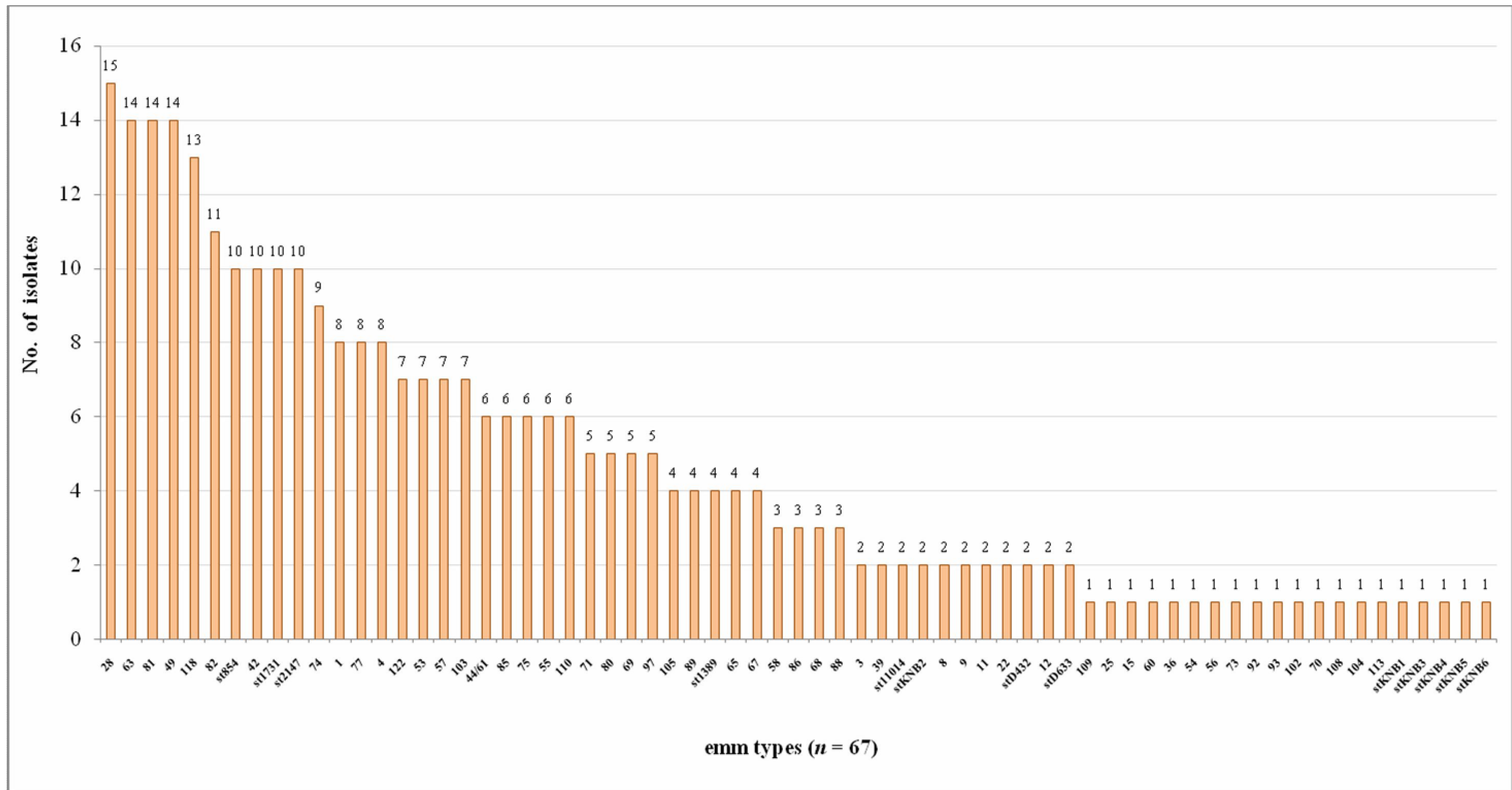


Fig. 8: Predominant *emm* types of PT, SK and NT.

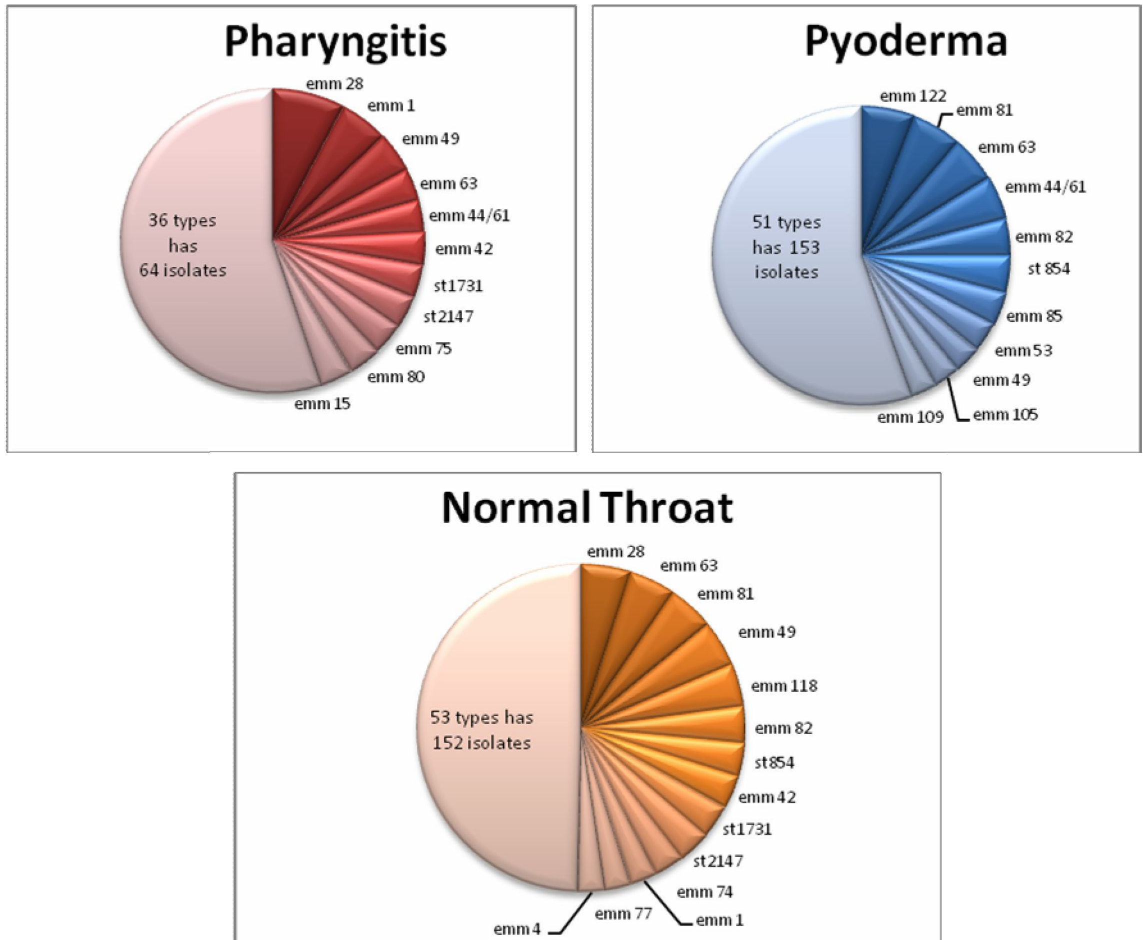


Fig. 9: Comparison of predominant *emm* types among PT, SK and NT isolates.

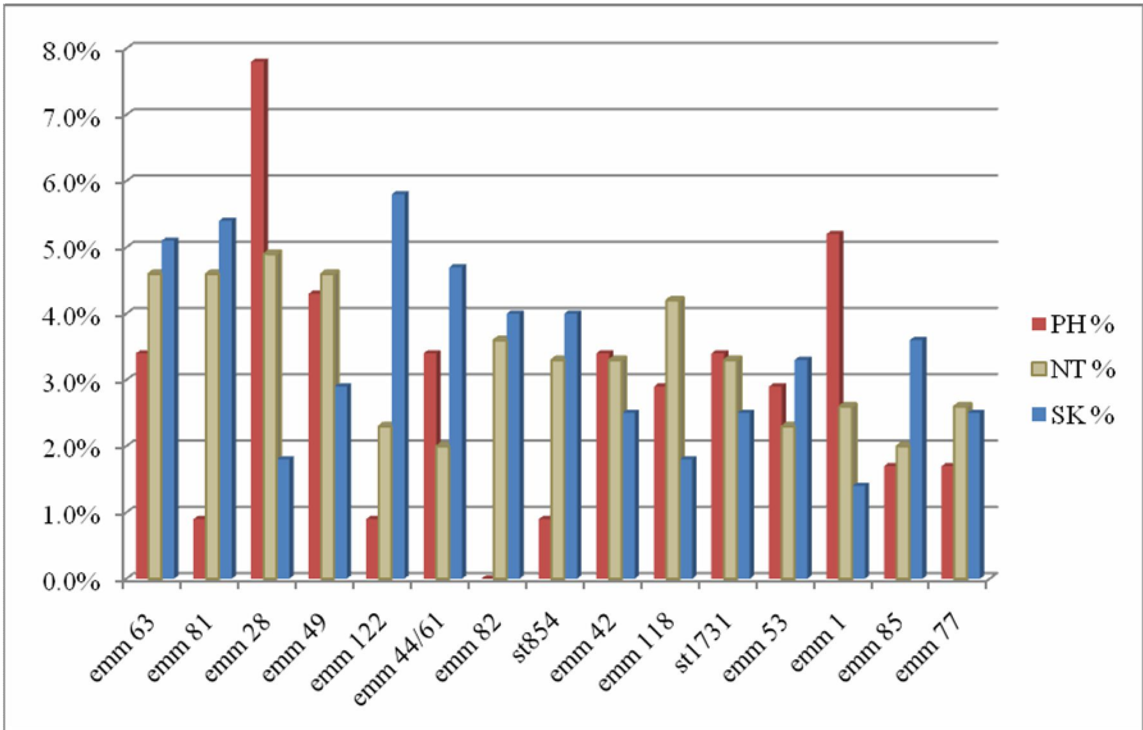


Fig. 14: Monthly distribution of *emm* types among PT isolates.

Sl.No:	*emm* type	Mar & Apr 02	Apr & Aug 02	Aug & Sep 02	Sep & Oct 02	Nov & Dec 02	Dec 02-Feb 03	Mar & Apr 03	Jun & Jul 03	Jul & Aug 03	Aug & Sep 03	Sep & Oct 03	Oct & Nov 03	Nov 03-Jan 04	Jan & Feb 04	Feb & Mar 04	
1	**58**	*															1
2	**36**	**															2
3	**97**	**															2
4	**25**	**															2
5	**1**	******															6
6	**st212**	*															1
7	**80**	**	**														4
8	**67**		*														1
9	**68**		*														1
10	**106**		*														1
11	**39**		*	*													2
12	**85**	*		*													2
13	**103**			**													2
14	**73**				*												1
15	**55**				***												3
16	**3**			*	**												3
17	**st2147**	*	*		**												4
18	**102**	*			*												2
19	**105**	*			**												3
20	**89**				*	*											2
21	**109**				*	*											2
22	**77**	*					*										2
23	**81**						*										1
24	**st1731**	**					**										4
25	**15**					**	*	*									4
26	**st11014**					*	*	*									3
27	**28**	******					*	**									9
28	**49**	*		**	*				*								5
29	**63**	**				*			*								4
30	**42**			*			**		*								4
31	**75**					**		*	*								4
32	**118**					**			*								3
33	**71**					*						*					2
34	**53**	*							*				*				3
35	**110**		**												*		3
36	**stKNB2**		*													*	2
37	**74**									*							1
38	**22**											*					1
39	**stD432**													*			1
40	**122**													*			1
41	**44/61**													***	*		4
42	**4**														*		1
43	**st854**														*		1
44	**st62**														*		1
45	**100**															***	3
46	**st1389**															*	1
47	**Non typable**															*	1
		33	**10**	**8**	**14**	**11**	**9**	**5**	**6**	**1**	0	**2**	**1**	**5**	**5**	6	116

Note: The redish colour indicates the presence of those *emm* types in the first half of the study. The orange colour indicates the presence of those *emm* type in the first survey of the second half also. Blue colour is the types that is present in the overlapping months of the two halves in the middle of the study. Tan colour is distributed widely. The green colour is presence of *emm* types in the latter half of the study.

Fig. 15: Monthly distribution of *emm* types among SK isolates.

Sl.No:	*emm* type	Mar & Apr 02	Apr & Aug 02	Aug & Sep 02	Sep & Oct 02	Nov & Dec 02	Dec 02-Feb 03	Mar & Apr 03	Jun & Jul 03	Jul & Aug 03	Aug & Sep 03	Sep & Oct 03	Oct & Nov 03	Nov 03-Jan 04	Jan & Feb 04	Feb & Mar 04	Total
1	36	*															1
2	st2147	*															1
3	54	***															3
4	103		**														2
5	67		*														1
6	stKNB1		*														1
7	79		*														1
8	106		*														1
9	39	***		*													4
10	1	**	*		*												4
11	55			*	**												3
12	73				*												1
13	st3211				*												1
14	57		**	*		*											4
15	st11014			*		*											2
16	87					*											1
17	77	*****	*			*											7
18	89			*	**	**											5
19	70						*										1
20	93						**										2
21	3			**	*	*	*										5
22	110		*				*										2
23	28	****						*									5
24	80	**						*									3
25	105	**	****		*			*									8
26	st1731	*				****	*	*									7
27	58		***			*			*								5
28	69					**			*	*							4
29	st854				**	*	*	******		*							11
30	75		**	*		*					*						5
31	60		*	*						*	*						4
32	92			*							*						2
33	st1389		**								*						3
34	118				*			*	**			*					5
35	25	****									**	*					7
36	53	**	***						***			*					9
37	4		*		*	*		*			*	*					6
38	109					*				****	***	*					8
39	82				*		*	**	**	*	***	*					11
40	63		**		*				**		******	**	*				14
41	42				***	*	*			*			*				7
42	122							**		*	******	******		*			16
43	44/61			*				*	***	*	*		**	****			13
44	15					*	*								*		3
45	81		***	***	*	*	***		*			*	*			*	16
46	56						*									*	2
47	71		*		*	*	*			*		*				*	7
48	86		**												*	***	6
49	49	**	*		*			**							*	*	8
50	85	**	**				*							*	**	**	10
51	18								*								1
52	108								*								1
53	9									*							1
54	11									*							1
55	8									**							2
56	stKNB2										*	*					2
57	31											**					2
58	76											*					1
59	74															*	1
60	100															*	1
61	st62													*		*	2
62	65								*	*						*	3
	Total	**34**	**38**	**14**	**21**	**22**	**16**	**19**	**18**	**17**	**27**	**20**	**5**	**7**	**5**	**13**	**276**

Note: Blue- *emm* types in first half only, Tan-distributed all over, Green- latter half only

Fig. 16: Monthly distribution of *emm* types among NT isolates.

Sl.No:	*emm* type	Mar & Apr 02	Apr & Aug 02	Aug & Sep 02	Sep & Oct 02	Nov & Dec 02	Dec 02-Feb 03	Mar & Apr 03	Jun & Jul 03	Jul & Aug 03	Aug & Sep 03	Sep & Oct 03	Oct & Nov 03	Nov 03-Jan 04	Jan & Feb 04	Feb & Mar 04	Total
1	25	*															1
2	36	*															1
3	54	*															1
4	104	*															1
5	stD633	**															2
6	stKNB4		*														1
7	1	****	****														8
8	60			*													1
9	stKNB1			*													1
10	39		*	*													2
11	102				*												1
12	109				*												1
13	73				*												1
14	12			*	*												2
15	89			**	**												4
16	58		*	*	*												3
17	55	**	*	*	**												6
18	70					*											1
19	93					*											1
20	56					*											1
21	3			*		*											2
22	15						*										1
23	67			*		**	*										4
24	st11014	*					*										2
25	108							*									1
26	105	*		*		*		*									4
27	st2147	*	**	**	**	*		**									10
28	69		*				**		**								5
29	42				***	***	*	**	*								10
30	80	***	*						*								5
31	103			**	***			*		*							7
32	28	*****	***		***	**		*		*							15
33	st1389			*	*	*					*						4
34	4	*	*		*				****			*					8
35	22			*								*					2
36	53	*	****						*			*					7
37	57			*	**			**				**					7
38	71				*	**				*		*					5
39	st1731	**			*	**	****					*					10
40	77	***	*		*		*		*			*					8
41	82	**	*		*			*	**	**	*	*					11
42	118	*			**			*	***	**	**	**					13
43	68					*				*			*				3
44	49					**		**	*****	**	**		*				14
45	110		*	*						*			**	*			6
46	97	**							*					**			5
47	88					*			*						*		3
48	63				*	*	**	*	**	*		**	*		***		14
49	st854						*	*	*	***			*		**	*	10
50	75				*	****										*	6
51	74				**	**	*		**		*					*	9
52	81		*	*			****	**	**	*	*	*				*	14
53	85	****	*													*	6
54	stKNB5								*								1
55	stKNB6								*								1
56	113									*							1
57	stKNB3									*							1
58	9									*	*						2
59	11								*			*					2
60	92														*		1
61	stKNB2													*	*		2
62	8											*			*		2
63	122										*	***	*	*	*		7
64	44/61													**	***	*	6
65	stD432											*				*	2
66	65								*							***	4
67	86								*							**	3
		39	**25**	**20**	**34**	**29**	**19**	**18**	**34**	**19**	**10**	**20**	**7**	**7**	**13**	**12**	**306**

Note: Blue- *emm* types in first half only, Tan-distributed all over, Orange- latter half only

. **Fig. 10:** Patterns type distribution among PT isolates

VAP 51	18-03-02	08-04-02	21-10-02	02-21-02	06-01-02	27-01-03	21-04-03	05-08-03	16-03-04		
	28	28	28	28	28	28	28	28	86.2		
VAP 110	13-03-02	10-04-02	17-04-02	19-06-02	10-10-02	06-11-02	13-11-02	08-01-03	13-11-03	26-02-04	10-03-04
	28	28	28	103	118	89	118	st1731.1	44/61	86.2	74
VAP 109	21-08-02	23-10-02									
	st2147	st2147									
VAP 191	12-09-02										
	49.4										

Fig. 11: Pattern of type distribution among SK isolates

VAP 139	19-06-02	07-08-02	20-08-02	29-08-02	17-09-02	26-11-02	03-12-02	10-12-02	10-02-03	25-02-03	26-06-03	24-07-03
	103	75.1	81.2	3.22	3.22	69.1	69.1	69.1	st854.1	81.2	65.1	8
VAP 195	23-08-02	12-09-02	18-09-02	10-10-02	18-10-02	11-11-02	04-04-03					
	89	89	st3211	89	118	89	st854.1					
VAP 150	04-09-03											
	44/61											

Fig. 12: Patterns of type distribution among NT isolates

VAP 57	15-03-02	04-10-02	10-12-02	08-04-03	18-06-03	16-07-03	29-01-04	26-02-04
	77	st1389.1	st854.1	st854.1	st854.1	st854.1	st854.1	st854.1
VAP 78	18-06-02	29-08-02	07-01-03	05-02-04				
	80	89	81.1	8				
VAP 87	20-08-02							
	81.2							

Fig. 13: Examples of transmission of same *emm* type from one site to other

VAP 194	16-03-02	19-08-02	23-08-02	06-09-02	12-09-02	18-09-02	18-09-02	29-11-02	05-12-02
	118	57	57	57	57	57	57	81.2	81.2
VAP 190	05-04-02	10-10-02	18-10-02	11-11-02	29-11-02				
	25.2	71	71	71	71				
VAP 176	14-03-02	26-03-02	26-03-02	04-04-02	22-08-02	28-08-02	23-10-02	11-02-03	17-04-03
	85	st1731	st1731	st1731	44/61	44/61	42.1	st11014	44/61
VAP 4	01-04-02	16-04-02	17-06-02	13-08-02	21-04-03	15-07-03	15-07-03	12-08-03	21-08-03
	53.1	85	110	110	28.5	82.1	82.1	82.1	63

Note: Red colour denotes pharyngitis, blue denotes impetigo and black denotes throat colonization. VAP # is the student's code. First row is the date of sample collection and the second row is the *emm* type. Date of sample collection should be read as day/month/year.

Fig. 17: Phylogenetic tree of *emm* gene for representative types.

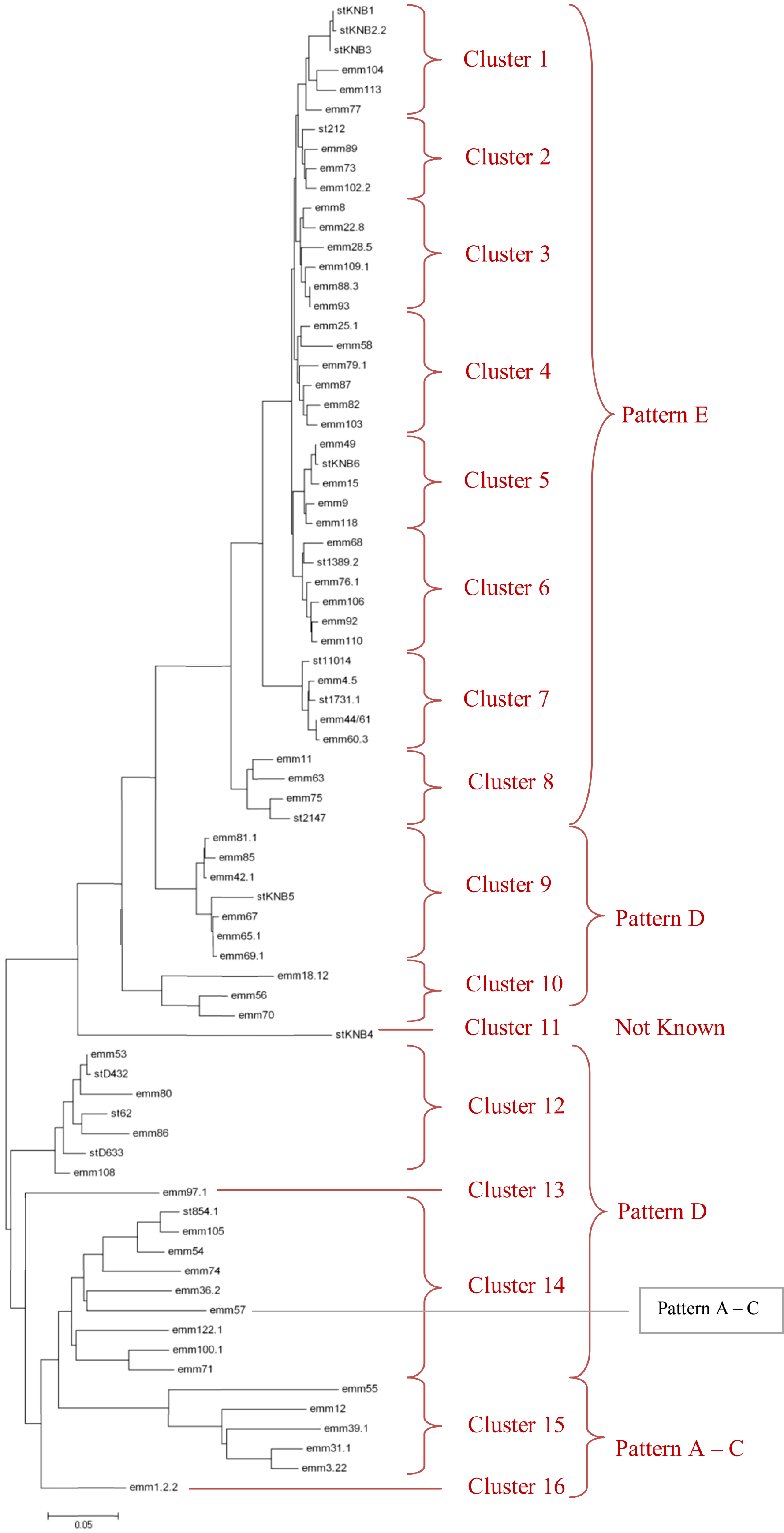


Fig 18: The matrix of the Mean distance between the Phylogenetic clusters.

Cluster	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0	.04	.04	.04	.04	.04	.09	.11	.18	.22	.34	.30	.32	.35	.46	.31
2	.04	0	.03	.03	.04	.03	.07	.10	.17	.23	.35	.30	.33	.35	.45	.31
3	.04	.03	0	.03	.04	.04	.07	.10	.16	.22	.34	.29	.32	.35	.45	.31
4	.04	.03	.03	0	.03	.04	.07	.10	.17	.23	.35	.30	.32	.36	.45	.31
5	.04	.04	.04	.03	0	.03	.08	.09	.16	.23	.35	.30	.32	.36	.44	.31
6	.04	.03	.04	.04	.03	0	.07	.09	.16	.23	.35	.30	.32	.35	.45	.30
7	.09	.07	.07	.07	.08	.07	0	.09	.16	.22	.36	.29	.35	.35	.46	.32
8	.11	.10	.10	.10	.09	.09	.09	0	.10	.23	.37	.30	.38	.37	.46	.32
9	.18	.17	.16	.17	.16	.16	.16	.10	0	.15	.31	.22	.28	.27	.35	.24
10	.22	.23	.22	.23	.23	.23	.22	.23	.15	0	.35	.15	.27	.22	.35	.21
11	.34	.35	.34	.35	.35	.35	.36	.37	.31	.35	0	.33	.26	.37	.45	.35
12	.30	.30	.29	.30	.30	.30	.29	.30	.22	.15	.33	0	.17	.16	.25	.12
13	.32	.33	.32	.32	.32	.32	.35	.38	.28	.27	.15	.33	0	.20	.29	.18
14	.35	.35	.35	.36	.36	.35	.35	.37	.27	.22	.27	.26	.17	0	.27	.14
15	.46	.45	.45	.45	.44	.45	.46	.46	.35	.35	.45	.25	.29	.27	0	.27
16	.31	.31	.31	.31	.31	.30	.32	.32	.24	.21	.35	.12	.18	.14	.27	0

Note: The tan shade denotes *emm* family pattern E, Blue denotes pattern D and Red denotes pattern A-C.

Fig. 19: Conserved region sequences of *emm* types.

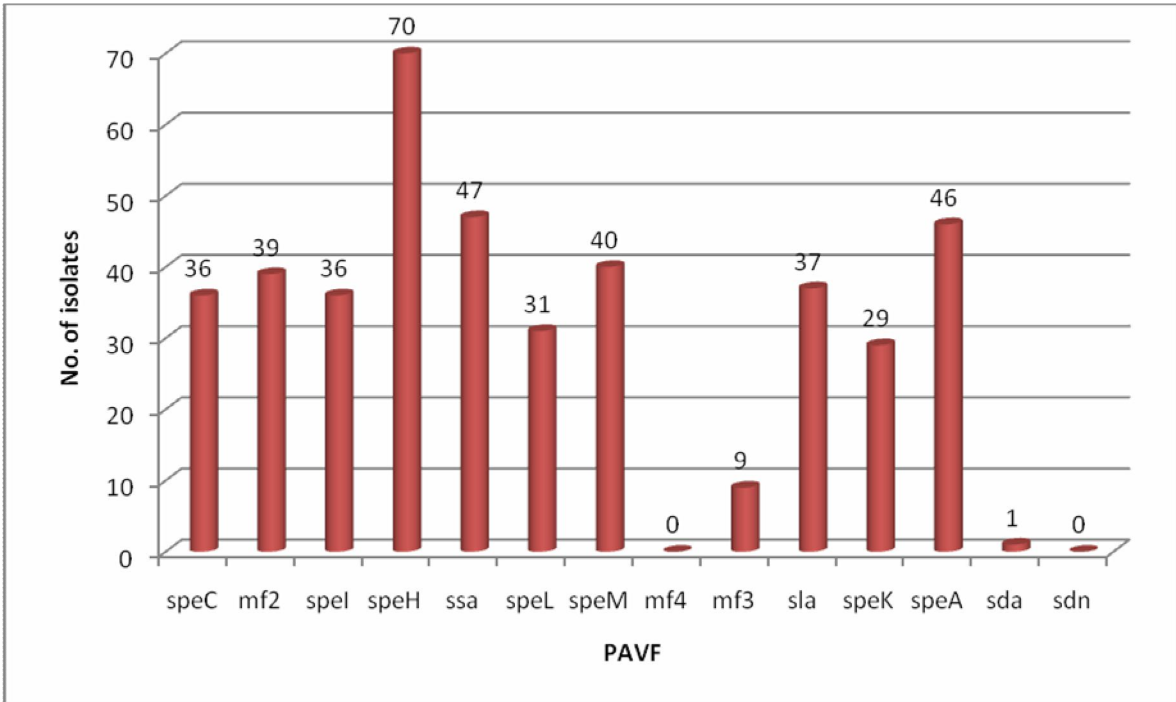
```
emm68           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 962
st1389.2        AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 931
emm104          AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 979
emm106          AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 982
emm110          AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 943
emm92           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 973
emm76.1         AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 985
stKNB1          AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 1045
stKNB2.2        AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 1054
stKNB3          AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 1015
emm25.1         AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 982
emm87           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 991
emm103          AAGCAAAAGCTCTTAAAGAGCAATTGACTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 979
emm79.1         AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 988
emm58           AAGCAAAAGCTCTTAAAGAACAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 1012
emm82           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 1015
emm113          AAGCAAAAGCTCTTAAAGAGCAATTAGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 964
emm49           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 962
stKNB6          AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 907
emm118          AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 916
emm9            AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 925
emm15           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 928
emm88.3         AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 994
emm93           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 994
emm8            AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCAAAACTAAAAG 997
emm22.8         AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCAAAACTAAAAG 994
emm109.1        AAGCAAAAGCTCTTAAAGAGCAATTAGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 979
emm28.5         AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 994
emm73           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 934
emm102.2        AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 895
st212           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 916
emm77           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 886
emm89           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 895
emm44/61        AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 1000
emm60.3         AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 1021
st1731.1        AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 958
st11014         AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 979
emm4.5          AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 973
emm11           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 952
emm63           AAGCAAGAGCTCTTAAAGAGCAATTAGCTAAACAAGCTGAAGAACTTGCTAAACTAAAAG 922
emm75           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 955
st2147          AAGCAAAAGCTCTTAAAGAGCAATTAGCTAAACAAGCTGAAGAACTTGCCAAACTAAAAG 946
                             <--- Class II ---              ---- ClassI ---->
emm42.1         AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCAAAACTAAGAG 970
emm85           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCAAAACTAAGAG 943
emm65.1         AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCAAAACTAAGAG 940
emm69.1         AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCAAAACTAAGAG 940
emm67           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCAAAACTAAGAG 937
stKNB5          AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCAAAACTAAGAG 811
emm81.1         AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCAAAACTAAGAG 934
emm18.12        AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1042
emm56           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1045
emm70           AAGCAAAAGCTCTTAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1018

stKNB4          AAGGAAAAGCCCTCAAAGAACAATTAGCAAAACAAACTGAAGAGCTTGCAAAACTAAGAG 876
emm53           AAGCAAAAGCACTCAAAGAACAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1015
stD432          AAGCAAAAGCACTCAAAGAACAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1035
emm80           CAGCAAAAGCACTCAAAGAACAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAAAG 1027
emm108          AAGCAAAAGCCCTCAGAGAACAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1060
stD633          AGGCAAAAGCACTCAAAGAACAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1016
st62            AGGCAAAAGCACTCAAAGAAAAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1022
emm86           AAGCAAAAGCACTCAAAGAACAATTAGCGAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1048
emm97.1         AAGCAAAAGCACTCAAAGAAAAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1021
emm3.22         AAGCAAAAGCACTCAAAGAACAATTAGCGAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1480
emm31.1         AAGCAAAAGCACTCAAAGAAAAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1417
emm12           AAGCAAAAGCCCTCAAAGAACAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1459
emm39.1         AAGCAAAAGCACTCAAAGAACAATTAGCGAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1291
emm55           AAGCAAAAGCACTCAAAGAACAATTAGCGAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1387
emm57           AAGCAAAAGCACTCAAAGAACAATTAGCGAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1195
emm71           AAGCAAAAGCACTCAAAGAACAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1160
emm1.2.2        AAGCAAAAGCACTCAAAGAACAATTAGCGAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1207
emm74           AAGCAAAAGCACTCAAAGAACAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1108
emm122.1        AAGCAAAAGCACTCAAAGAACAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1108
emm105          AAGCAAAAGCACTCAAAGAACAATTAGCGAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1041
st854.1         AAGCAAAAGCACTCAAAGAACAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1180
emm54           AAGCAAAAGCACTCAAAGCACAATTAGCAAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1063
emm100.1        AAGCAAAAGCACTCAAAGAGCAATTGGCTAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1084
emm36.2         AAGCAAAAGCACTCAAAGAAAAATTAGCGAAACAAGCTGAAGAACTTGCAAAACTAAGAG 1021
                  * ** *** ** * **   ****  * ****** ******* ***** ******* **
```

Fig. 20: Sequence comparison of clonally related new *emm* types.

```
stKNB2.0        GCTTTGACTGTTTTAGGAGCAGGCTTTACAAACCAAACAGAAGTTAAGGCAGCTGAAGCA 60
stKNB2.1        GCTTTGACTGTTTTAGGAGCAGGCTTTACAAACCAAACAGAAGTTAAGGCAGCTGAAGCA 60
stKNB2.2        GCTTTGACTGTTTTAGGAGCAGGCTTTACAAACCAAACAGAAGTTAAGGCAGCTGAAGCA 60
stKNB1          GCTTTGACTGTTTTAGGAGCAGGCTTTACAAACCAAACAGAAGTTAAGGCAGCTGAAGCA 60
stKNB3          GCTTTGACTGTTTTAGGAGCAGGCTTTACAAACCAAACAGAAGTTAAGGCAGCTGAAGCA 60
                ************************************************************

stKNB2.0        CCAGCAGCACCAGCAGCACCAGCAGCACCAGCAGCACCAGCACCAGCAGCAGCAAAAGCA 120
stKNB2.1        CCAGCAGCACCAGCAGCACCAGCAGCACCAGCA---------CCAGCAGCAGCAAAAGCA 111
stKNB2.2        CCAGCAGCACCAGCAGCACCAGCAGCACCAGCA---------CCAGCAGCAGCAAAAGCA 111
stKNB1          CCAGCAGCACCAGCAGCACCAGCAGCACCAGCA---------CCAGCAGCAGCAGCAAAA 111
stKNB3          CCAGCAGCACCAGCAGCACCAGCAGCACCAGCA--------------------------- 93
                *********************************

stKNB2.0        GCAGCAGCAAAAGCAGCAGCAGCAAAAGCACCAGCAGCAAAAGCACCAGCAGCAAAAGCA 180
stKNB2.1        GCAGCAGCAAAAGCAGCAGCAGCAAAAGCACCAGCAGCAAAAGCACCAGCAGCAAAAGCA 171
stKNB2.2        GCAGCAGCAAAAGCACCAGCAGCAAAAGCACCAGCAGCAAAAGCACCAGCAGCAAAAGCA 171
stKNB1          GCACCAGC---AGCAAAGCAGCAAAAGCACCAGCAGCAAAA------GCAGCAAAAGCA 162
stKNB3          GCACCAGC------ACCAGCAGCAAAAGCACCAGCAGCAAAA---------------GCA 132
                *** ****      *  *************************               ***

stKNB2.0        CCAGCAGCAAAAGCAAGAGAAAAACAGCTACTAGAAGAGTATAGAAAGTTAGAAGAAGGT 240
stKNB2.1        CCAGCAGCAAAAGCAAGAGAAAAACAGCTACTAGAAGAGTATAGAAAGTTAGAAGAAGGT 231
stKNB2.2        CCAGCAGCAAAAGCAAGAGAAAAACAGCTACTAGAAGAGTATAGAAAGTTAGAAGAAGGT 231
stKNB1          CCAGCAGCAAAAGCAAGAGAAAAACAGCTACTAGAAGAGTATAGAAAGTTAGAAGAAGGT 222
stKNB3          CCAGCAGCAAAAGCAAGAGAAAAACAGCTACTAGAAGAGTATAGAAAGTTAGAAGAAGGT 192
                ************************************************************

stKNB2.0        TATTTTAATTTAGAAGACCGTATAGAAAAACTAGGATCAGATTATGAAAAGTTAGAAAAA 300
stKNB2.1        TATTTTAATTTAGAAGACCGTATAGAAAAACTAGGATCAGATTATGAAAAGTTAGAAAAA 291
stKNB2.2        TATTTTAATTTAGAAGACCGTATAGAAAAACTAGGATCAGATTATGAAAAGTTAGAAAAA 291
stKNB1          TATTTTAATTTAGAAGACCGTATAGAAAAACTAGGATCAGATTATGAAAAGTTAGAAAAA 282
stKNB3          TATTTTAATTTAGAAGACCGTATAGAAAAACTAGGATCAGATTATGAAAAGTTAGAAAAA 252
                ************************************************************

stKNB2.0        GAAAATAAAGAGTACGCAAGTCAACTTGATAAAAATCAAGAAGAACGCGAAAAATTAGAG 360
stKNB2.1        GAAAATAAAGAGTACGCAAGTCAACTTGATAAAAATCAAGAAGAACGCGAAAAATTAGAG 351
stKNB2.2        GAAAATAAAGAGTACGCAAGTCAACTTGATAAAAATCAAGAAGAACGCGAAAAATTAGAG 351
stKNB1          GAAAATAAAGAGTACGCAAGTCAACTTGATAAAAATCAAGAAGAACGCGAAAAATTAGAG 342
stKNB3          GAAAATAAAGAGTACGCAAGTCAACTTGATAAAAATCAAGAAGAACGCGAAAAATTAGAG 312
                ************************************************************
```

Note: The homologous sequences were indicated by stars at the bottom. The gaps were indicated by hyphens. The numbers at the last of each line is the number of the nucleotide positions.

Fig. 21: Frequency of individual PAVF in the isolates of this study (n = 218).



Note: Numbers on top of each bar indicates the number of GAS isolates positive for each PAVF

Fig. 22: Percentage of isolates with PAVF for various categories.

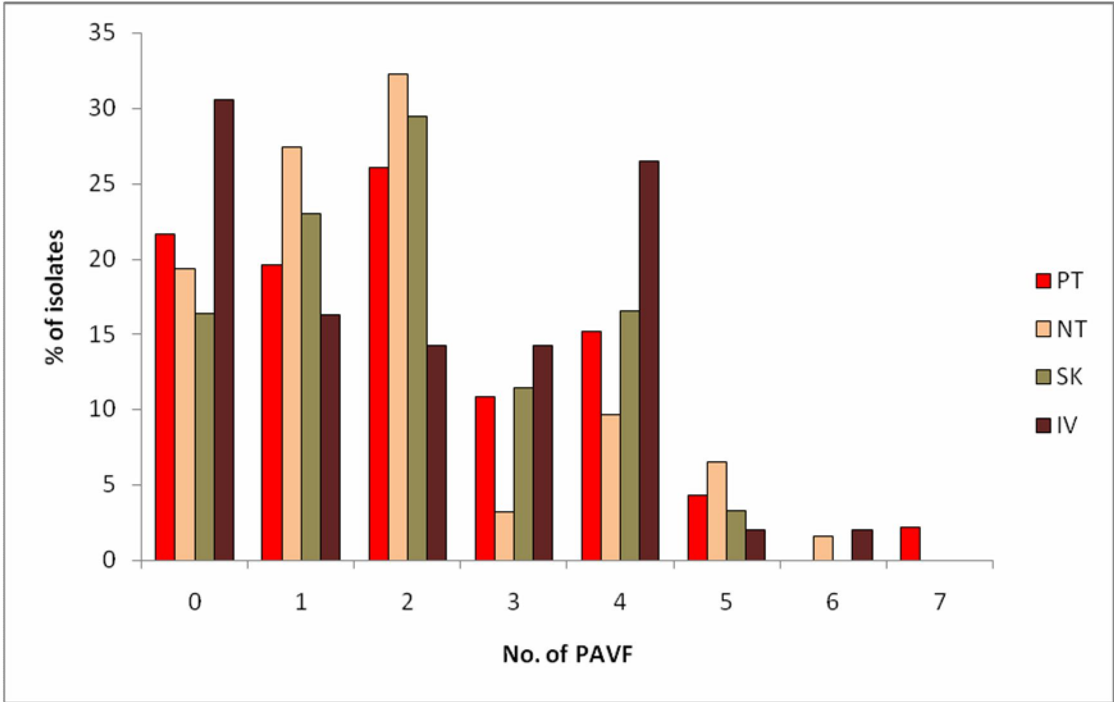


Fig. 23: Distribution of PAVF among PT, NT, SK and Invasive isolates.

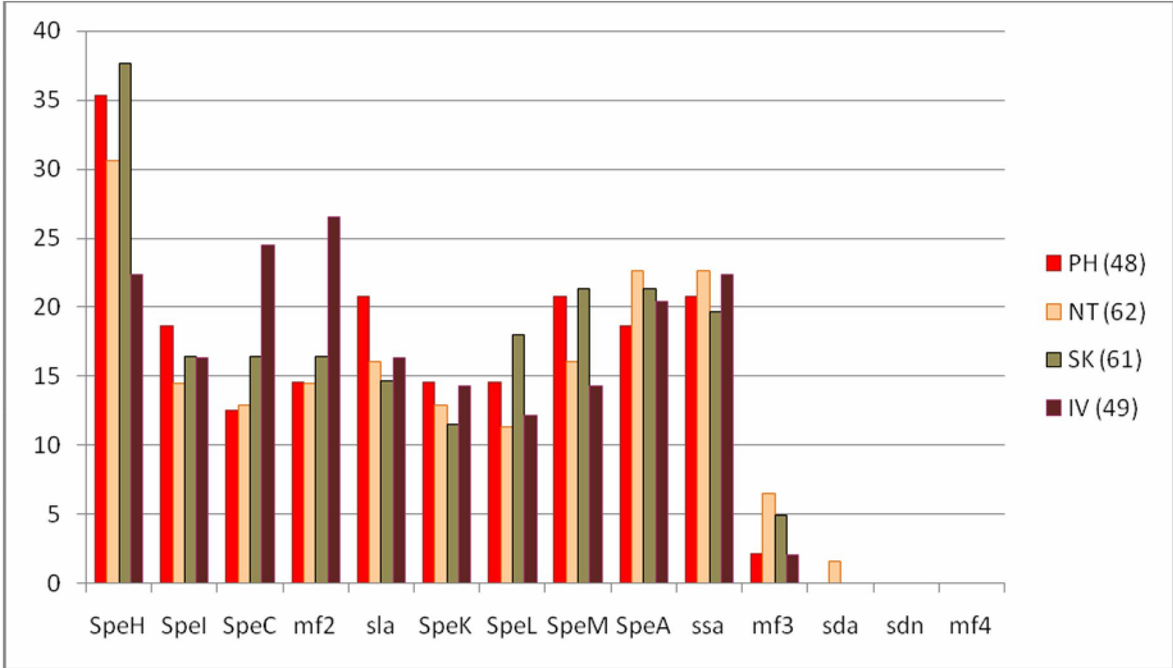


Fig. 24: Prevalence of individual CCVF among GAS isolates.

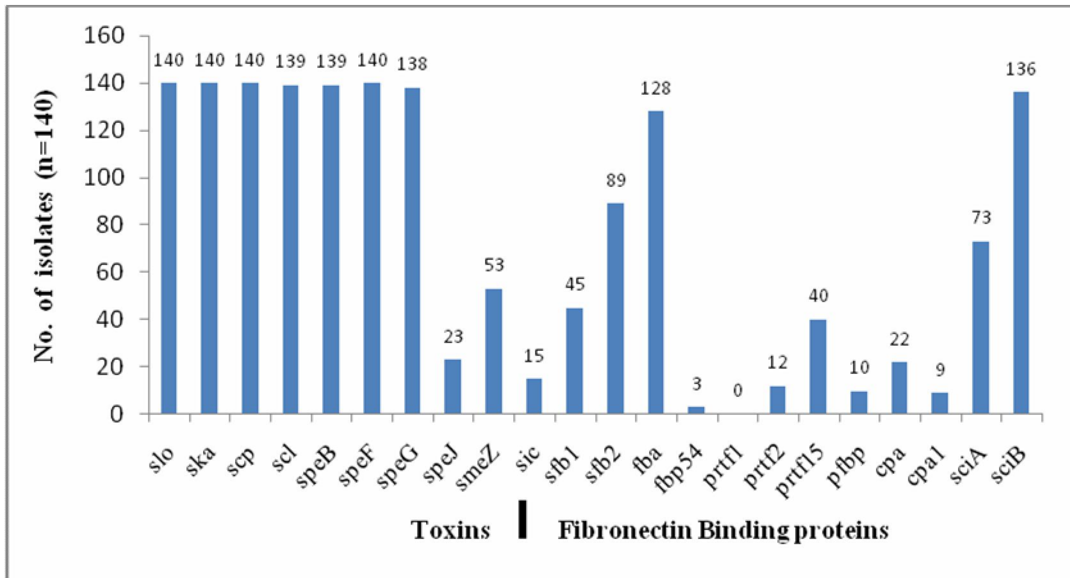


Fig. 25: Comparison of CCVF in PT, NT and SK.

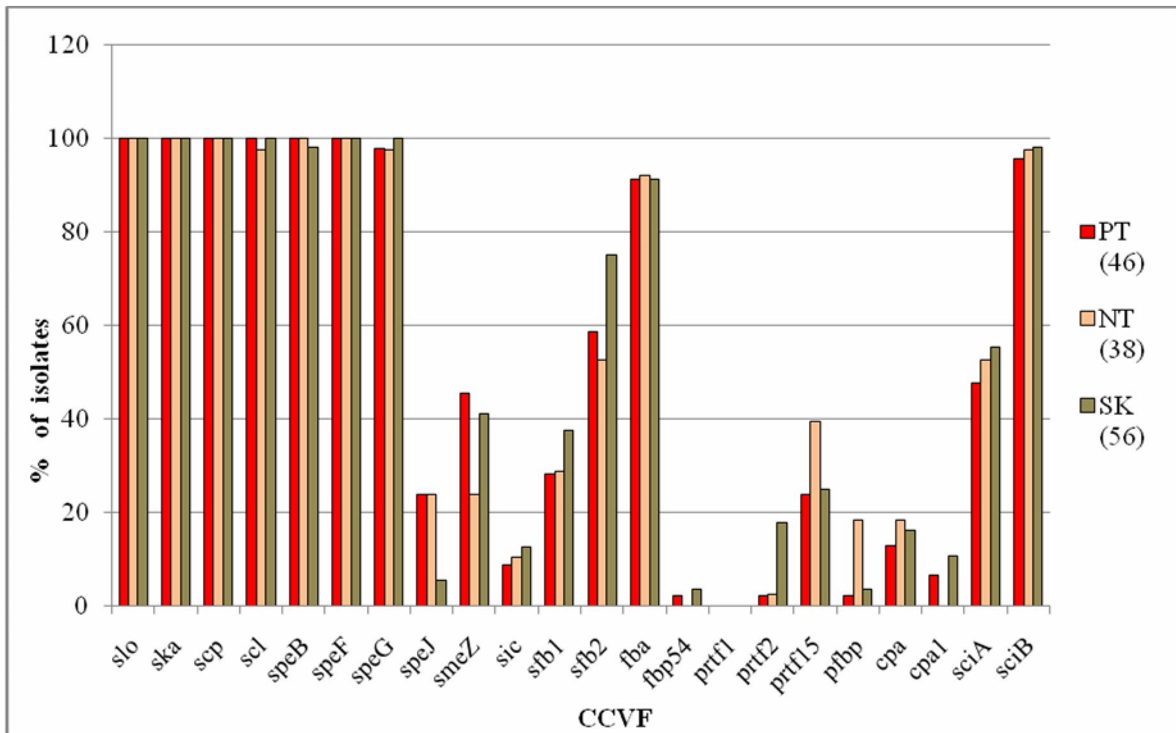


Fig 26: Pie diagram showing the extent of *emm* type diversity in South India (n = 698).

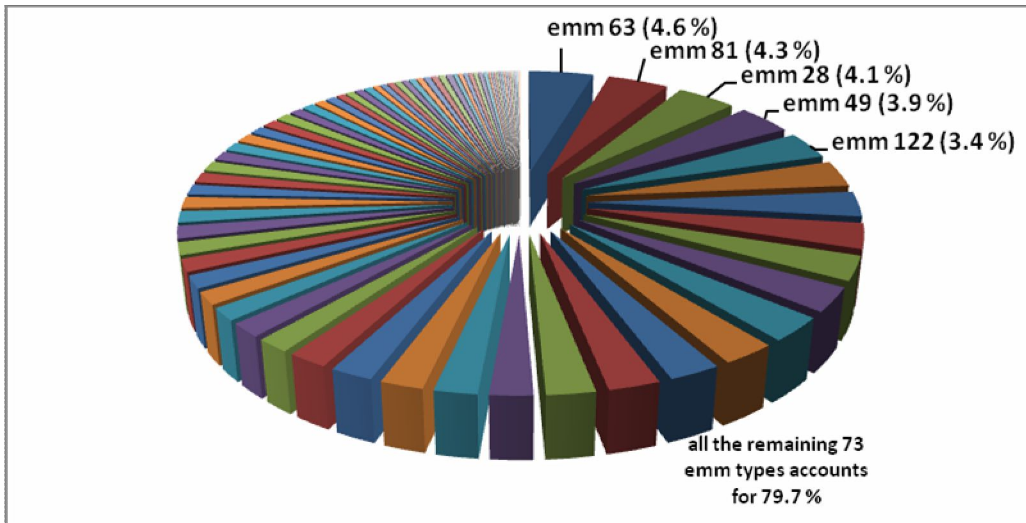


Fig 27: Pie diagram showing less diversity of *emm* type in USA (n = 1975).

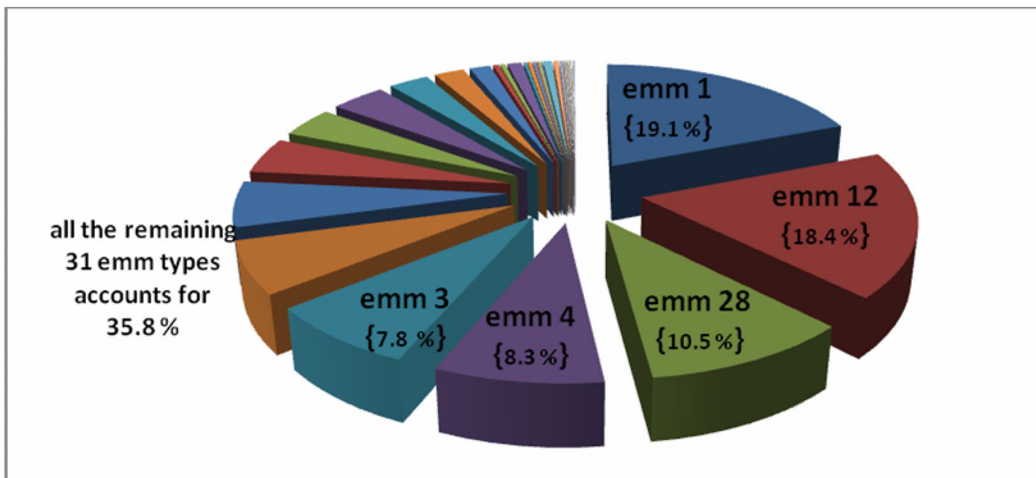
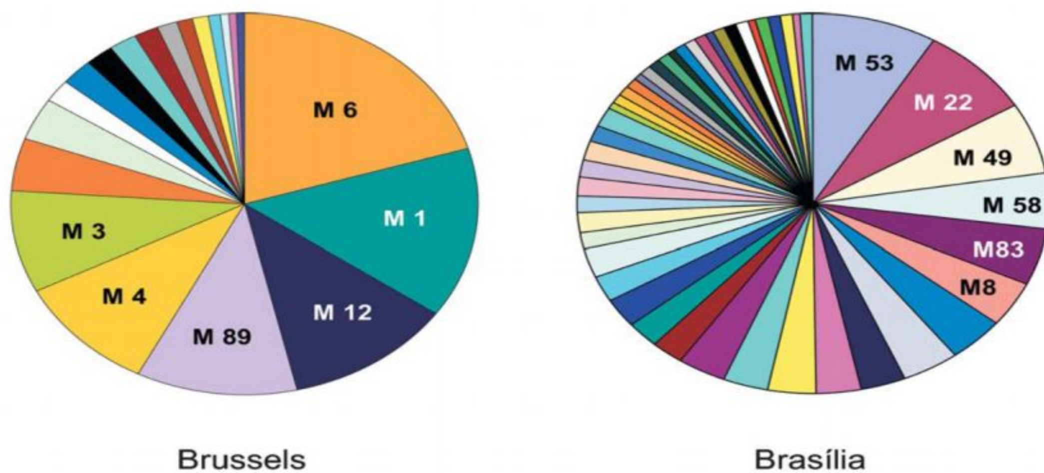
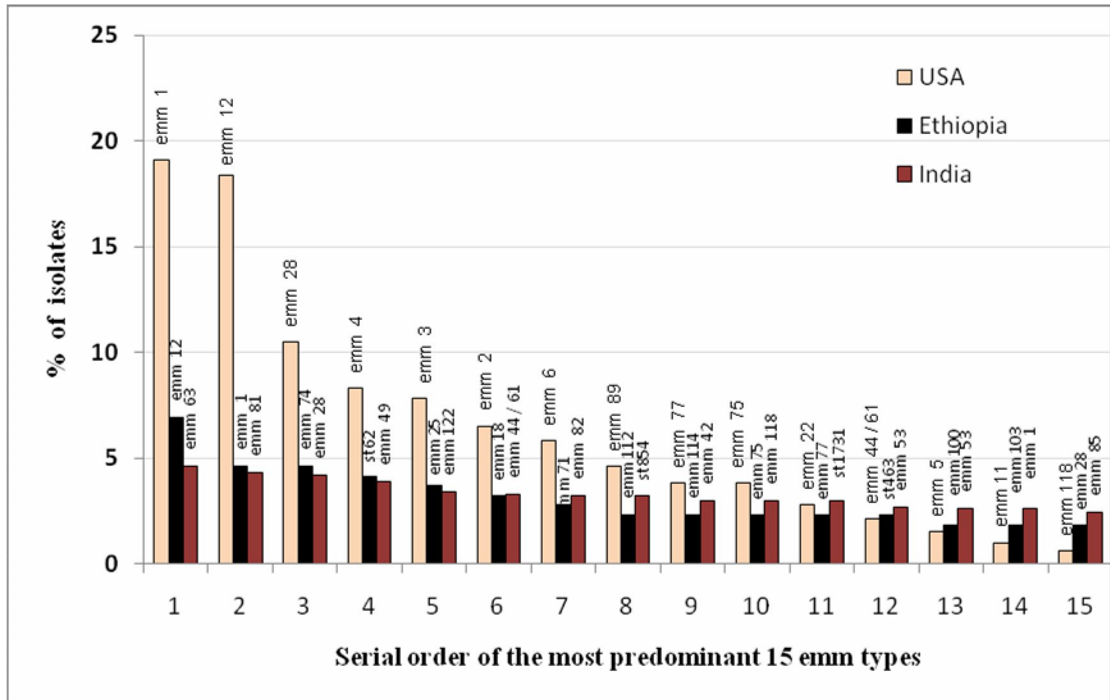


Fig 28: Comparison of *emm* type distribution from Belgium and Brazil.



Smeesters P.R., et al., [265].

Fig 29: Comparison of the predominance of *emm* types from India, Ethiopia and USA



Note: Given data is from this study and from Tewodros W. and Kronvall G. [16], And Shulman S.T., [15].