Phenotypic, Transcriptomic, and Proteomic Comparison of Clinically-Derived Group A Streptococcus Isolates in The United Kingdom Population

Alsahag, Mansoor

Department of Clinical Infection, Microbiology and Immunology

University of Liverpool

This dissertation is submitted for the degree of

Doctor of Philosophy

February 2019

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text. This dissertation contains less than 65,000 words including appendices, bibliography, footnotes, tables and equations and has less than 150 figures.

Contribution	**Names of individuals**
Measurement of capsular hyaluronic acid content	Dr. Jon van Aartsen, Post Doctoral Researcher, Institute of Infection and Global Health and Jenny Clarke, PhD student, IGH.
Transcriptomic analysis	Dr. Adrian Cazares, Research Fellow, Institute of Infection and Global Health
Proteomic data construction	Dr. Stuart Armstrong, Post Doctoral Researcher, Institute of Infection and Global Health

Abstract

Group A Streptococcus (GAS), also known as *Streptococcus pyogenes*, is an important cause of human diseases ranging from minor conditions such as pharyngitis to severe, life-threatening infections such as necrotising fasciitis. Between 2010 and 2012, a GAS outbreak (n=14 cases) caused by the *emm*32.2 GAS subtype occurred in Merseyside, which led to severe invasive disease with 29% case fatality. The outbreak was confined to adults and IV drug users; homelessness and alcohol abuse were identified as risk factors. We sought to investigate the genomic features of this subtype to explain its clinical phenotype. A first objective in my work was to carry out *in vitro* examination of the *emm*32.2 outbreak strain to identify whether phenotypic differences may be associated with specific genomic features hence explaining the epidemiological features of this *emm* subtype. A series of *in vitro* experiments were exploited and developed to determine whether and how *emm*32.2 might behave differently from other non-outbreak strains. We carried out a comparative examination of invasive *emm*32.2 isolates (n=14), against invasive and non-invasive isolates, i.e., *emm*6.0, *emm*89.0, and *emm*1.0, with respect to their capsule thickness, resistance to complement deposition, proteinase, NAD+-glycohydrolase (NADase) and Streptolysin O (SLO) activity, and biofilm formation. Our results demonstrated that phenotypic heterogeneity exists within the *emm*32.2 isolates as evidenced by variability in capsule thicknesses, proteolytic activity, SLO Activity, and biofilm formation. Interestingly, NADase activity was found to be below detection in all *emm*32.2 strains. More importantly, the *emm*32.2 isolates were less susceptible to complement deposition which may be translated to reduced phagocytosis in the host. In addition, RNA sequencing (RNA-Seq) and shotgun mass spectrometry approaches were used in order to identify differences in gene expression and expressed proteins levels of *emm*32.2 strain (112327) compare to *emm*1.0 (101910 invasive) and *emm*89.0 (127746 non-invasive) strains collected during the early, mid, and late exponential growth phases. My results displayed that a total of 76 GAS genes showing significantly different gene expression (p-value =< 0.01; $\log_2$-fold change > 4 or<- 4) in *emm*32.2 vs. *emm*1.0 and *emm*89.0. Notably, our data highlighted 16 virulence associated genes which showed different patterns of expression. This list included known virulence factors such as *slo* and *nga* which showed higher expression in *emm*32.2 isolates when compared to *emm*1.0 and *emm*89 isolates. In addition, proteomic analysis exhibited several recognised GAS virulence factors which were more abundant ($q < 0.05$ and fold change values ≥ 2) in *emm*32.2 compared to *emm*1.0 and *emm*89 isolates, including SLO and streptococcal NAD+ glycohydrolase (SPN). My project provides evidence that virulence factors such as SLO and SPN are essential for the virulence of *emm*32.2 GAS. Overall, characterisation of group A streptococcus outbreak isolates using high-throughput phenotypic, transcriptomic and proteomic analysis reveal genes associated with highly invasive strains and gives new insights into the development of potential GAS vaccine strategies.

Acknowledgements

I would like to acknowledge and thank colleagues who have supported and helped me during my PhD. First and foremost, I would like to express my deep gratitude and appreciation to Prof. Aras Kadioglu and Prof. Neil French, who have given me their valuable time, advice and guidance right from the start of my PhD up to writing this thesis. Aras and Neil, your guidance, enthusiasm and support have been outstanding over the past four years and I look forward to continuing to work together in the future.

I would also like to thank all the members of the bacterial pathogenesis and immunity group (past and present) who have taught and helped me during the years of my study. I would especially like to thank Jenny Clarke, Dr. Marie Yang, and Dr. Laura Jacques for their endless support and encouragement. My sincere thanks also go to Dr. Adrian Cazares who helped me with transcriptomic analysis.

I would like to give a special thanks to my wonderful mother and father for their help, encouragement and prayer, and also for making me the person that I am today, thank you for everything. I would also like to thank both my brothers and my sisters for their support and encouragement. I would like to express my deep thanks to my beloved Grandmother, for her love, prayer, and support. I hope you are proud Dad, Mum, and family.

In this very special moment, I express a deep sense of gratitude to my family, especially to my wife, who has always supported, encouraged, and stood by me through all the good and bad times.

Finally, I would like to dedicate this thesis to my beloved Baby boy Abdallah who sadly passed away in the middle of my PhD. Your love laughs, and smiles will stay with me forever.

Contents

List of Figures

List of Tables

Abbreviations

°C	Degrees Celsius
Ab	Antibody
amyA	Alpha-amylase
APC	Allophycocyanin
ARF	Acute Rheumatic Fever
BAB	Blood agar base
BHI	Brain heart infusion
C3	Complement component 3
C4BP	C4b-binding protein
CFU	Colony forming units
COGs	Clusters of Orthologous Groups
CovR/S	Component of virulence responder/sensor
CSPs	Culture supernatant proteins
CsrR	Two-component sensoregulator
CV	Crystal violet
D	Day
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
EEP	Early exponential phase
ELISA	Enzyme-linked immunosorbent assay
EndoS	Endo-β-N -acetylglucosaminidase
FabT	Fatty Acid Synthesis
FDR	False Discovery Rate
FHL-1	Factor H-like protein 1
FI	Fluorescence Index
Flincr	Florescent index increase
FITC	Fluorescein isothiocyanate
G	Gram
GO	Gene Ontology
GAS	Group A streptococcus
GP	General Practitioner
GpoA	Glutathione peroxidase
GRAB	Protein-G related-α2-macroglobulin-binding protein
H	Hour
HA	Hyaluronic acid

Has	Hyaluronic acid synthesis
HylA	Hyaluronidase lyate
HylP	Hyaluronidase
Ig	Immunoglobulin
IL 8	Interleukin 8
Kb	Kilobase
kDa	Kilodaltons
KEGG	Kyoto Encyclopedia of Genes and Genomes
LEP	Late exponential phase
M	Minutes
MEP	Mid-exponential phase
MF2	Mitogenic factor
MFI	Mean fluorescence intensity
Mga	Multiple gene activator
MGE	Mobile genetic elements
mL	Milliliter
MS	Mass spectrometry
Myr	M protein RNA
NAD+	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide hydrogen
NETs	Neutrophil extracellular traps
NF	Necrotizing Fasciitis
OD	Optical density
PBS	Phosphate buffer saline
PCA	Principal-component analysis
PCR	Polymerase chain reaction
PHE	Public Health England
PMNL	Polymorphonuclear leukocyte
RBCs	Red Blood Cell
RHD	Rheumatic heart disease
Rlog	Regularised logarithm
RNA	Ribonucleic Acid
rocA	Regulator of Cov
RopB	Protease B regulator
RVPBRU	Vaccine Preventable Bacteria Reference Unit
S.E.M.	Standard error of the mean
ScpA	C5a peptidase
Sda1	Bacteriophage-encoded extracellular streptodornase D
SibA	Secreted immunoglobulin binding protein
SIC	Secreted inhibitor of complement

SipA	putative signal peptidase
SLE	Systemic lupus erythematosus
SLS	Streptolysin S
SodA	Superoxide dismutase
SpeA	Exotoxin type A
SpeB	Streptococcal pyrogenic exotoxin B
SpeJ	Streptococcal pyrogenic exotoxins
SPN	Streptococcal NAD+ glycohydrolase
STSS	Streptococcal Toxic Shock Syndrome
SsE	Streptococcus secreted esterase
TCA	Trichloroacetic acid
THB	Todd-Hewitt Broth
THY	Todd-Hewitt Broth plus 0.5% Yeast Extract and Glucose
UK	United Kingdom
USA	United states of America
V/V	Volume/volume %
VST	Variance stabilizin transformationg
W/V	Weight per volume
WGS	Whole-genome sequence
µg	Microgram
µL	Microliter

Chapter 1 Introduction

1.1 Group A streptococcus (GAS)

1.1.1 Differentiation and classification of streptococci

Widely distributed in nature, *streptococci* are a heterogeneous group of gram-positive spherical bacteria that grow in pairs or chains [1]. Commonly found in the normal human microflora, they are also important causes of life-threatening human diseases. *Streptococci* have been classified on the basis of clinical and epidemiological studies such as colony morphology, haemolytic activity, and cell wall or capsular antigens specificity, for example, Lancefield antigens [1]. With the increased understanding of bacteria and the introduction of new characterisation means including the advances brought by molecular genetics, the methods of classification have now expanded and fall under multiple criteria [1].

One classification approach relies on the haemolytic activity of *Streptococci,* a phenomenon which is visible upon the growth of microorganisms on blood agar plates [2]. Three categories of haemolysis have been described: (i) alpha haemolysis, where the red blood cell haemoglobin is reduced and the area surrounding the bacterial colony is partially clear ; (ii) beta haemolysis, which corresponds to the full lysis of red blood cells in the area surrounding the bacterial colony; and, (iii) gamma-haemolysis, which refers to non-hemolytic microorganisms, where no clearing of the

area surrounding the colony is noted [2]. Due to variations in the assay conditions, such as species and length of storage of the red blood cells, haemolysis is not a robust classification method for *Streptococci;* however, it is recognised as a classic, useful and rapid screening method for identification of Group A streptococcus (GAS) [2].

Clinical samples are routinely screened on blood agar plates to determine the presence of beta-haemolytic colonies when the presence of GAS is suspected. After a 24-hour incubation at 35 – 37^0C, GAS typically appear as moist grey-white colonies surrounded by a clear zone where full red blood cell lysis has occurred. The cell wall of GAS is known to be composed of multiple repeats of common bacterial peptidoglycans [2]. First developed by Rebecca Lancefield, the antibody-based detection of the polysaccharide cell wall antigens of beta-haemolytic streptococci (i.e., Lancefield antigens) has become the standard method of identification for the differentiation of streptococcal species [3]. Due to the broad availability of commercial Lancefield antigen classification sera offered by multiple manufacturers, this classification method of beta-haemolytic streptococci is widely used in microbiology labs. Commercial assays classically contain both antigen extraction substrates and antigen-specific antibodies, which standardly target Lancefield antigens A, B, C, F and G and cause agglutination in the presence of the appropriate antigen. While all GAS strains exhibit type A antigen, with the exception of rare mutants, other strains of Streptococci also show this antigen. Thus, a positive result for Lancefield antigen A alone does not guarantee the presence of GAS. In the case of Lancefield antigen A detection, reports have shown that this antigen can also be found in *Streptococcus*

anginosus [4], as well as in rarely occurring *Streptococcus dysgalactiae* subsp. *equisimilis isolates* [5].

An increased specificity in the identifications of GAS isolates may be achieved by combining the Lancefield antigen A and exposure to bacitracin tests. Compared to other beta-haemolytic streptococci, that may express the group A antigen, GAS is sensitive to bacitracin while the other strains are bacitracin-resistant. Bacitracin sensitivity assay is also beneficial when differentiating GAS from other beta-haemolytic streptococcal strains, including *S. iniae* and *S. porcinus* [6-8]. In cases of bacitracin-resistant GAS strains, the possible identification of GAS strain may be made upon the combined criteria of colony morphology, beta-haemolysis, absence of catalase, activity of the enzyme pyrrolidonyl arylamidase (PYR), and presence of group A antigen. [4]

1.1.2 Description

GAS is a gram-positive coccoid-shaped bacterium that grows in pairs or chains. On a blood agar plate (Figure 1-1), it is visible as white/grey colonies surrounded by a zone of beta-haemolysis; however, though it is rare, some strains of GAS are not haemolytic [8]. Over 100 GAS serotypes have been described, based on their combination of M-protein cell wall antigens. Serotyping of the M-antigen was classically carried out using polyclonal antibody sera; however, this method has gradually been replaced by molecular tools, such as polymerase chain reaction (PCR) and DNA sequencing, in which the *emm* gene is targeted, that encodes the M-antigen. Today, over 130 distinct

M serotypes have been identified using this method, many of which show parallel characteristics with the previously serum-defined serotypes [9, 10].

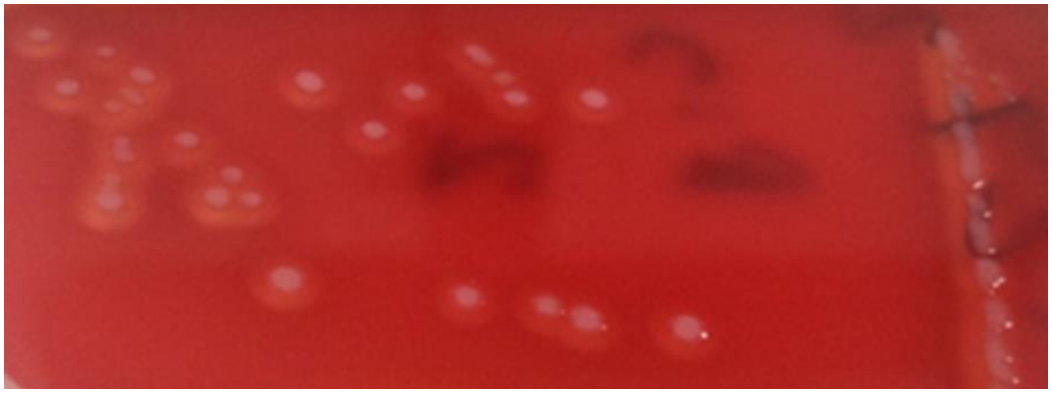


Figure 1-1 Group A Streptococcus causes beta-haemolysis in blood agar plate.

1.1.3 Clinical importance of GAS infection

GAS, which is found in all ages and population demographics, is responsible for a range of both minor and life-threatening invasive diseases [11]. Some groups of people are more susceptible to GAS infection [12], including those in care homes [13, 14], while outbreaks occur at lower incidence in the general community [15, 16]. GAS is commonly found as a coloniser of the epithelial surfaces of the nasopharynx, skin, and vagina, from where it can spread and cause a plethora of infections. These infections range from relatively minor conditions such as pharyngitis and impetigo to more invasive and severe infections such as bacteraemia, necrotising fasciitis and streptococcal toxic shock syndrome [17].

1.1.4 GAS disease transmission

Transmission of GAS disease occurs through multiple routes, including mainly inhalation of exhaled droplets and skin-to-skin contact with an infected person [18]. Studies suggest that transmission also occurs via direct contact with contaminated surfaces or objects, or through dust particles [18]. The effect of crowding on the facilitation of disease transmission has also been documented whereby overcrowding was shown to promote the spread of GAS infection and GAS disease outbreaks in military facilities [19].

1.1.5 Epidemiology of GAS disease

Globally, GAS-related disease epidemiology and distribution is variable. While a decline was observed in a number of developed countries, GAS-related infections, such as pharyngitis and invasive diseases, remain a significant area of public health concern [10], with the majority of fatalities resulting from invasive diseases [20]. In low-income settings, which frequently have poorer infrastructure, half a million GAS-associated deaths occur annually because of high incidence of diseases, predominantly caused by rheumatic heart disease (RHD). RHD is an immune disorder, resulting from damages to one or more heart valves caused by either a single acute episode or multiple recurrent episodes of acute rheumatic fever (ARF) [20]. Due to their weaker immunity and greater exposure levels in school and nursery environments, disease incidence rates are usually higher in children [20]. The highest pharyngeal infection rates are seen in children over the age of three, while it is infrequent in neonates due to passive immunity acquired from the mother [21].

Recent epidemiological studies conducted in industrialised countries reported GAS invasive disease frequency of 2 to 4 per 100,000 (0.002 – 0.004%). In contrast, infection rates were found to be substantially higher at 12 to 83 per 100,000 (0.012 – 0.083%) in developing countries [22]. Globally, the burden of GAS invasive disease is high with reports of 663 000 new cases and 163 000 deaths annually [23].

The global incidence of non-invasive GAS diseases, such as pharyngitis and tonsillitis, are also significant and result in substantial economic impact. Of over 600 million annual cases of pharyngitis, most of them (i.e., 550 million cases) occur in low and middle-income countries [24], and 11 million cases are reported in the United States of America. In European countries, 8% of all general practitioner (GP) consultations are for pharyngitis, with 15% of school-aged children in developed countries developing strep throat (pharyngitis) annually, impacting on economy through loss of school and work days [24]. In developing countries, the incidence of pharyngitis in school children are five- to ten-fold higher [23]. While pharyngitis is not solely caused by GAS - since many cases are viral-related - it was reported that 15 – 30% of cases are GAS-related [25].

In developing countries, GAS infection is no longer considered life-threatening although an increased incidence was reported over the last decade [20]. Following pharyngitis infection, rheumatic heart disease (RHD) represents the most frequent GAS disease with 15 million cases reported globally, leading to 233,000 deaths annually. In developing countries, RHD levels reach 2.4 million annual cases in children [26], affecting a total of 15.6 – 19.6 million individuals globally [23].

1.1.6 Molecular epidemiology

In the past two decades, the molecular epidemiology of GAS has shifted from the study of single genes to population-based genomic comparisons, leading to new findings on the genetic basis of GAS pathogenesis, clonal emergence, and strain genotype–disease phenotype relationships [27]. A recent study based on whole-genome sequence (WGS) combined with phenotypic studies identified the sequential changes that have contributed to the global epidemic caused by *emm*1 (M1T1) GAS clonal type. The study reported that this clone had acquired three regions of heterologous DNA: a 36-kb chromosomal region and two bacteriophages [28].

1.1.7 GAS disease outbreak

GAS infection with high variations observed in disease rates and associated outbreaks, as seen in the recent 2013/2014 United Kingdom outbreak [29] and the 2011–2016 China outbreak of scarlet fever where infection dramatically increased [30]. GAS outbreaks were documented to be associated with the acquisition of prophages carrying virulence genes [31, 32] or the acquisition of genetic changes producing novel *emm* genotypes [33, 34]. The horizontal transfer of mobile genetic elements (MGE) has been an important factor in the evolution of *emm*12 and *emm*1 which have both been associated with cases of scarlet fever in Hong Kong. Previous studies have shown that the predominantly found *emm*12 type acquired MGE conferring tetracycline and macrolide antibiotic resistance as well as prophages encoding superantigens SSA, SpeC and the DNase Spd1 [31, 35].

Similarly, the horizontal transfer of MGE has also played a major role in the evolution of the M1T1 clone which has acquired three regions of heterologous DNA: a region that encodes the toxins Streptolysin O (SLO) and NAD-glycohydrolase (SPN) , a bacteriophage-encoded extracellular streptodornase D (Sda1) and an exotoxin type A (SpeA) contributing to enhanced virulence and colonisation, respectively [36]. Other findings showed that the early *emm*89 clonal populations differed from the current *emm*89 strains found in the United Kingdom by the acquisition of the NADase and streptolysin O locus[34]. The *emm*89 variant showed an enhanced expression of SPN and SLO resulting in a potentially more pathogenic strain with a selective advantage over other *emm*89 variants, translated by its changing epidemiology [34].

1.2 GAS Disease spectrum

Diseases caused by GAS (Figure 1-2) are often the combined result of an (i) acute infection, and (ii) host-mediated immune responses.

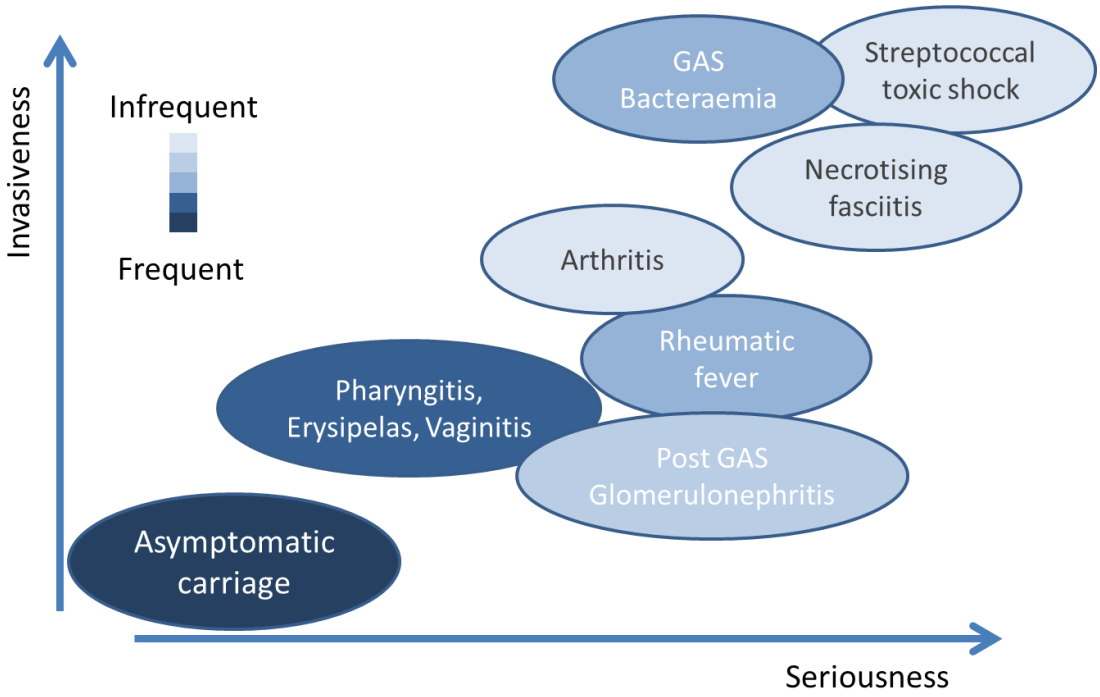


Figure 1-2 Group A Streptococcus Disease Spectrum.

1.2.1 Pharyngitis

The most common manifestation of GAS infection is pharyngitis (Figure 1-3), commonly called strep or sore throat [37]. More common in the Winter and Spring, 20-40% of GAS pharyngitis cases occur in children and 5-15% cases are in adults [38, 39]. Clinical symptoms of GAS pharyngitis include sore throat accompanied by sudden fever, often seen as swelling of the tonsils and pharynx with patchy white secretions [40]. Other symptoms include tiredness, headache, abdominal pain, nausea and vomiting [25, 41]. The transmission of this disease occurs via both direct and indirect contact with an infected person, e.g., contaminated objects, respiratory secretion droplets [37]. Under certain circumstances, autoimmune antibodies may be raised due to molecular mimicry with cardiac and joint tissue antigens [42, 43], whereby the individual's immune response can lead to non-infectious conditions, such as acute rheumatic fever and post-streptococcal glomerulonephritis [37].

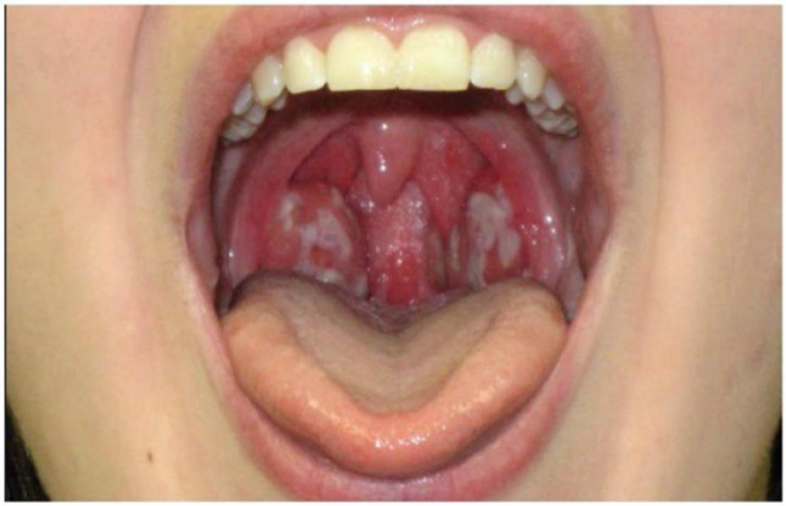


Figure 1-3 Infection with pharyngitis. ***(Source: [44])***

1.2.2 Acute rheumatic fever (ARF)

Acute rheumatic fever (ARF) is an inflammatory condition that primarily affects the joints and cardiac tissues and is most commonly found in children and young adults between the ages of 5 and 14 years old [45, 46]. It usually occurs three weeks post GAS-pharyngitis infection [23] with initial symptoms usually beginning with the presentation of painful arthritis in the medium and large joints in 60-80% of patients [45].

1.2.3 Necrotising fasciitis (NF)

Necrotising fasciitis (NF) is a severe soft tissue infection with a mortality rate of 70 – 80% [47]. Also called Streptococcal gangrene, it is a rapidly spreading, and is characterised by extensive necrosis of the skin, deeper subcutaneous tissues and facia [47]. Observed in patients with certain medical conditions, Streptococcal gangrene may occur post-operatively at the incision site and occur spontaneously or after a minor trauma [48].

1.2.4 Streptococcal toxic shock syndrome (STSS)

Streptococcal toxic shock syndrome (STSS) is a systemic toxicosis which may occur because of a streptococcal infection, although only rarely following streptococcal pharyngitis [49-51]. It results in the sudden onset of shock and leads to organ failure [47]. This is the most severe complication of streptococcal bacteraemia. In community studies where there is a higher incidence of STSS, M-type 1 or 3 strains have been predominantly found in throat isolates [52-54].

1.3 Therapeutic approaches

For many decades, penicillin-based antibiotics have been the treatment of choice for GAS infection and are still effective. Irrespective of the long-term and frequent use of penicillin-based treatments, GAS has remained extremely sensitive to treatment and has not developed any resistance. This extraordinary continued susceptibility is likely due to a lack of beta-lactamase expression or to the fact that beta-lactamase is toxic to the bacteria. Alternatively, it is also possible that the optimal conditions for GAS resistance to develop have not yet occurred [55].

In patients presenting penicillin allergy, erythromycin is used as an alternative treatment, as are first generation cephalosporins; however, these are restricted to patients who do not exhibit immediate-type hypersensitivity to β-lactam antibiotic [56].

1.4 Vaccination

Despite intensive investigations from public health authorities and scientists to develop a successful vaccine to combat GAS infection, efforts have been unsuccessful. Facing complex challenges, including avoidance of cross-reactivity with human tissue and design of clinical trials that demonstrate efficacy of a vaccine against not only skin infections but also the rarer manifestations of GAS infection, has been a challenge [26].

The growing understanding of the molecular basis of GAS infection has led to the discovery of several vaccine candidates that are at various stages of development. A leading vaccine candidate is the M-antigen, which was shown to not only confer protection in animal models, but antibodies in human sera protect against homologous GAS infection clinical trials [57]. A highly complex, multivalent 26-valent vaccine has been genetically engineered against 26 antigens, which showed no cross-reactivity with human tissue, and well tolerated by 30 healthy adult volunteers the vaccine [26, 57, 58]. The vaccine is now being considered for paediatric trials. Vaccine development efforts are focused on the prevention of GAS disease in developing countries and on the design of vaccines targeting ARF and RHD. Safe and effective GAS vaccines could have a substantial impact on the health of millions of individuals [59].

1.5 Biofilm

The capacity of GAS to form biofilms both *in vivo* and *in vitro* is not well understood and their impact on human disease remains obscure, when compared to biofilm producing microorganisms such as *Pseudomonas aeruginosa* and *Staphylococcus sp.*

A bacterial biofilms represents a sensitive community of microorganisms and biofilm development is described as a five-stage process [60, 61]. In brief, stage 1 consists of single floating cells (planktonic cells) briefly adhering to a surface. At this early stage, small quantities of extra polymeric materials are linked to the attached cells and many of the cells can perform movement independent of the colony [62]. Attachment

becomes more stable during stage 2 when cell production of extracellular polymer increase. The biofilm structure becomes more established and matures during stages 3 and 4. Stage 5 promotes discontinuation of the biofilm cycle, during which individual cells or discrete cell colonies are being detached from the biofilm structure. A mature GAS biofilm contains proteins, DNA and a polysaccharide-containing component [63-65]. Current literature provides evidence of GAS biofilm production, with researchers identifying GAS microcolonies in the skin lesions of impetigo patients [65]. Similar findings were observed in experimental skin infections in zebrafish models [66]. *In vitro* studies using a variety of materials, including polystyrene microtiter plates, plastic coverslips and flow chambers, covering both static and flow conditions, also demonstrated GAS biofilm production [64, 67].

Biofilm production may be a key explanation for antibiotic therapy failure against GAS infection where antibiotic sensitivity was shown *in vitro*. Studies indicate that non-invasive strains of *Streptococcus* are more likely to form biofilms compared to invasive strains [68, 69]. Similar findings were described for erythromycin-sensitive and erythromycin-resistant isolates of GAS [70]. These findings strongly support the hypothesis that biofilm production functions as a microbial defensive mechanism against both antibiotics and host immune responses [71].

While biofilm production was observed in 90% of strains, both invasive and non-invasive, research suggested that biofilm formation is likely to be a strain-associated trait rather than a general serotype characteristics [70]. The complex process of GAS biofilm production is driven by external conditions such as pH, but also by virulence

factors such as capsule, Streptococcal pyrogenic exotoxin B (SpeB), and M protein [72].

1.6 Main virulence factors

The extensive range of GAS infections is attributable to its comprehensive array of virulence factors which facilitate adherence, immune system evasion, stimulation or degradation of specific host components, and direct cell lysis. The virulence factors depicted in Figure 1-4 do not represent a complete list, but represent a selection of well-characterised factors.

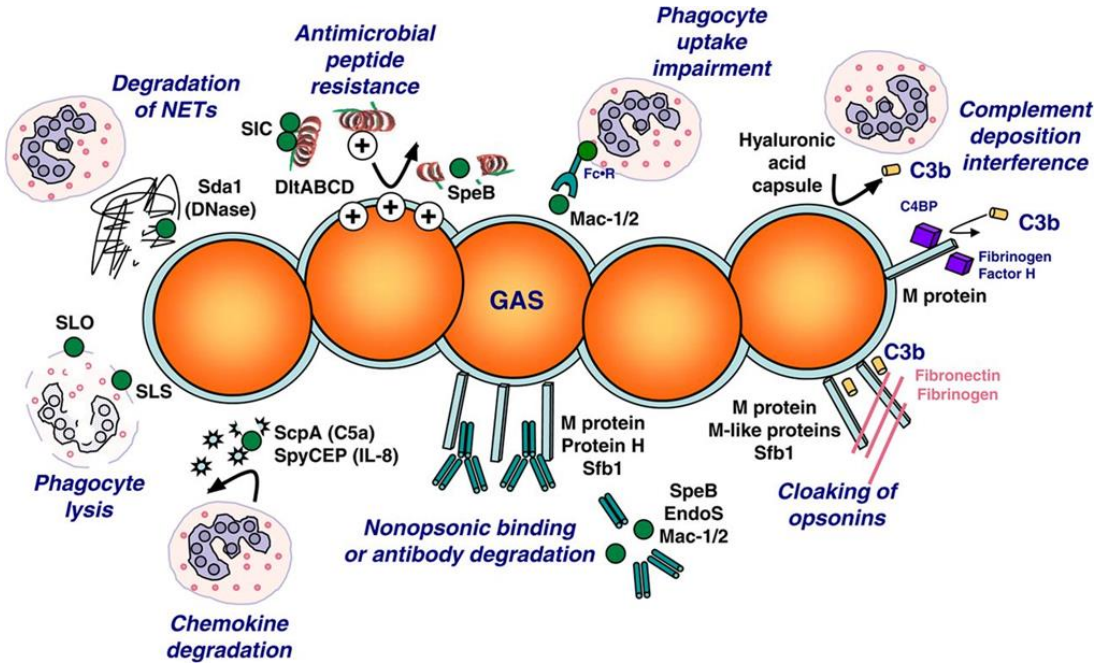


Figure 1-4 Group A Streptococcus virulence factors.

The collection of virulence factors that GAS expresses in order to evade the host innate immune response. (Source: [11]).

1.6.1 Hyaluronic acid capsule

Hyaluronic acid capsule is a critical virulence factor for a number of GAS strains which has the role to shield them from the extracellular environment[11]. Hyaluronic acid capsule facilitates pathogenesis by increasing resistance to phagocytosis [73, 74], GAS *produces* a hyaluronic acid- capsule that is structurally indistinguishable from the hyaluronic acid produced in human connective tissue [75]. Most GAS clinical isolates produce hyaluronic acid. Hyaluronic acid polymers are not covalently connected to GAS cell wall, but rather are associated with the cell surface as they are synthesised by a cell membrane-associated polymerase. The enzymes responsible for capsule production are regulated by a highly conserved three-gene operon, regulated by an upstream promoter. The genes and enzyme products are: *hasA, hasB,* and *hasC which* encodes hyaluronate synthase, UDP-glucose dehydrogenase, *and* UDP-glucose pyrophosphorylase, respectively [74, 75]. This operon is regulated by a bacterial two-component system of virulence responder/sensor (CovR/S; formerly designated CsrR/S) which has a major role in regulating the expression of GAS virulence factor in response to environmental signals and which negatively controls capsule expression [74-76]. A noteworthy regulatory protein, RocA, plays a significant role in capsule expression ,and demonstrated that the inactivation of mutations in this protein reduces expression of the two-component sensor regulator (CsrR), which is linked to higher capsule production [77]. Previous studies have shown that RocA negatively regulates the ability of serotype M1 GAS to express capsule [78]. Difference in capsule production can be attributed to promoter elements and it is notable that the *has* operon

in the hyper-encapsulated *emm*18 strain is three times more active than the *has* operon promoter found in a poorly encapsulated strain [74, 76].

Capsule has been associated with GAS virulence since its initial discovery [79, 80]. In previous study, it was reported that the GAS capsule may obstruct antibody's access to surface epitopes [81], and it was also demonstrated that capsule also promotes the ability of GAS to inhibit complement deposition [73]. Other authors have reported that GAS capsule facilitates opsonophagocytic resistance [73, 82] and improves survival in neutrophil extracellular traps (NETs) [83]. A number of *in vivo* studies reported that the expression of GAS capsule was a requirement for full virulence [79, 84, 85].

1.6.2 M protein

The cell wall of GAS carries an important and characteristic antigenic part called the M protein which is encoded by the *emm* gene [11]. It consists of long fibrils shaped into a relatively-conserved C terminal anchored to the cell wall with an N-terminal hypervariable region [86, 87]. With their ability to bind multiple human proteins, including fibronectin, and C4b-binding protein (C4BP) [88, 89], M proteins (Figure 1-5) promote adherence to host cells, and evasion of complement binding and phagocytosis [90]. Due to the importance of M proteins, GAS strains were classified by *emm* types - over 200 have been described so far - as identified by the hypervariable region of the N-terminal sequence of M protein [9, 90, 91]. A correlation was found between specific *emm* types and individual streptococcal diseases, for example, *emm*1

(M1) strains are associated with invasive disease while *emm* types 3, 5, 6, 18, and 89 are pharyngitis-associated; and *emm* types 1, 5, 6, 14, and 18 which are correlated with acute rheumatic fever [11, 36, 92]. Studies have also reported that specific *emm* types may also be geographically restricted [10, 93].

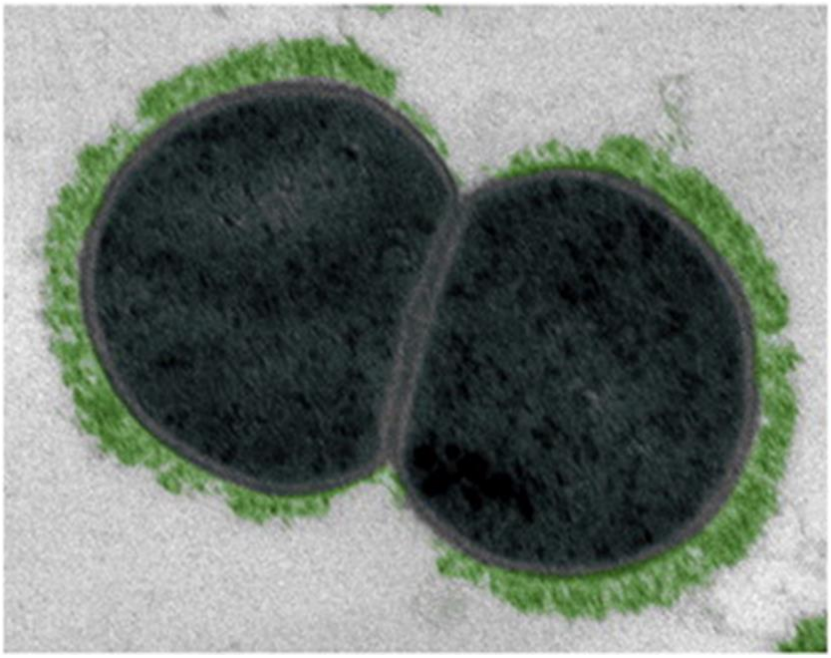


***Figure 1-5 The appearance structure of M protein molecules under electron microscope** (Source: [94]).*

1.6.3 Streptococcal pyrogenic exotoxin B (SpeB)

Streptococcal pyrogenic exotoxin B (SpeB) is an anchorless adhesin, which functions as a cysteine protease and is present in the majority of GAS isolates [95]. SpeB is capable of cleaving both GAS factors and host molecules [95]. One proposed scenario is that SpeB cleaves tight junction proteins to allow the passage of bacteria across the epithelium which facilitates progression to invasive disease [96]. *In vivo* studies [97] propose that SpeB modulates host inflammatory response through a number of mechanisms including degradation of complement factors (C3b and formation of the membrane attack complex) [97, 98]. In addition, SpeB also degrades multiple GAS virulence factors, including M1, streptolysin O (SLO), secreted inhibitor of complement (SIC) and streptokinase. Aziz *et al.* suggested that these virulence factor degradation is controlled by altered SpeB expression during *in vivo* infection [97].

1.6.4 Streptolysin O (SLO)

Streptolysin O (SLO) (Figure 1-6) is a secreted, oxygen-labile, thiol-activated toxin that binds cholesterol in cell membranes [99]. SLO aids in the pathogenesis of GAS infection by polymerising in membranes to form pores that lyse erythrocytes, macrophages, lymphocytes, and platelets [99-101]. As such, SLO contributes to the cytotoxicity of host cells [102]. SLO also delivers NAD-glycohydrolase (NADase, Nga, or SPN) to the host cell cytoplasm in a process known as cytolysin-mediated translocation [103].

In addition to the remarkable function of streptococcal NAD-glycohydrolase transfer into host epithelial cells [103], studies have determined that multiple SLO functions were directly related to GAS pathogenesis. Varying expression levels of SLO were shown to be associated with virulence [11], with enhanced efficacy and disease causation associated with higher expression levels in invasive isolates in comparison to non-invasive isolates [104]. Tissue damage induced by SLO activity generates increased mucosal inflammation [105, 106]. The immune response to SLO during infection is confirmed by the presence of serum anti-streptolysin O antibodies[107].

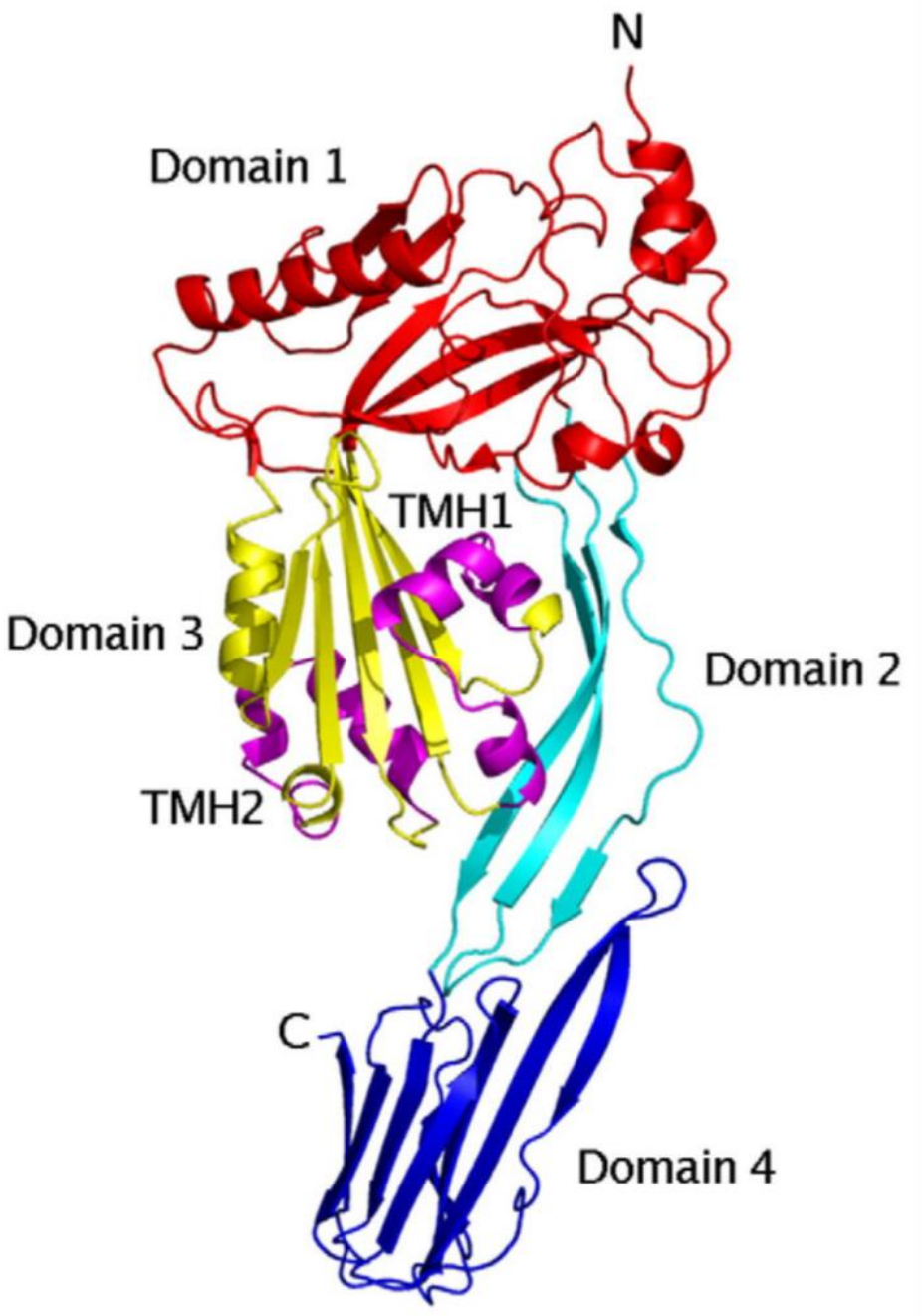


Figure 1-6: Domain structure of streptolysin O.

The figure gives a ribbon schematic of the SLO. Domains (D) 1, 2, 3, and 4 are indicated by D1 in red, D2 in cyan, D3 in yellow, and D4 in dark blue, respectively.D3 provide the transmembrane spanning regions of the toxin. D4 (initial anchoring points) is critical for binding to cholesterol-containing membranes (Source: [108]).

1.6.5 Streptococcal NAD+ glycohydrolase (SPN)

The streptococcal NAD+ glycohydrolase (SPN) is a secreted protein that is encoded by the *nga* gene [109]. The role of NADase is not fully understood, but data suggests its function in pathogenesis as a NAD-glycohydrolase [109]. When NADase is present in the cytoplasm, NADase activity depletes host cell energy stores [110]. Host cell biology is dramatically altered by delivery of NAD-glycohydrolase into the human cell cytoplasm, which results in increased streptococcal pathogenicity and intercellular survival [111, 112]. To date, all GAS strains have been shown to contain the *nga* gene ,however, some isolates were recently reported to produce a protein lacking NADase activity [113]. While inactive NADase is known to be toxic to *E. coli* strains, this is not the case with GAS [110]. This difference is believed to be linked to a currently undefined NADase-independent function of the protein. In addition, while NADase is reported to be associated with cytotoxicity [110], the role for this form of the protein in GAS pathogenesis remains undefined. Interestingly, a correlation was proposed between tissue tropism and NADase subtype. It was described that *emm* types that infect both throat and skin have a propensity to express the NADase-active form of the protein, and the *emm* types that are tissue-specific - infecting either throat or skin only, but not both - have a tendency to produce NADase-inactive proteins [114].

1.6.6 Streptolysin S (SLS)

The cytolytic toxin streptolysin S (SLS), is a β-haemolysin produced by the majority of *GAS* strains [115]. Its name derives from the serum-based culture used to grow the streptococcal cells from which it was initially extracted. SLS is an oxygen-stable toxin encoded within a nine-gene operon (sagA - sagI) [116]. SLS induces hydrophilic pores in a wide range of cell types, of both innate and adaptive immune system, including erythrocytes, leukocytes, and platelets, causing osmotic cell lysis, and aiding in virulence [117, 118]. SLS contributes to the pathogenesis of GAS in a variety of ways which include cytotoxicity, activation inflammatory response, and phagocytosis inhibition [119]. *In vitro* and *in vivo* laboratory experiments showed that SLS-mutants exhibit reduced virulence [120].

1.6.7 DNases

DNases (also known as streptodornases) are a highly conserved family of DNA degrading enzymes. GAS is known to produce up to four DNases as extracellular products (DNaseA-D) [121]. DNaseB is chromosomally encoded and present in 100% of strains, while the other classical extracellular DNases are reportedly phage-encoded and present in 20-50% of strains [122].The bacteriophage-encoded DNase Sda1, also known as streptococcal DNase D2 (SdaD2) which is a key contributor to virulence and is utilized solely by the *emm*1 (M1T1) [11, 28, 123]. Through degradation of DNA-based neutrophil extracellular traps, the DNase protects GAS from neutrophil-mediated killing [124, 125].

1.6.8 Streptococcal inhibitor of complement (SIC)

SIC, one of the most polymorphic bacterial proteins known to date, is a 31 kDa protein that interferes with complement-mediated lysis by inhibiting the binding of the membrane attack complex (MAC) onto bacterial cell membranes [126-129]. In addition, SIC is able to disrupt the activity of other innate immune mediators such as lysozyme [126, 130].

The reported function of SIC in restricting the complement system via blocking C3a activity, and in inhibiting the formation of antimicrobial-peptides via and blocking binding site of Human kininogens (HK) to the bacterial surface [131, 132], suggests that SIC-producing GAS strains may have a selective advantage resulting in enhanced virulence and wider global disease incidence [132].

Furthermore, due to the changes occurring in cellular processes induced by SIC genes, SIC contributes to GAS pathogenesis in various ways including enhancement of bacterial survival and adherence, as well as host colonisation [133, 134]. It was proposed that exposure to the host immune response during infection drives variation in the SIC gene which leads to enhanced bacterial survival [135].

1.6.9 Superoxide dismutase

Superoxide dismutase (SodA) is a secreted protein which converts superoxide anions to oxygen and hydrogen, subsequently detoxified by peroxidases [136]. SodA is detected on both cell surfaces and cell culture supernatants [137]. The host white blood cells produce an oxidative burst in response to detection of these pathogenic bacteria,

and SodA plays an important role in detoxifying this attack, particularly as GAS does not produce a catalase. GAS has been considered susceptible to host oxidative burst killing. However, recent studies have identified a glutathione peroxidase (GpoA) produced by GAS that supports the bacterium's ability to survive oxidative stress during phagocytosis. This was shown to contributes to virulence in several GAS infection animal models [138, 139].

1.6.10 Immunoglobulin binding proteins

Immunoglobulin binding proteins are important virulence factors which have been associated with invasive GAS isolates [11]. GAS exhibits a variety of surface-linked, immunoglobulin binding proteins, which include the M-proteins, M-related proteins and M-like proteins [11]. The M and M-like proteins primarily target IgA, which is found predominantly in mucosal secretions, while the M-related proteins target IgG, which is mainly located in the bloodstream [11, 140, 141]. It is postulated that these proteins have an active role in immune system evasion; however, the underlying mechanisms remain to be elucidated [11]. A highly conserved secreted immunoglobulin binding protein (SibA), a 45kDa protein targeting both the Fc and Fab regions of IgA, IgG and IgM [11, 142] was confirmed to be present in the majority of GAS strains [142].

1.6.11 EndoS (endo-β-*N* -acetylglucosaminidase)

Endo-β-N -acetylglucosaminidase (EndoS), encoded by the *ndoS* gene, is a secreted endoglycosidase that specifically degrades immunoglobulin G *via* removal of its carbohydrates [11, 143]. By reducing binding of IgG to the Fc receptors and inhibiting complement activation, this enzyme improves survival of GAS in the host [144]. EndoS is therefore believed to inhibit phagocytic uptake of GAS by impairing the functionality of the IgG molecule. In contrast, previous *in vivo* studies also showed that EndoS had no significant influence on GAS phagocyte resistance when they examined a highly virulent *emm*1 type [145]. This suggests that there may be variation amongst GAS strains in relation to their immunoglobulin inactivating activities.

1.7 Research Background

One invasive GAS (iGAS) case was identified after the admission of an adult male at a secondary healthcare unit in Liverpool, Merseyside in September 2012. The infection was caused by the isolate *emm*32.2, which was identified using Vaccine Preventable Bacteria Reference Unit (RVPBRU) molecular protocols (http://www.cdc.gov/streplab/mproteingenetyping.html). Researchers identified two more patients who had been diagnosed with *emm*32.2 iGAS from the same secondary healthcare unit. Therefore, Public Health England (PHE) conducted an inquiry to determine the existence of common sources of exposure between the three cases. The inquiry found that each of the patients acquired *emm*32.2 iGAS in the community.

Furthermore, after reviewing the typing database maintained by RVPBRU, investigators noted that other *emm*32.2 iGAS cases had emerged in the community since January 2010. In the period going from January 2010 to September 2012, 14 *emm*32.2 iGAS cases were identified by the RVPBRU, 12 of which were reported in Liverpool and 2 in the greater Merseyside region. No subsequent *emm*32.2 cases have since emerged in the region, while 30 non- *emm*32.2 iGAS infections had been identified over the same period in Liverpool [146].

We reported that the novel GAS *emm*32.2 emerged as an outbreak strain in Liverpool, Merseyside, U.K [146]. A key characteristic of this strain is the fact that it suggested that the acquisition of super-antigens, or multi-drug resistance, did not drive the *emm*32.2 outbreak. As was the case with other iGAS outbreaks [11], but was instead related to the emergence of a new *emm* type. We hypothesised about the factors that

could have facilitated the dissemination of the disease into the population, including the collection of virulence factors identified in the *emm*32.2 outbreak strains - in particular, those not observed in different iGAS strains of the same population over the same period. Based on the results of genetic analysis, 19 genes unique to the outbreak isolates have been observed. Of the accessory genes which were found to be unique to *emm*32.2, the following 5 have been previously associated with bacterial virulence: namely, multiple gene activator (Mga)-like regulatory protein, M protein RNA yield (Myr) positive regulator, a putative signal peptidase, (SipA also called LepA), a putative trypsin resistant surface protein T6, and hyaluronidase, (HylP) [146].

Here, the Mga-like regulatory protein is classified as a DNA-binding protein, and the literature indicates that it supports activation of the expression of several key virulence factors [147]. As for the Myr regulator protein, studies have demonstrated that it plays a role in phagocytosis *via* the manipulation of antiphagocytic M protein transcription [148]. Regarding the potential signal peptidase LepA (commonly referred to as SipA), it was shown to be required for pilus formation and *tee6* gene expression [149]. It is also worth noting that hyaluronidases are important streptococcal virulence factors, and studies have found evidence to suggest that they promote bacterial proliferation within host tissues [150].

1.8 Aims and Objectives

1.8.1 Aims

The aim of this project was to characterise a subset of GAS *emm*32.2 outbreak isolates using in vitro assays and cell-based models. High-throughput transcriptomic and secretome analyses of one *emm*32.2 isolate were also conducted in comparison with *emm*1.0 and *emm*89 isolates collected from within the same people community and over the same period, with *emm*1.0 as an invasive GAS isolate representative and *emm*89.0 as a non-invasive representative.

1.8.2 Objectives

- To conduct *in vitro* phenotypic analyses of the *emm*32.2 outbreak strain compared to *emm*1.0, *emm*89.0 and *emm*6.0 with respect to their capsule thickness, immune complement deposition, proteinase activity (SpeB), NADase / SLO activity, and biofilm formation. This will contribute towards gaining insight into the virulence properties that may be specific to *emm*32.2.
- To determine the differential RNA gene expression of one *emm*32.2 isolate in comparison to *emm*1.0 (invasive) and *emm*89.0 (non-invasive) types during their *in vitro* growth in broth, using transcriptomic analysis. This will aid to define the transcriptomic features that have contributed to the dissemination of *emm*32.2 and its emergence within the Merseyside population.
- To identify the *in vitro* secretome of *emm*32.2 isolates in comparison to *emm*1.0 and *emm*89.0 types using mass spectrometry proteomics analysis. This will

complement the results obtained in objective 2/3 and holds the potential to identify novel virulence factors specific to the pathogenicity of *emm*32.2 type.

In a recent high-resolution pan-genome analysis, the authors reported that the *emm*32.2 isolates encoded a unique set of 19 genes that were not identified in any of the other isolates collected during the same outbreak period [146] . A key observation is that, unlike previously reported GAS outbreak [11], emergence of these *emm*32.2 outbreak strains was not due to the acquisition of super-antigens or multi-drug resistance genes, but was rather the result of a new genetically distinct *emm* type. Here, I hypothesize that this unique subset of 19 virulence genes has facilitated the emergence and dissemination of the *emm*32.2 GAS type within a disease susceptible group of individuals. To verify this hypothesis, a range of in vitro assays have been performed with the ultimate goal to dissect the virulence properties unique to *emm*32.2 outbreak strains and further the understanding of GAS pathogenicity.

Chapter 2 Materials and Methods

2.1 Microbiology

2.1.1 Bacterial strains

The GAS isolates used in this project are listed in Table 2-1. The strains were clinical isolates and were provided by Professor Neil French. All test isolates (including *emm*32.2, *emm*1.0, *emm*89.0 and *emm*6.0) in this PhD project were collected as part of an outbreak inspection in Liverpool United Kingdom (UK) that was by performed Public Health England (PHE). Ethical approval was not mandatory in this study and the chosen isolates were subjected to whole-genome sequencing (WGS). The *emm*1.0 and *emm*89.0, and *emm*6.0 invasive isolates used in the study were randomly selected from 30 invasive GAS isolates which were identified in Liverpool at the same time as the outbreak. As the local microbiology laboratories do not routinely keep non-invasive isolates, these were selected from a set of 20 non-invasive pharyngitis GAS isolates which were provided by 2 hospitals in Liverpool namely the Royal Liverpool University Hospitals Trust and Alder Hey Children's Hospital.

Table 2-1 Group A Streptococcus strains used in this PhD study

Number	Strain	Age*	Sex	Sample type	*emm* type	Date	Outcome
1	101008 - i	47	M	Blood Culture	*emm*32.2	05/01/2010	Discharged
2	101452 - i	39	F	Blood Culture	*emm32.2*	10/03/2010	Death
3	101700 - i	29	M	Blood Culture	*emm32.2*	25/04/2010	Discharged
4	101967 - i	87	F	Blood Culture	*emm32.2*	09/06/2010	Discharged
5	102029 - i	43	M	Blood Culture	*emm32.2*	20/06/2010	Discharged
6	102434 - i	41	F	Blood Culture	*emm32.2*	23/08/2010	Discharged
7	102920 - i	58	F	Blood Culture	*emm32.2*	13/11/2010	Discharged
8	103045 - i	45	M	Blood Culture	*emm32.2*	05/12/2010	Death
9	111617 - i	53	F	Blood Culture	*emm32.2*	25/04/2011	Discharged
10	112327 - i	56	F	Blood Culture	*emm32.2*	10/09/2011	Discharged
11	112844 - i	42	M	Blood Culture	*emm32.2*	12/12/2011	Discharged
12	121324 - i	51	M	Blood Culture	*emm32.2*	01/03/2012	Death
13	121511 - i	71	M	Blood Culture	*emm32.2*	31/03/2012	Discharged
14	122397 - i	18	M	Blood Culture	*emm32.2*	08/09/2012	Discharged
15	101910 - i	37	M	Left hip pus	*emm1.0*	25/05/2010	─
16	127251 - n	3	M	─	*emm1.0*	11/05/2012	─
17	126215 - i	─	─	─	*emm6.0*	─	─
18	137027 - n	39	F	─	*emm6.0*	02/05/2012	─
19	101724 - i	34	F	Blood Culture	*emm89.0*	30/04/2010	─
20	137072 - n	43	M	─	*emm89.0*	06/06/2012	─
21	127785 - n	35	F	─	*emm6.0*	24/05/2012	─
22	127784 - n	35	M	─	*emm6.0*	14/05/2012	─
23	137081 - n	44	M	─	*emm89.0*	11/03/2013	─
24	127746 - n	29	F	─	*emm89.0*	08/01/2013	─

M=Male, F=Female, i= invasive, n=non-invasive.*= Years. ─ = information was not available.

2.1.2 Standard media

Blood agar base (BAB)

16 grams (g) of BAB medium (Sigma, Dorset UK) was mixed with 400 ml of distilled water and autoclaved. Following autoclaving, the media was left at room temperature until reaching 56°C. 20 ml of Sterile Defibrinated Horse Blood (Sigma, Dorset UK) was added and evenly mixed. The media was then poured into sterile petri dishes (90 mm) and left to dry overnight. The plates were inverted and stored at 4°C for 1-2 weeks. One bottle of media produced approximately 25 plates.

Brain heart infusion (BHI)

20 g of BHI medium (Sigma, Dorset UK) was mixed with 400 ml of distilled water and autoclaved. After the media reached 56°C, it was then poured into sterile petri dishes (90 mm) and left to dry overnight. The plates were inverted and stored at 4°C for 1-2 weeks. One bottle of media produced approximately 25 plates.

Todd Hewitt broth +0.5% yeast (THY)

14.4 g of Todd Hewitt broth (Sigma, Dorset UK) plus 2 g of yeast (Sigma, Dorset UK) was mixed with 400 ml of distilled water an autoclaved. The medium was stored at room temperature.

Columbia blood agar base

7.8 g of Columbia blood agar (Oxoid, Thermo Scientific, Hampshire UK) was mixed with 200 ml of distilled water an autoclaved at 121°C for 15 minutes.

Columbia blood agar-1.5% milk

200 ml 3% (w/v) skim milk autoclaved and add to 200 ml Columbia blood agar base autoclaved.

2.1.3 Solutions

Phosphate buffer saline (PBS)

To prepare PBS, 10 tablets of phosphate buffer saline (PBS, Oxoid, Thermo Scientific, Hampshire UK) were dissolved in 1L of distilled water and autoclaved at 121°C for 15 minutes.

40 % Glucose

To prepare 40% Glucose, 40 g of glucose added and mixed with 100 ml of distilled water and autoclaved.

3% (w/v) skim milk

To prepare 3% (w/v) skim milk, 6g skim milk powder added and mixed with 100ml of distilled water and autoclaved.

2.1.4 Preparing stocks of GAS

To obtain a single colony blood, agar plates were streaked GAS isolates from laboratory bead stocks. Plates contained 5% horse blood and were incubated in a CO_2 GAS jar overnight at 37ºC. The next day, a single fresh colony was inoculated into a

universal tube which had 7.9 ml of Todd-Hewitt Broth (THB) and 5% of extract yeast along with 100 µl of 40 % glucose which was added to the tube prior to adding the bacteria. The universal tube was incubated statically for 16-18 hours at 37ºC. The following morning, 200 µl from the overnight culture was added to 8 ml of fresh THYG broth. Subsequently, the broth was incubated for 4-6 hours to reach an optical density of OD_{600nm} 0.5. After that, 20% (v/v) of glycerol was added to the broth. Next, 500 µl was transferred into an Eppendorf tube and then stored at -80°C. The Miles and Misra method was used to determine the number of colony forming units (CFU) in the bacterial stocks [151]. Both the Bacitracin disc (sterile, 6mm diameter filter paper discs impregnated with 0.04 Units of Bacitracin 08382 Bacitracin Disks, Thermo Scientific, Hampshire UK) and Streptococcal grouping kit (PathoDxtraTM Strep Grouping Kit DR0700M, Thermo Scientific, Hampshire UK) were used as rapid identification tests for GAS.

2.1.5 Determination of GAS viable counts on blood agar plates

According to standard Miles and Misra methodology [151], sterile round-bottomed 96 well plates (Thermo Scientific, Hampshire UK) were used to determine the number of bacteria in liquid culture. Serial dilutions from 10^1-10^6 were performed by adding 20 µl of the liquid culture to 180 µl of sterile PBS (PBS, Sigma, Dorset UK). The blood agar plates were divided into six sections and performed in triplicate. 60 µl (approx. 3x20 µl) of each dilution was spotted on duplicate blood agar plates corresponding sector. When dry, the plates were inverted and incubated at 37^0C 5% CO2 for 16-18 hours. The optimal numbers of colonies (between 30-300 colonies) were identified

and the number of bacteria was determined as colony forming units (CFUs) per ml, using the following equation: CFU/ml = (Total number of colonies counted in sector) ÷ 60 (total volume) X 100.

2.2 Quantification of capsule production: FITC-Dextran exclusion

Capsule thickness was measured using the FITC-dextran zone of exclusion method, similar to that described by Hathaway and colleague [152]. Bacteria were cultured overnight with Todd-Hewitt Broth (THB) and 5% of extract yeast medium. Afterwards, 200µl was subcultured into 8 µl of fresh medium and cultured again until OD_{600nm} 0.5± 0.1. After centrifuging at 3000g for 10 minutes the pellet was re-suspended with 500 µl of PBS. 10 µl of bacterial suspension was mixed with 2 µl of 2000 kDa FITC-dextran (10 mg/ml; Sigma-Aldrich, Dorset UK), pipetted onto a microscope slide and coverslip applied. The slides were viewed using a Nikon Eclipse 80i fluorescence microscope with 100x objective lens and photographed by a Hamamatsu C4742-95 camera. ImageJ was used to determine the zone of exclusion (area in pixels), a value proportional to capsular thickness. Each strain was prepared twice on different days and analysis done on the blind pictures.

2.3 Measurement of capsular hyaluronic acid content

Capsular hyaluronic acid was extracted as per established protocols [153]. Briefly, GAS isolates were cultured as for the measurement of capsular thickness. Hyaluronic

acid was extracted by adding a 2:5 volume of chloroform, vortexing for 30 seconds and placing in a tube inverter for 15 minutes. The aqueous phase was used to determine hyaluronic acid concentration, as per manufacturer's instructions (HA Test Kit, Corgenix, Colorado USA). Data was normalised by CFU counts and was presented as hyaluronic acid concentration per 10^3 CFU. This assay is done kindly by J van Aartsen (postdoc) and Jenny Clarke (PhD student).

2.4 Complement deposition

The complement deposition assay was roughly based on a previously published method [154]. $2.5x10^6$ CFU of bacteria from frozen bacterial stocks were added to 7 ml of BHI, incubated at 37°C for 15 minutes, and centrifuged at 3000 g for 10 minutes. The supernatant was removed and the pellet was washed using 1 ml of PBS and re-pelleted at 3000 g for 10 minutes. The pellet was resuspended in 100 µl of 20%-HSVB (PBS with 20% human serum (pooled from five individuals) and 1% gelatin veronal buffer) and incubated at 37ºC for 30 minutes. Next, the tube contents were washed by adding 900 µl of PBS, centrifuging at 16 000 g for 3 minutes and removing the supernatant. This was followed by resuspending the bacteria in 100 µl of mouse-anti-human-C3 in PBS (1:300; Abcam ab17455 Ms Ab to C3 [8G4], Abcam, Cambridge UK) and incubation at 37°C for 30 minutes. Washing was repeated, and the contents were in resuspended in 100 µl of anti-Mouse IgG2a-APC in PBS (1:400; EBioscience 17-4210-80 anti-Mouse IgG2a APC, Abcam, Cambridge UK) and incubated at 4°C for 30 minutes in the absence of light. A final washing step was performed; the

remaining bacteria were resuspended in 100ul of PBS and then incubated for 5 minutes at room temperature with thiazole orange (BD Cell Viability kit, Fisher Scientific, Loughborough UK) (42uMol/L). Control samples without complement or antibodies were run in parallel to experimental samples. Samples were stored in the absence of light at 4°C until acquisition. Samples were acquired using the Accuri C6 flow cytometer. Only thiazole orange positive signals were included for analysis.

2.5 Casein-plate assay

Group A Streptococcus expression of extracellular cysteine protease, is also referred to as streptococcal pyrogenic erythrogenic toxin B (sepB), was determined by a plate assay as previously described[155]. Briefly, single GAS colonies were stab-inoculated into Columbia agar base plates containing 3% w/v skim milk. Protease-expressing strains produced a translucent zone surrounding the site of inoculation after 24-h aerobically incubation at 37 °C. The appearance of the casein hydrolysis zone was considered as protease activity positive. Two independent assays were performed for each strain.

2.6 Biofilm assays

GAS biofilms were measured using previously described methods with minor modifications [156]. Briefly, streptococci was grown overnight in THY, diluted 1:40 in THY and 100 µl of the suspension was added to wells of a polystyrene, microtiter plate (Corning Costar 3598, Sigma, Dorset UK). The plate was incubated for 24, 48,

96 and 168 hours at 37°C. Afterwards, the plates were gently washed 3 times with PBS and the plate was left to dry at room temperature for one hour. Crystal violet solution (125 µl) was then added to each well and after 10 minutes, the wells were washed with distilled water. Next, 150 µl of 95% ethanol was added, the plate was shaken for 1 minute by hand, and the absorbance at 595 nm was recorded. Wells incubated with THY without streptococci were used as the blank. To assess biofilm sensitivity to trypsin, 40 µl of 10x trypsin (2.5% solution of 1:250 trypsin; Sigma-Aldrich, Dorset UK) was added to wells after 22 hours of incubation. Isolates were analysed in eight replicate wells, in three independent experiments.

2.7 SLO activity

SLO activity in culture supernatant was measured as described elsewhere [36], with minor modifications. Isolates were cultured to an OD_{600nm} of 0.25 ± 0.05. Bacteria-free supernatants were prepared by centrifuging at 3000 g for 10 minutes and filtering through a 0.25 µm filter. Supernatants were incubated room temperature for 10 minutes with 20 mmol/l of dithiothreitol (Sigma-Aldrich, Dorset UK). For each strain, 500 µl of supernatant was aliquoted into two tubes; 25 µg of water soluble cholesterol (inhibitor for SLO activity) was added to one tube. Both tubes were incubated at 37°C for 30 minutes, followed by the addition of 250 µl of 2% sheep erythrocytes/PBS suspension to each sample and further incubation at 37°C for 30 minutes. After 500 µl of PBS was added to each tube the samples were centrifuged at 3000 g for 5 minutes.

200 µl of each sample was transferred to a 96-well plate and the OD_{541nm} was measured. Each isolate was tested in three independent experiments.

2.8 NADase activity in culture supernatant

NADase activity in culture supernatant was assayed as described elsewhere [112], with minor modifications. Overnight bacterial cultures were centrifuged at 3000 g for 10 minutes and the supernatant was filtered through a 0.25 µm filter. Supernatants were two-fold serially diluted with PBS in 96 well plates, and 0.67 mmol/l NAD+ (100 nmol; Sigma-Aldrich, Dorset UK) was added to each well. The plate was incubated for 1 hour at 37°C, and developed by adding NaOH to a final concentration of 2 M followed by further incubation for 1 hour at room temperature in the dark. Plates were visualised using a microplate reader (Fluorostar Omega, BMG Labtech) with excitation/emission detection profile wavelengths set at 355 nm / 520 nm. Results were reported as the highest 2-fold dilution capable of fully hydrolysing 100 nmol of NAD. Each isolate was tested in three independent experiments.

2.9 Bacterial growth curve

Growth curves were produced to compare the growth kinetics of different GAS strains and to determine the three phases of growth: early exponential phase (EEP), mid-exponential phase (MEP), late exponential phase (LEP). For each strain, The OD_{600} of 200 µl of 10^5cfu/ml bacteria in THY media (supplemented with glucose) was

determined every 30 minutes for a total of 24 hours using a microplate reader (Fluorostar Omega, BMG Labtech).

2.10 Genome-wide transcriptome-sequencing (RNA-seq) analysis

2.10.1 Sample collection

Growth curves of *emm*32.2 (112327), *emm*1.0 (101910) and *emm*89.0 (127746) were performed by inoculating 8 ml of THBY-Glucose with 1 x10^5 bacteria and incubating them at 37ºC. Samples were taken every 2 hours between 0 and 12 hours to determine the CFU. Results were recorded and analysed to plot the growth curve respective of each strain of interest. This experiment was done in duplicate and plotted to determine the three phases of growth: early exponential phase (EEP), mid-exponential phase (MEP), late exponential phase (LEP).

2.10.2 RNA extraction protocol

Briefly, for the preparation of Ribonucleic Acid (RNA) from each growth phase, GAS strains were cultured in 8 ml THY-Glucose in a universal tube solution containing 1x10^5 of each strain, from frozen aliquots. Each strain was grown once at 37ºC, and samples were collected after 4h, 6h, and 10h, corresponding to early exponential phase, mid-exponential phase, and late exponential phase, respectively. The supernatant was discarded, and the bacterial pellet was used immediately for RNA extraction.

2.10.3 RNA isolation

RNA was isolated as previously described [157] with minor modifications, briefly, at the designated time points, separate aliquots were collected from the cultures for transcriptome analysis. The bacterial pellet was added to 2-ml tubes containing lysis matrix B (MP Biomedicals, UK) and lysed using Bead beating equipment (The BeadBug™). The RNA was extracted with an RNeasy minikit (Qiagen, Manchester UK) following the manufacturer's instructions. The extracted RNA was subjected to DNase treatments with Turbo DNA-free (Thermo Scientific, Loughborough UK). The quantity of RNA was determined using the Qubit® Fluorometer (Thermo Scientific, Loughborough UK). A 2100 Bioanalyzer (Agilent Technologies, Stanta Clara, USA) was used to assess the quality of the RNA. Once confirmed that the RNA was not degraded the samples were sent to Vertis Biotechnologie AG (Freising, Germany) for Illumina NextSeq 500 sequencing (Illumina).

2.10.4 Transcriptome analysis

Transcriptomic data analysis was performed by Dr. Adrian Cazares, at the University of Liverpool, Liverpool - U.K. Briefly Raw Illumina sequencing reads in FastQ format were processed for both adapter removal and low-quality sequence filtering with TrimGalore v0.4.4. (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) using the default settings for single-end reads. This step aims to remove primer and low-quality sequences to avoid biases in further steps of the analysis as carrying non-biological sequences or error bases. Filtering was subsequently verified using FastQC v0.11.5.

(https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). In average, 98.78% of the reads survived the filtering process. Sequencing reads were mapped (i.e., aligned) to the genome of the GAS strain MGAS5005 (CP000017; GenBank) using the Bowtie2 algorithm [158] from the EDGE-pro software[159] to determine which genes in the bacterial genome they belong . Estimation of the levels of gene expression was also performed by following the EDGE-pro pipeline. From 86 to 96% of the reads were mapped to the reference genome, except for the *emm*89.0.L sample which featured 54% of the reads mapped. Gene count tables were calculated with the "edgeToDeseq.perl" script, part of the EDGE-pro package.

An exploratory differential expression analysis of the samples was performed in the R environment [160] with the DESeq2 package [161] by using the gene count tables with tRNA and rRNA genes masked. Genes were considered differentially expressed when the transcript level change was at least 4-fold between samples and the p-value was less than 0.01. As sample replicas were not available, putative differential expression was mainly considered based on the observed log2 fold change values. Regularised logarithm (rlog)- and Variance stabilising (VST)-transformed values of the count data used both for performing the principal component analysis (PCA) of the samples and generating the heatmaps in the R environment, were obtained from the DESeq2 analysis.

Prediction of the operon structure was carried out with the Rockhopper v2.0.3 package [162] from the alignment files in order to identify groups of contiguous genes displaying similar levels of expression.

Assignment of the putative differentially expressed genes to Clusters of Orthologous Groups (COGs) was recovered from a previous study [163] whereas the corresponding COGs functional categories were accessed from the NCBI COG table for GAS M1. (ftp://ftp.ncbi.nlm.nih.gov/pub/COG/COG2014/static/lists/listStrpyo.html).

The presence/absence of putatively down-regulated genes in the genome of the *emm*32.2 strain 112327 was assessed by using blastn. The nucleotide sequences from the genes of interest were extracted from the refence genome of the GAS strain MGAS5005 (CP000017; GenBank) and compared to the collection of genes reported for the strain 112327 by Cornick et al 2017. The nucleotide sequences from *emm*32.2 (112327) genes were extracted from the genes clusters deposited at https://old.datahub.io/dataset/liverpool-gas. Matches covering more than 50% of the queried gene were considered as evidence of the presence of the gene in the genome of the *emm*32.2 strain (112327).

2.11 Preparation of secretome samples for label-free quantitative proteomics.

2.11.1 Sample collection

At the designated time points, separate aliquots were collected from the cultures for proteomic analysis as described elsewhere [164], with minor modifications. GAS strains were cultured in 8 ml THY-Glucose in a universal tube solution containing

$1x10^5$ of each strain, from frozen aliquots. Each strain was grown in duplicate at 37ºC, and samples were collected after 4h, 6h, and 10h corresponding to early exponential phase, mid-exponential phase, and late exponential phase, respectively. The cultures were harvested by centrifugation (3,000 rpm, 10 min, 4°C). The supernatant containing the secreted GAS proteome was collected and sterile filtered (0.22 µm, Sartorius, UK). The supernatant was stored at -80°C until protein extraction could be performed.

2.11.2 Preparation of secretome samples for label-free quantitative proteomics

Proteomic data construction was performed by Dr. Stuart Armstrong, at the University of Liverpool, U.K. Briefly, secreted proteins were precipitated by addition of trichloroacetic acid (TCA) (final TCA conc. 14%, v/v) and incubated on ice for 2 hours. The sample was centrifuged at 12,000 x g for 10 minutes at 4°C and the supernatant discarded. The pellet was washed three times with acetone and allowed to dry at RT. The pellet was then solubilized in 50mM ammonium bicarbonate, 0.1% (w/v) RapiGest (Waters). Total protein concentration was measured by Pierce™ Coomassie Plus (Bradford) Assay Kit assay (Thermo scientific, Hampshire UK). Sample protein content and volume was normalised with 50mM ammonium bicarbonate. Samples were then heated at 80^0C for 10 minutes, reduced with 3 mM dithiothreitol (Sigma, Dorset UK) at 60^0C for 10 minutes then alkylated with 9 mM iodoacetamide (Sigma, Dorset UK) at RT for 30 minutes in the dark. Proteomic grade trypsin (Sigma, Dorset UK) was added at a protein: trypsin ratio of 50:1 and samples incubated at 37^0C overnight. Rapigest was removed by adding TFA to a final

concentration of 1% (v/v) and incubating at 37ºC for 2 hrs. Peptide samples were centrifuged at 12,000 x g for 30 min (at 4°C) to remove the precipitated Rapigest.

2.11.3 NanoLC MS ESI MS/MS analysis

Peptides were analysed by on-line nanoflow LC using the Ultimate 3000 nanosystem (Dionex/Thermo Fisher Scientific). Samples were loaded onto a trap column (Acclaim PepMap 100, 2 cm × 75 μm inner diameter, C18, 3 μm, 100 Å) at 5 μl min-1 with an aqueous solution containing 0.1 %(v/v) TFA and 2%(v/v) acetonitrile. After 3 min, the trap column was set in-line an analytical column (Easy-Spray PepMap® RSLC 50 cm × 75 μm inner diameter, C18, 2 μm, 100 Å) fused to a silica nano-electrospray emitter (Dionex). The column was operated at a constant temperature of 30°C and the LC system coupled to a Q-Exactive mass spectrometer (Thermo Fisher Scientific, UK). Chromatography was performed with a buffer system consisting of 0.1 % formic acid (buffer A) and 80 % acetonitrile in 0.1 % formic acid (buffer B). The peptides were separated by a linear gradient of 3.8 – 50 % buffer B over 90 minutes at a flow rate of 300 nl/min. The Q-Exactive was operated in data-dependent mode with survey scans acquired at a resolution of 70,000. Up to the top 10, most abundant isotope patterns with charge states +2 to +5 from the survey scan were selected with an isolation window of 2.0Th and fragmented by higher energy collisional dissociation with normalised collision energies of 30. The maximum ion injection times for the survey scan and the MS/MS scans were 250 and 100ms respectively, and the ion target value was set to 1E6 for survey scans and 1E5 for the MS/MS scans. MS/MS events were

acquired at a resolution of 35,000. Repetitive sequencing of peptides was minimised through dynamic exclusion of the sequenced peptides for 20s.

2.11.4 Protein Identification and Quantification

Thermo RAW files were imported into Progenesis QI for proteomics (version 4.1, Nonlinear Dynamics). Progenesis aligns LC-MS/MS data across samples and compares them against a reference sample, peptide ions are detected, and a 2D map is created for each sample. The abundance of each peptide ion is then calculated. Runs were time aligned using default settings and using an auto-selected run as reference. Peaks were picked by the software using default settings and filtered to include only peaks with a charge state between +2 and +7. Spectral data were converted into .mgf files with Progenesis QI for proteomics and exported for peptide identification using the Mascot (version 2.3.02, Matrix Science) search engine. The mascot software attempts to find the best match in the database for each peptide mass and associated MS/MS spectra. Each entry in the database is digested in silico and the masses of the expected peptides calculated. If a calculated peptide mass matches the experimental one, the mass values expected to result from the fragmentation of the peptide are calculated, and the degree of matching to the peaks in the MS/MS spectrum scored. Mascot uses probability based matching for protein identification. It calculates the probability that the observed match between the experimental data, and mass values calculated from a candidate peptide or protein sequence, is a random event. This allows rejection of anything with a probability greater than a chosen threshold, e.g., 0.05 or 0.01. In addition to the calculated peptide scores, Mascot also estimates the

False Discovery Rate (FDR) by searching against a decoy database. The search is repeated using identical search parameters, against a database in which the sequences have been reversed or shuffled. There should be no real matches from the decoy database. The number of matches that are found in the decoy database is a good estimate of the number of false positives in the results from the target database. We use a FDR <1%. Tandem MS data were searched against translated ORFs from GAS M1 SF370 pan proteome (6,309 sequences; 1,519,142 residues, UniProt, May 2017). The search parameters were as follows: precursor mass tolerance was set to 10 ppm, and fragment mass tolerance was set as 0.05Da. Two missed tryptic cleavages were permitted which is standard practice. Carbamidomethylation (cysteine) was set as a fixed modification and oxidation (methionine) set as variable modification. Mascot search results were further validated using the machine learning algorithm Percolator embedded within Mascot. Percolator distinguishes between a subset of the high-scoring matches from the target database, assumed correct, and the matches from the decoy database assumed incorrectly. Using Percolator improves the discrimination between correct and incorrect spectrum identifications and can give very substantial improvements in sensitivity. The Mascot decoy database function was utilised, and the false discovery rate was <1%, while individual percolator ion scores >13 indicated identity or extensive homology ($p < 0.05$). Mascot search results were imported into Progenesis QI for proteomics as XML files. Peptide intensities were normalised against the reference run by Progenesis QI for proteomics, and these intensities were used to highlight relative differences in protein expression between sample groups. Only proteins with 2 or more identified peptides were included in the dataset.

Statistical analysis (ANOVA) of the data was performed using Progenesis QI for proteomics to identify significantly ($p<0.05$, $q \leq 0.05$, relative fold change ≥ 2) differentially expressed proteins. PCA plots were created using Progenesis. Hierarchal clustering and heatmaps were created with the Morpheus online tool https://software.broadinstitute.org/morpheus/.

2.12 Statistics

Unless otherwise stated, the data was analysed in GraphPad Prism and compared by one-way, two-way ANOVA and t-test. The Tukey-Kramer and Kruskal-Wallis multiple-comparison test was used to identify significant differences between individual groups. Data is expressed throughout as mean ± standard error of the mean (S.E.M.). Differences between data were considered significant if $p < 0.05$ Moreover, Pearson correlation coefficient was used to measure the linear correlation between two variables.

Chapter 3 Phenotypic Analysis

3.1 Introduction

Infections caused by invasive GAS are notifiable to Public Health England. These records showed that Merseyside had an outbreak of invasive GAS involving *emm*32.2 strains between 2010 and 2012. The outbreak was unusual because it had a 29% mortality rate which is typically higher than usual, was confined to adults and had a propensity to occur in IV drug users, the homeless or in alcoholics. In total, 14 *emm*32.2 strains have been isolated and sequenced. Genetic analysis has identified 19 genes that were unique to the outbreak isolates [146]. Noteworthy, the phenotypic attributes of the *emm*32.2 strains which have been identified as the cause of the outbreak are not understood. Hence, the purpose of the present chapter was to use *in vitro* models to gain insight into *emm*32.2's capability to cause invasive disease.

The underlying aim of the research was to conduct *in vitro* examinations of the *emm*32.2 outbreak strain to identify whether there are differences in bacterial virulence that can be linked to the genomic data and may explain the epidemiological features of this *emm* subtype. In order to determine the virulent phenotype of outbreak strains (*emm*32.2), I evaluated several virulence factors as potential candidates responsible for the induction of invasive GAS infections. The investigation initially began by performing capsule assay, complement deposition assay, proteinase activity assay,

NADase and SLO activity assay, and biofilm formation assay, of 14 invasive isolates of *emm*32.2 to invasive and 10 non-invasive isolates (including *emm*6.0, *emm*89.0, and *emm*1.0).

3.2 Results

3.2.1 Capsule thickness assay

The bacterial capsule is essential virulence factor for GAS and most GAS *emm* types are coated within a capsule composed of hyaluronic acid (HA) [73]. Measuring the capsule thickness may assist to determine the relative ability of the isolate to adhere, colonise and invade the host. The assay presented below is used to determine bacterial cell size, which correlates with capsule thickness. I compared capsule sizes of 24 GAS strains from diverse *emm* types such as *emm*32.2, *emm*6.0, *emm*89.0 and *emm*1.0.

As shown in Figure 3-1A, FITC-Dextran assays are a method of providing a measurement of two-dimensional surface area of bacteria, theoretically positively correlates with capsule thickness [152]. Results obtained demonstrated differences in capsule thickness of among the strains of GAS, and there was a statistically significant difference between strains (One-way ANOVA, followed by Kruskal-Wallis post hoc tests p=< 0.0001) (Table 3-1). We observed more than one group in *emm*32.2 based on pixel measurement: high capsule (>400) (112844 and 112327) and low capsule (<200) (102920, 111617, 121324, 101967, 101008, 101700, 102029, 102434, 121511, and 122397). In the *emm*6 type, there was no statistically significant difference between invasive (126215) and non-invasive strains (127785, 127784, and 137027). However, *emm*89.0 showed a significant difference between invasive isolate (101724) and non-invasive strains (127746, 137081, and 137072) (One-way ANOVA, followed by Kruskal-Wallis post hoc test p=0.02). Similarly, a significant difference between

invasive isolate (101910) and non-invasive isolate (127251) was presented in *emm*1.0 (Student t-test, p=0.004).

In conclusion, although *emm*32.2 strains belong to the same *emm* type, they displayed variation in capsule.

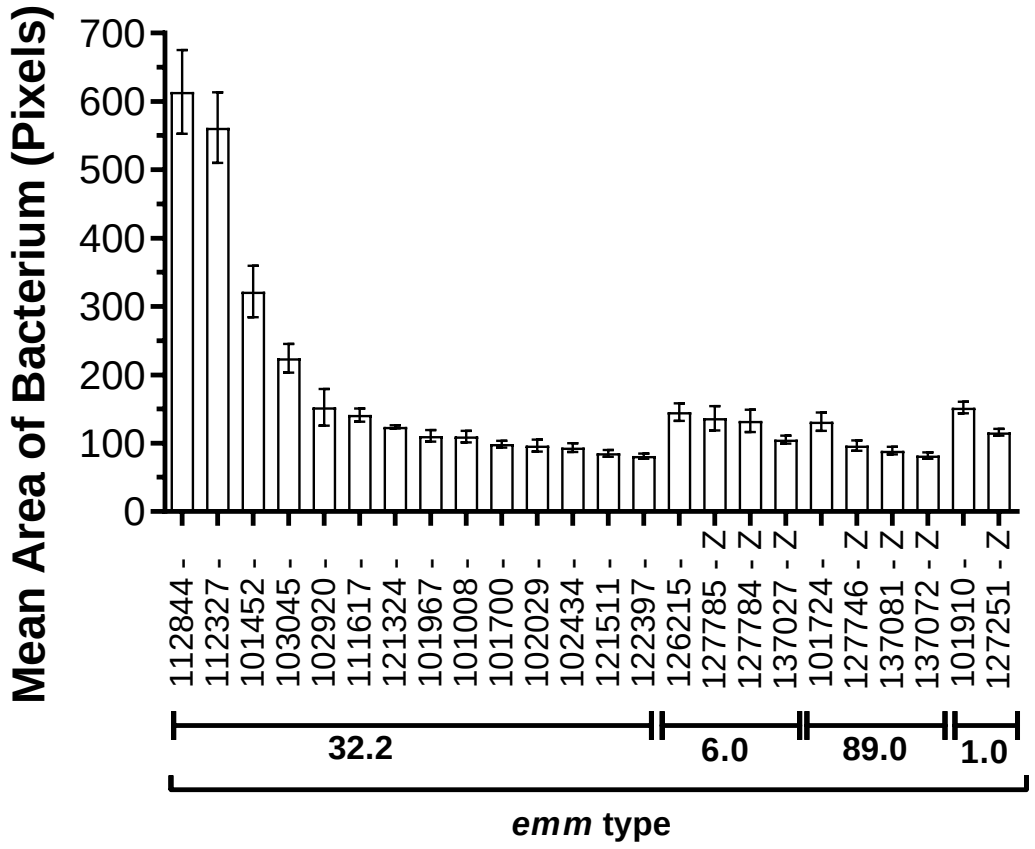


Figure 3-1 FITC-Dextran exclusion assay.

(A) FITC-dextran exclusion assay of *emm*32.2, *emm*6.0, *emm*89.0 and *emm*1.0 *In vitro* capsule thickness as measured using FITC-dextran. Acquired images were analysed in a blinded fashion by a single assessor. Images were acquired in two independent experiments. In each, three slides were produced for each strain, with every slide yielding three images from different areas of the slide. The area in pixels was calculated for individual (i.e., non-clumped) cells; an average number of 442 individual cells per strain were analysed (range: 347 – 597, total: 10754). Results are presented as mean ± S.E. (One-way ANOVA, followed by Kruskal-Wallis post hoc tests p=< 0.0001). All strains analysed in triplicate in tow independent assays. N, non-invasive.

Table 3-1 Pair wise comparison of isolates used in capsule thickness, analysed using a one-way ANOVA and Kruskal-Wallis multiple comparisons test. ns= not significant

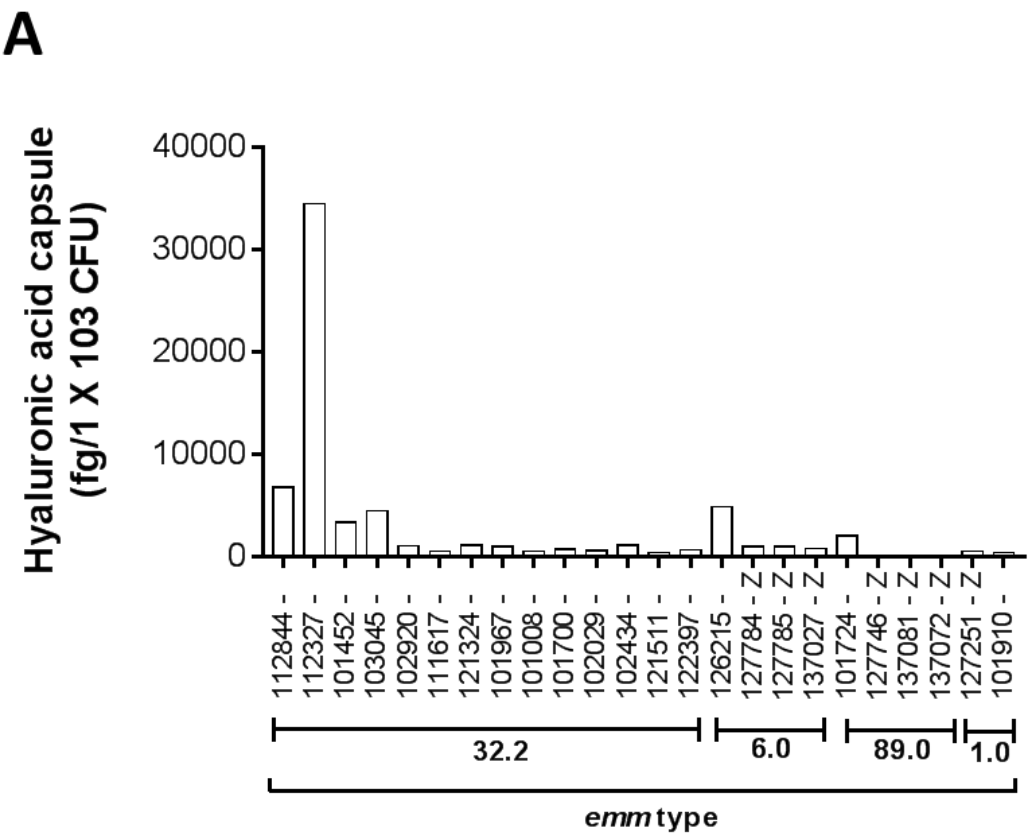
	112844	112327	101452	103045	102920	111617	121324	101967	101008	101700	102029	102434	121511	122397	126215	127785	127784	137027	101724	127746	137081	137072	101910	127251
112844		ns	ns	ns	ns	ns	ns	ns	0.044	0.019	0.01	<.001	<.001	ns	ns	ns	ns	ns	0.023	0.002	<.001	ns	ns	ns
112327	ns		ns	ns	ns	ns	ns	ns	0.05	0.021	0.012	<.001	<.001	ns	ns	ns	ns	ns	0.026	0.003	<.001	ns	ns	ns
101452	ns	ns		ns	ns	ns	ns	ns	0.05	0.021	0.012	0.005	0.002	ns	ns	ns	ns	ns	ns	0.016	0.002	ns	ns	ns
103045	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	0.023	0.008	ns	ns	ns	ns	ns	ns	ns	0.009	ns	ns	ns
102920	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
111617	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
121324	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
101967	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
101008	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
101700	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
102029	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
102434	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
121511	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
122397	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
126215	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns
127785	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns
127784	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns
137027	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns
101724	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns
127746	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns
137081	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns
137072	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns
101910	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns
127251	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	

3.2.2 Quantification of capsular hyaluronic acid

Next, we sought to compare a validated, but expensive ELISA-based method of assessing capsule thickness with an in-house and cheaper fluorescence microscopy-based method. Therefore, the amount of Hyaluronic Acid (HA) in the capsule of *emm*32.2, *emm*1.0, *emm*89.0, and *emm*6.0 GAS strains were quantified using an ELISA (Corgenix HA Test Kit).

We observed GAS isolates present a highly heterogeneous HA content distribution and a high level of variation occurred within *emm*32.2 strains (Figure 3-2A). When comparing FITC dextran and ELISA-based method, a correlation between FITC dextran capsule results and HA content results was observed (Persons Rank: r=0.7543, 95%Cl: 0.4887-0.8832, r^2 0.5554, p<0.0001) (Figure 3-2B). However, 3 *emm*89.0 type isolates were found not to be aligned with the linear fit trend curve.

Overall, our findings confirm previous observation that the *emm*32.2 outbreak cluster was phenotypically heterogeneous.


A
Hyaluronic acid capsule (fg/1 X 103 CFU)
40000
30000
20000
10000
0
112844
112327
101452
103045
102920
111617
121324
101967
101008
101700
102029
102434
121511
122397
126215
127784 - Z
127785 - Z
137027 - Z
101724
127746 - Z
137081 - Z
137072 - Z
127251 - Z
101910
32.2
6.0
89.0
1.0
emm type

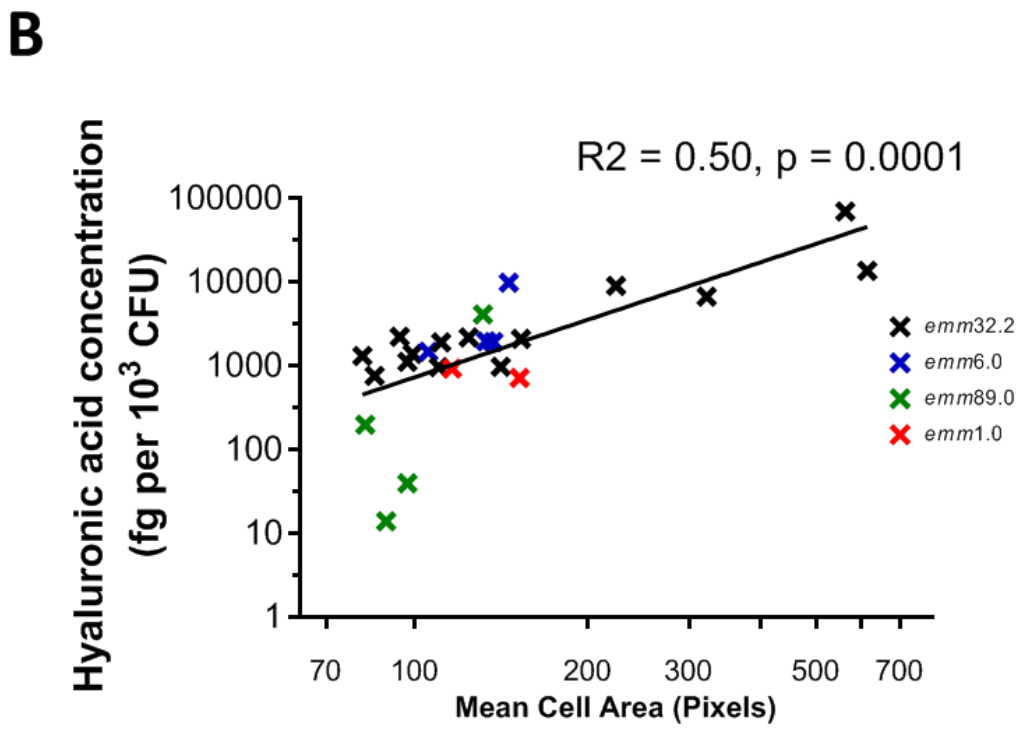


Figure 3-2 Quantification of capsular hyaluronic acid.

(A) Production of hyaluronic acid was measured for *emm*32.2, *emm*6.0, *emm*89.0 and *emm*1.0 strains using an ELISA-based assay. Hyaluronic acid capsule was measured as femtograms per 1×10^3 CFU. Results are presented as mean ± S.E. All strains were analysed in duplicate in one assay. B). Comparison plot between capsule thickness measured by zone of exclusion of FITC-dextran and capsular hyaluronic acid content, as plotted on $\log_{10}$ axes. The Pearson correlation coefficient (r) was equal to 0.7085 (95% CI: 0.4271 to 0.8648; r2: 0.5020; p = 0.0001). N, non-invasive.

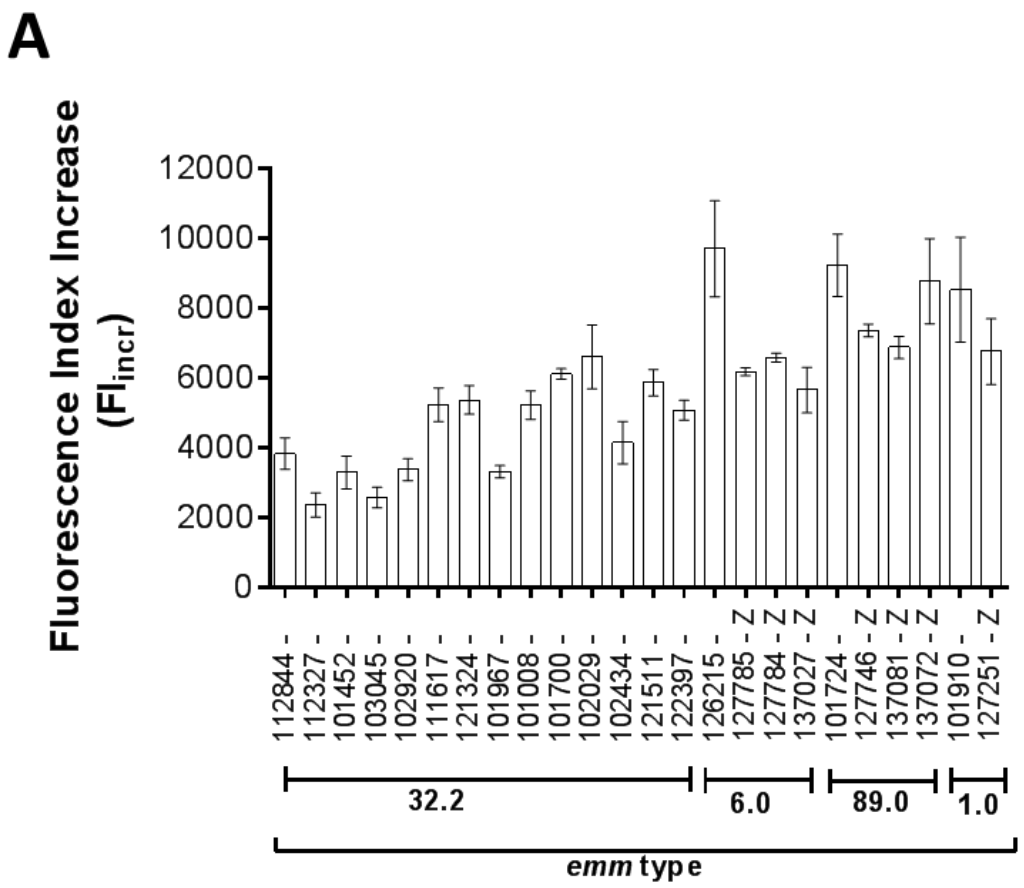
3.2.3 Resistance to complement deposition

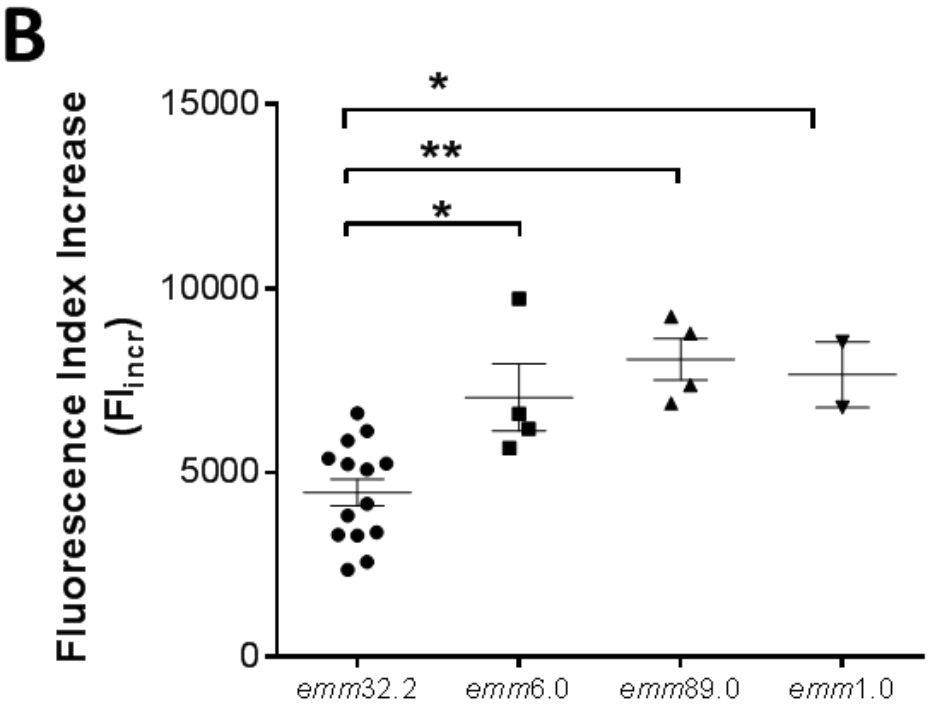
The immune complement mediates opsonophagocytosis, and GAS was documented for its ability to inhibit complement deposition through its HA capsule [11]. A small amount of complement deposition would thus suggest that the pathogen in question is more likely to evade immune phagocytosis.

Our results demonstrated that there were wide variations in complement deposition between GAS strains (***Figure*** **3-**3A), with statistically significant differences (Table 3-2) (One-way ANOVA, followed by Tukey's post-hoc test, $p<0.0001$). The result also illustrated that there were wide variations in sensitivity to complement deposition between *emm*32.2 strains.

Interestingly, *emm*32.2 strains would be divided into three groups depending on the amount of complement deposition. The strains from the first group (low: < 4000) (112844, 112327, 103045, 101452, 102920, and 101967) were more complement-resistant, while the second group (medium 4000-6000) of strains (1012434, 121324, 122397, 111617, 101008, and 121511,) showed considerably higher amounts of complement deposition. Taken together, these results suggested that there might be variations of virulence among *emm*32.2 strains. Unlike *emm*89.0 and *emm*1.0 type, *emm*6 type revealed a statistical difference between invasive (126215) and non-invasive strains (127785, 127784, and 137027) (One-way ANOVA, followed by Tukey's post-hoc test, $p=0.003$).

Overall, the C3-binding activity on the bacterial surface of *emm*32.2 outbreak strain was significantly lower than those in *emm*1.0, *emm*89.0, and *emm*6.0 (***Figure*** 3-3B) .Resistance to complement deposition was found to be statistically significantly correlated with capsule thickness when comparing *emm*32.2 strains and *emm*1.0, *emm*89.0, and *emm*6.0 strains (***Figure*** 3-3C) ($p = 0.0027$). All in all, *emm*32.2 were determined to be most resistance to complement deposition, which may thus explain its virulence properties.


A
Fluorescence Index Increase
(FIincr)
12000
10000
8000
6000
4000
2000
0
112844 -
112327 -
101452 -
103045 -
102920 -
111617 -
121324 -
101967 -
101008 -
101700 -
102029 -
102434 -
121511 -
122397 -
126215 - Z
127785 - Z
127784 - Z
137027 - Z
101724 - Z
127746 - Z
137081 - Z
137072 - Z
101910 -
127251 - Z
32.2
6.0
89.0
1.0
emm type


B
Fluorescence Index Increase
(FIincr)
15000
10000
5000
0
*
**
*
emm32.2
emm6.0
emm89.0
emm1.0

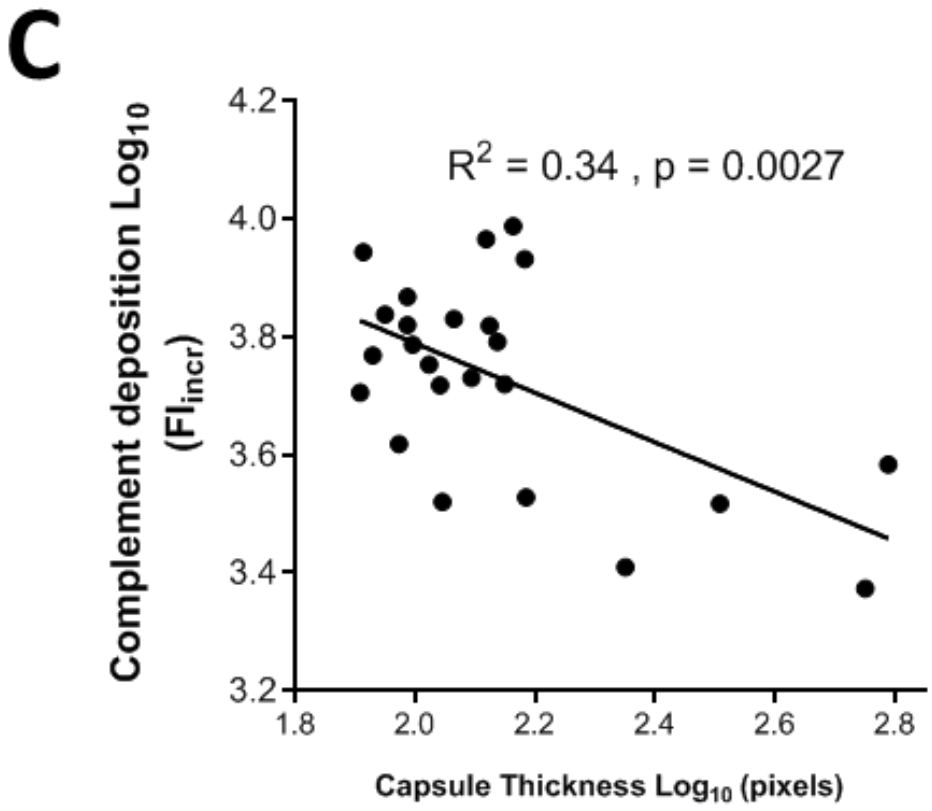


Figure 3-3 Complement deposition on Group A Streptococcus surface.

(A) Complement deposition of *emm*32.2, *emm*6.0, *emm*89.0 and *emm*1.0 Results are expressed as Fluorescence Index increase (FI increase) = Mean fluorescence intensity (MFI) x % bacteria positive for C3b depositions. This calculation takes into account both binding intensity and proportion of bacteria affected. All strains were analysed in triplicate in three independent assays. (One-way ANOVA, followed by Tukey's post-hoc test, $p<0.0001$) (B) Comparison of mean FI_{incr} between *emm*32.2 strains and *emm*6.0, *emm*89.0, and *emm*1.0 strains. Values are plotted as mean FI_{incr} (One-way ANOVA, followed by Tukey's post-hoc test, $p=0.0003$). (C) Comparison between capsule thickness (Pixels) and complement deposition (mean FI_{incr}) on a logarithmic scale plot. The Pearson correlation coefficient (r) after log_{10} transformation was -0.5845 (95% CI: -0.7994 to -0.2369; $r^2 = 0.34$; $p = 0.0027$). N, non-invasive.

Table 3-2 Pair wise comparison of isolates used in complements deposition, analysed using One-way ANOVA, followed by Tukey's post-hoc test. Significance indicated by adjusted p value. ns=not significant

	112844	112327	101452	103045	102920	111617	121324	101967	101008	101700	102029	102434	121511	122397	126215	127785	127784	137027	101724	127746	137081	137072	101910	127251
112844		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	<.001	ns	ns	ns	<.001	0.043	ns	<.001	<.001	ns	
112327			ns	ns	ns	ns	ns	ns	0.019	0.003	ns	0.047	ns	<.001	0.015	0.003	ns	<.001	<.001	<.001	<.001	<.001	0.001	
101452				ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	<.001	ns	ns	ns	<.001	0.006	0.035	<.001	<.001	ns	
103045					ns	ns	ns	ns	0.041	0.007	ns	ns	ns	<.001	0.033	0.007	ns	<.001	<.001	0.002	<.001	<.001	0.003	
102920						ns	ns	ns	ns	ns	ns	ns	ns	<.001	ns	ns	ns	<.001	0.008	0.047	<.001	<.001	ns	
111617							ns	ns	ns	ns	ns	ns	ns	0.001	ns	ns	ns	0.008	ns	ns	0.042	ns	ns	
121324								ns	ns	ns	ns	ns	ns	0.002	ns	ns	ns	0.013	ns	ns	ns	ns	ns	
101967									ns	ns	ns	ns	ns	<.001	ns	ns	<.001	<.001	0.006	0.038	ns	<.001	ns	
101008										ns	ns	ns	ns	<.001	ns	ns	ns	0.007	ns	ns	0.04	ns	ns	
101700											ns	ns	ns	0.0035	ns	ns	ns	ns	ns	ns	ns	ns	ns	
102029												ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	
102434													ns	<.001	ns	ns	ns	<.001	ns	ns	<.001	0.001	ns	
121511														0.014	ns	ns	ns	ns	ns	ns	ns	ns	ns	
122397														<.001	ns	ns	ns	0.004	ns	ns	0.024	ns	ns	
126215															0.044	ns	0.006		ns	ns	ns	ns	ns	
127785																ns	ns	ns	ns	ns	ns	ns	ns	
127784																	ns	ns	ns	ns	ns	ns	ns	
137027																		0.037	ns	ns	ns	ns	ns	
101724																			ns	ns	ns	ns	ns	
127746																				ns	ns	ns	ns	
137081																					ns	ns	ns	
137072																						ns	ns	
101910																							ns	
127251																								

3.2.4 Casein-plate assay

Cysteine proteinase such as SepB is one of the most studied GAS virulence factors, but its role in pathogenesis is still not fully understood [95]. In order to examine the significance of proteinase, I aimed to determine the expression of extracellular cysteine protease via a plate assay.

A simple plate assay based on digestion of skim milk contained within Columbia agar base medium was found to be a sensitive and reproducible method for detection of proteinase production by individual surface has grown colonies of GAS [155]. Using this method, proteinase activity was determined on 24-h anaerobic GAS cultures: 16 (67 %) of the 24 GAS strains tested were proteinase-positive, while no protease activity was not detected in the other 8 (33 %) (Table 3-3). Although our results showed a high proportion of the GAS strains in our collection present proteinase activity, we found that 8 out of 14 *emm*32.2 strains were protease-negative.

Table 3-3 Proteolytic activity of Group A Streptococcus.

Strain	*emm* type	Disease category	Outcome
112844	**32.2**	**invasive**	**Positive**
112327	**32.2**	**invasive**	**Negative**
101452	**32.2**	**invasive**	**Positive**
103045	**32.2**	**invasive**	**Positive**
102920	**32.2**	**invasive**	**Positive**
111617	**32.2**	**invasive**	**Negative**
121324	**32.2**	**invasive**	**Positive**
101967	**32.2**	**invasive**	**Negative**
101008	**32.2**	**invasive**	**Negative**
101700	**32.2**	**invasive**	**Negative**
102029	**32.2**	**invasive**	**Negative**
102434	**32.2**	**invasive**	**Negative**
121511	**32.2**	**invasive**	**Positive**
122397	**32.2**	**invasive**	**Negative**
126215	**6.0**	**invasive**	**Positive**
127785	**6.0**	**non-invasive**	**Positive**
127784	**6.0**	**non-invasive**	**Positive**
137027	**6.0**	**non-invasive**	**Positive**
101724	**89.0**	**invasive**	**Positive**
127746	**89.0**	**non-invasive**	**Positive**
137081	**89.0**	**non-invasive**	**Positive**
137072	**89.0**	**non-invasive**	**Positive**
101910	**1.0**	**invasive**	**Positive**
127251	**1.0**	**non-invasive**	**Positive**

3.2.5 NADase and SLO activity

GAS produces many virulence factors that facilitate transmission, colonisation, invasion, and bacterial dissemination [11]. streptococcal NAD+ glycohydrolase (SPN) and streptolysin O (SLO) are two potent cytotoxins secreted by GAS, and multiple roles in virulence have been attributed to each protein. To characterise SPN and SLO production, we measured and compared the NADase Activity and SLO activity present in culture supernatants obtained from our collection of 24 GAS strains using NADase assay and SLO haemolytic assay which previously described [36].

3.2.6 NADase Activity Assay

As shown in Figure 3-4, NADase activity was present in the supernatants of *emm*1.0, *emm*89.0, and *emm*6.0 strains versus no activity detected in the *emm*32.2 strains (One-way ANOVA, followed by Tukey's post-hoc test, P= < 0.0001) (Table 3-4).There was heterogeneity and significant variation between the *emm*89.0 strains (p <0.0001). In the *emm*6.0 and *emm*1.0 strains, no significant differences were determined. The distributions of GAS strains with and without NADase activity were not proportional within the invasive strains: 88 % of the invasive strains lacked NADase activity. Since all *emm*32.2 type strains presented no detectable NADase activity, we concluded that this might be one of the factors contributing to the invasiveness of the outbreak strain.

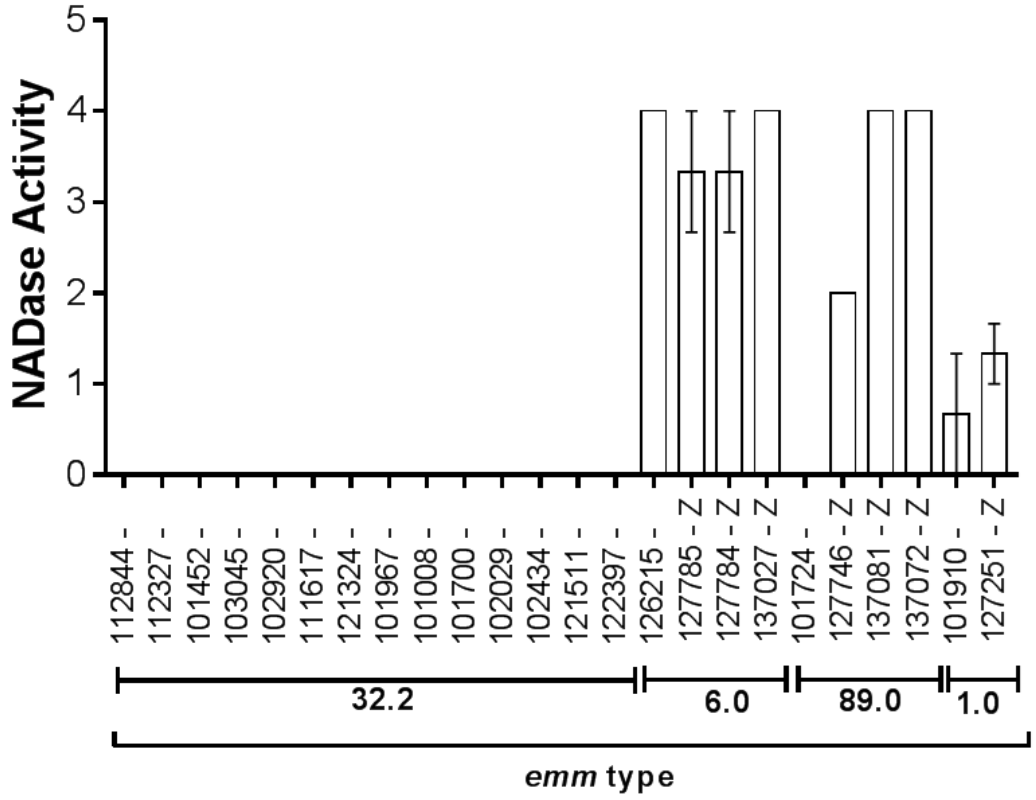


Figure 3-4 NADase activities of 24 Group A Streptococcus strains

The results are represented as means ± S.E.M of three experiments and corresponds to the highest 2-fold dilution that is able to fully hydrolyse 100 nmol of NAD+ (One-way ANOVA, followed by Tukey's post-hoc test, $p<0.0001$

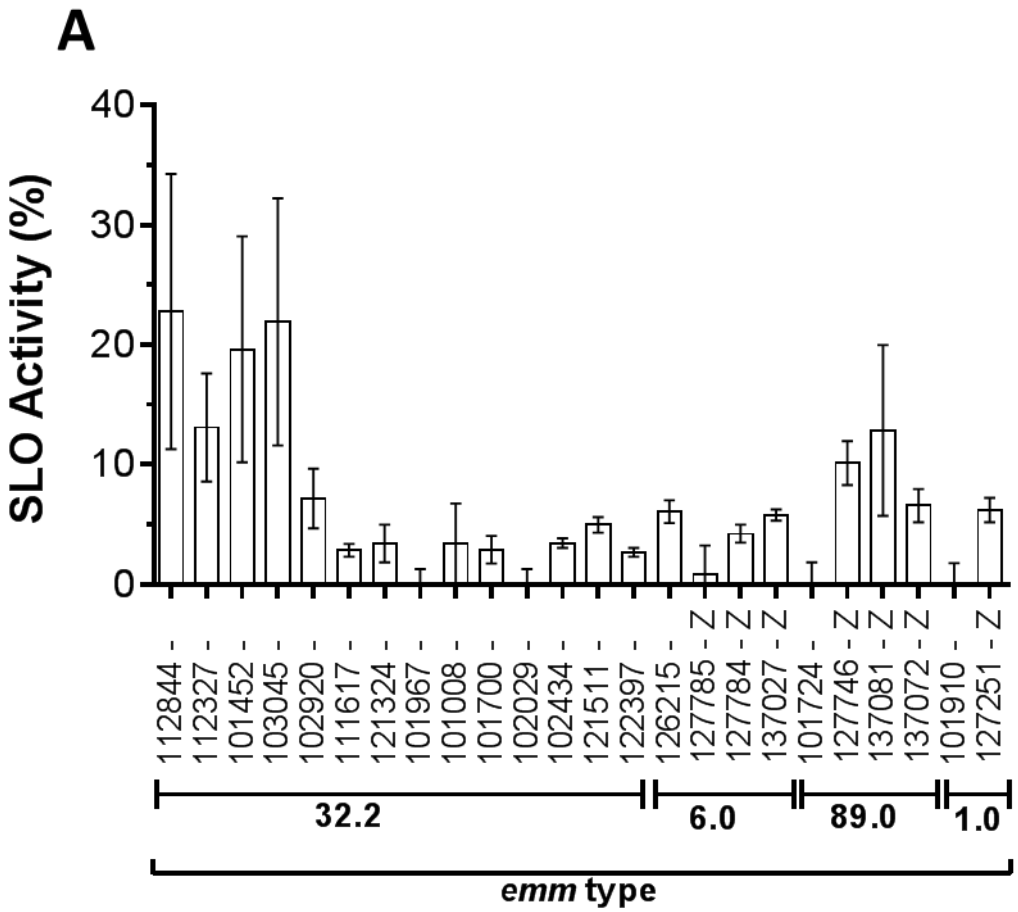
Table 3-4Pair wise comparison of isolates used in NADase Activity, analysed using One-way ANOVA, followed by Tukey's post-hoc test. Significance indicated by adjusted p value. ns=not significant

	112844	112327	101452	103045	102920	111617	121324	101967	101008	101700	102029	102434	121511	122397	126215	127785	127784	137027	101724	127746	137081	137072	101910	127251
112844		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	<.001	<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
112327			ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	<.001	<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
101452				ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	<.001	<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
103045					ns	ns	ns	ns	ns	ns	ns	ns	ns	<.001	<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
102920						ns	ns	ns	ns	ns	ns	ns	ns	<.001	<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
111617							ns	ns	ns	ns	ns	ns	ns	<.001	<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
121324								ns	ns	ns	ns	ns	ns	<.001	<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
101967									ns	ns	ns	ns	ns	<.001	<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
101008										ns	ns	ns	ns	<.001	<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
101700											ns	ns	ns	<.001	<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
102029												ns	ns	<.001	<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
102434													ns	<.001	<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
121511														<.001	<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
122397															<.001	<.001	<.001	ns	<.001	<.001	<.001	ns	ns	ns
126215																ns	ns	<.001	<.001	ns	ns	<.001	<.001	<.001
127785																	ns	<.001	ns	ns	ns	<.001	<.001	<.001
127784																		ns	<.001	ns	ns	<.001	<.001	<.001
137027																			<.001	<.001	<.001	<.001	<.001	<.001
101724																				<.001	<.001	ns	ns	ns
127746																					<.001	ns	ns	ns
137081																						ns	<.001	<.001
137072																							<.001	<.001
101910																								ns
127251																								

3.2.7 SLO Activity Assay

The haemolytic activity of Streptolysin O produced by *emm*32.2 strains was compared to that produced by other *emm* type strains. SLO haemolytic activity relies on the toxin's ability to lyse Red Blood Cell (RBCs). In Figure 3-5A, the results revealed that there were wide variations in SLO activity amongst the GAS tested strains. As a result, the strains would be divided into three groups depending on the percentage of SLO haemolytic activity: low (<5%) - medium (5%-10%) – high (>10%). The strains from the first group (111617, 121324, 101967, 101008, 101700, 102029, 102434, 121511, 122397, 127785, 127784, 101724, and 101910) produced low amount of SLO, while the third group (high) of strains (112844, 112327, 101452, 103045, 127746, and 137081) showed higher amounts of SLO.

There were significant differences between 24 strains (One-way ANOVA, followed by Tukey's post-hoc test, $p=0.004$) as well as significant differences within the *emm*32.2 strains (One-way ANOVA, followed by Tukey's post-hoc test, $p=0.02$). While there were no significant differences within the *emm*6.0 and *emm*89.0 strains, there were important differences among the *emm*1.0 strains (Student t-test, $p=0.03$). We found a clear correlation between capsule thickness and SLO haemolytic activity whereby a thicker capsule was associated with a higher haemolytic activity ($p=0.0002$). Overall, because of the large variations in SLO activity amongst the *emm*32.2 strains, we could not make any association between SLO and induction of *emm*32.2 outbreak.


A
SLO Activity (%)
40
30
20
10
0
112844 -
112327 -
101452 -
103045 -
102920 -
111617 -
121324 -
101967 -
101008 -
101700 -
102029 -
102434 -
121511 -
122397 -
126215 -
127785 - Z
127784 - Z
137027 - Z
101724 -
127746 - Z
137081 - Z
137072 - Z
101910 -
127251 - Z
32.2
6.0
89.0
1.0
emm type

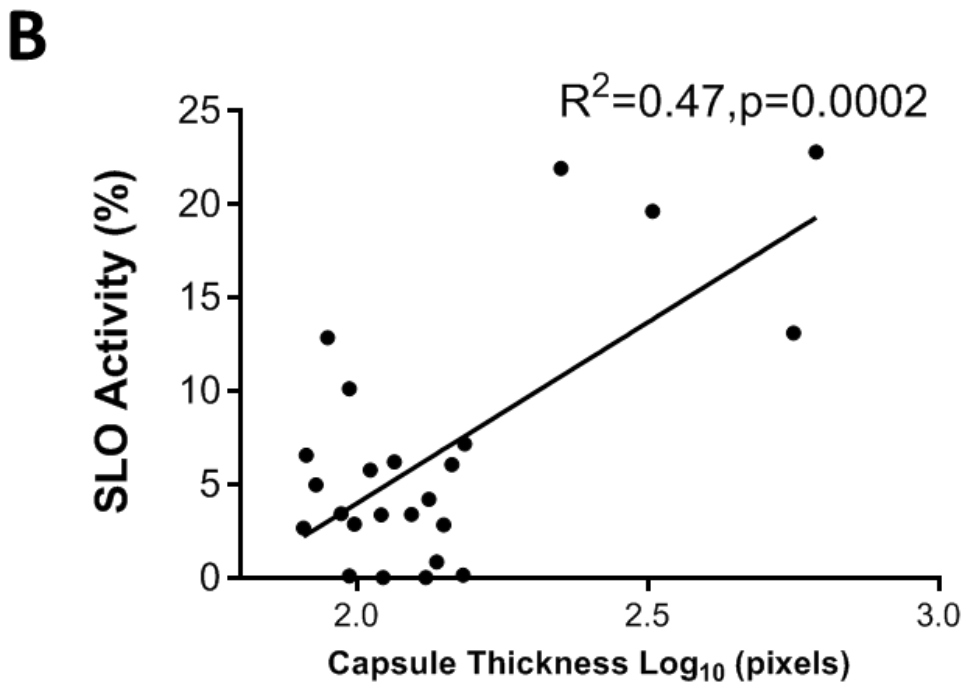


Figure 3-5 Streptolysin O haemolytic activity of Group A Streptococcus strains

Streptolysin O (SLO) activity was measured in accordance with its ability to lyse sheep erythrocytes. The results shown represent the percentage of activity relative to that of the 100% lysis positive control (H2O) and with 0% being equivalent to the activity measured in strain 101967, for which no lysis was observed. Results are reported as means ± S.E.M from three independent experiments (One-way ANOVA, followed by Tukey's post-hoc test, p=0.0043). (B) Upon comparison of the capsule thickness of the various *emm* types (*emm*32.2, *emm*6.0, *emm*89.0 and *emm*1.0), a statistically significant correlation was found whereby a thicker capsule presented higher haemolytic activity (p=0.0002). N, non-invasive.

3.2.8 GAS biofilm formation

Biofilm formation significantly contributes to GAS pathogenesis [72]. Not only is biofilm formation considered a protective mechanism but it also is a critical mechanism for bacteria to survive and proliferate in hostile environments thus facilitating the maintenance of the infectious process [165]. As little is known about GAS biofilms and their contributions to human disease, we aimed to determine whether there are differences between *emm*32.2 and *emm*1.0, *emm*89.0, and *emm*6.0 strains in biofilm formation at different incubation time points, i.e., 24h, 48h, 96h, and 168h / 7d.

3.2.9 Biofilm assay at 24h

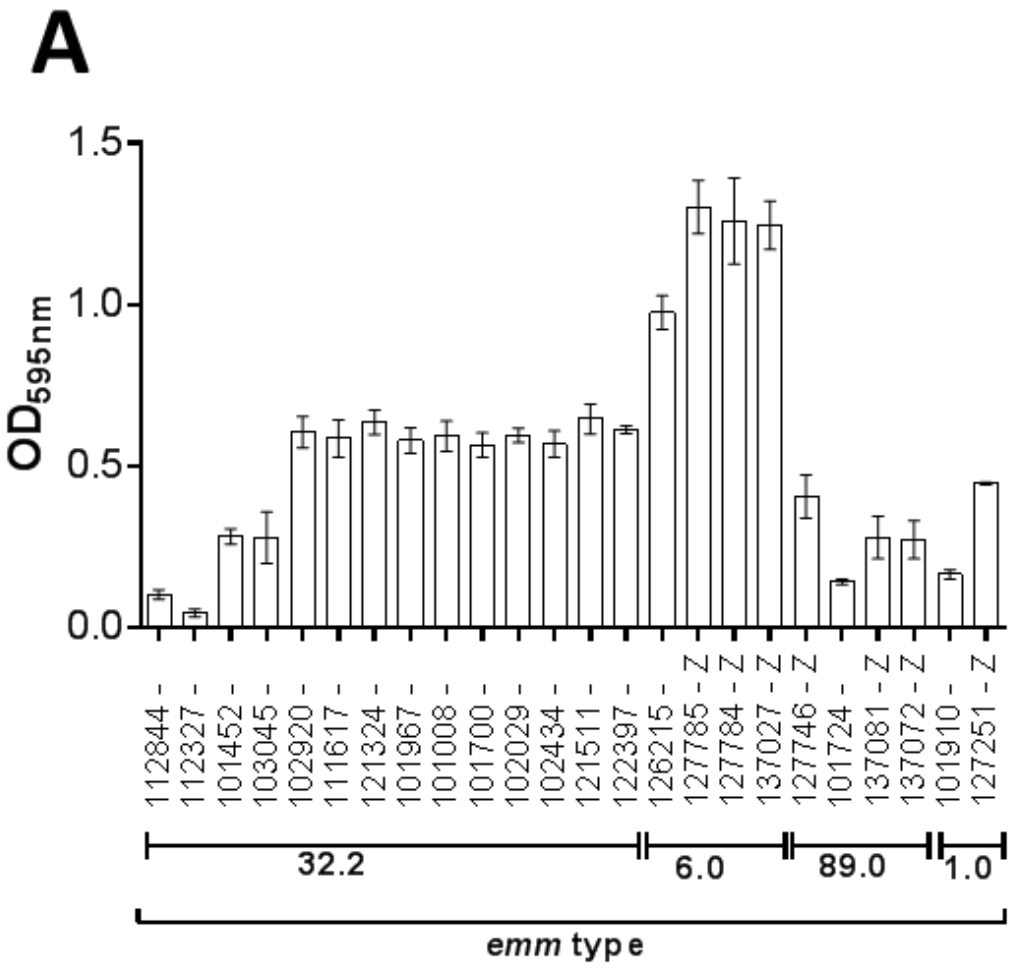
To date, it is not known whether *emm*32.2 strains included in this study are able to produce biofilms. To address this, biofilm formation was quantified by a standard microtiter plate crystal violet (CV) assay.

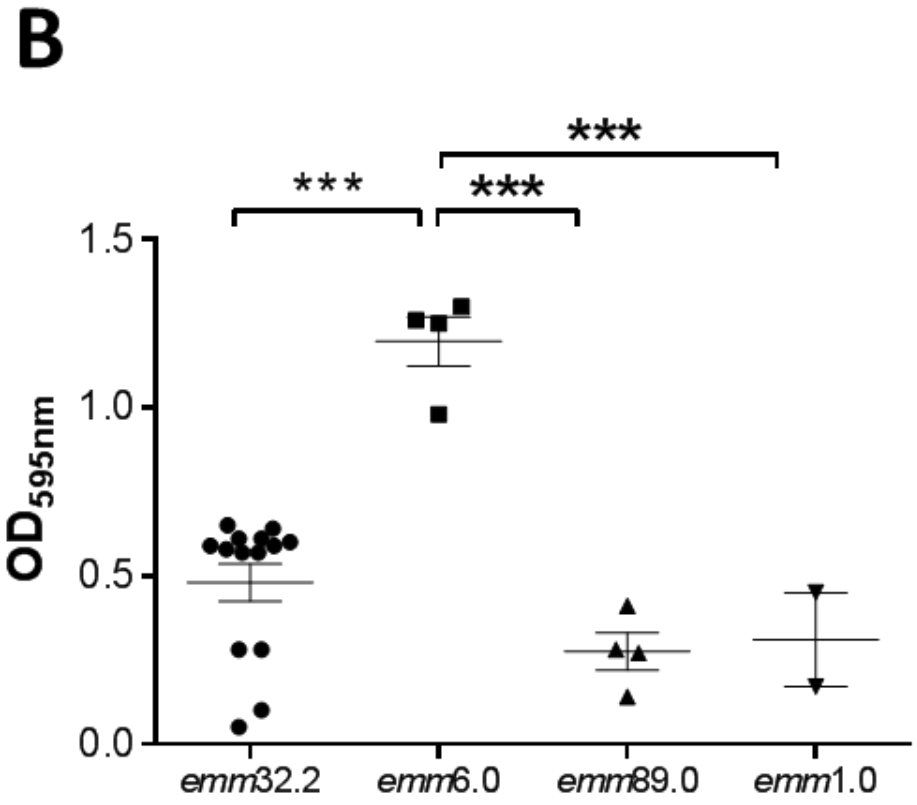
Figure 3-6 A is shown the detectable amounts of biofilm formed after 24h incubation in THB, on the basis of optical density readings. There was an obvious difference in the biofilm formation among *emm*32.2 strains in comparison to the other *emm*type strains, which may be categorized as biofilm producers. Although they had the same *emm* type, the 32.2*emm* strains exhibited phenotypic heterogeneity at statistically significant levels (One-way ANOVA, followed by Tukey's post-hoc test, $p< 0.0001$). The heterogeneity was such that we were able to divide the 32.2*emm* into 3 groups on the basis of OD_{595nm} measurement: low (< 0.2) (112844 and 112327) vs. medium (0.2-

0.5) (101452 and 103045) vs. high (>0.5) (102920, 111617, 121324, 101967, 101008, 101700, 102029, 102434, 121511, and 122397).

A different pattern of biofilm formation was seen in the *emm*6 type, which exhibited the highest levels of biofilm formation amongst all four *emm* types investigated. However, no significant differences were found between the 4 strains of the *emm*6 type at 24h.

A significant difference was seen between the invasive (101910) and non-invasive (127251) *emm*1.0 strains (One-way ANOVA, followed by Tukey's post-hoc test, $p<0001$). Finally, there were also variations in biofilm formation within the *emm*89.0 strains: the invasive (101724) strain exhibited a low amount of biofilm formation compared to non-invasive strains (127746, 137081 and 137072),however, there was no statistically significant difference between *emm*89.0 strains at 24h.


A
1.5
1.0
0.5
0.0
OD595nm
112844 -
112327 -
101452 -
103045 -
102920 -
111617 -
121324 -
101967 -
101008 -
101700 -
102029 -
102434 -
121511 -
122397 -
126215 -
127785 - Z
127784 - Z
137027 - Z
127746 - Z
101724 -
137081 - Z
137072 - Z
101910 -
127251 - Z
32.2
6.0
89.0
1.0
emm type


B
***
***
***
1.5
1.0
0.5
0.0
OD595nm
emm32.2
emm6.0
emm89.0
emm1.0

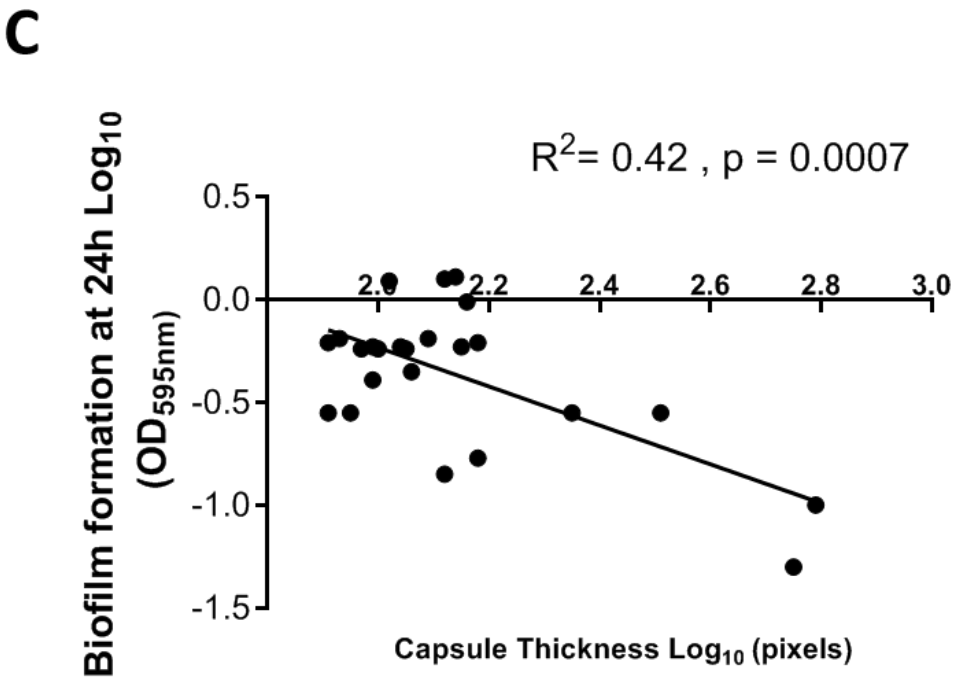


Figure 3-6 Biofilm formation of Group A Streptococcus at 24h

A total of 24 strains of *emm*32.2, *emm*6.0, *emm*89.0 and *emm*1.0 grown in THBY medium supplemented with glucose was assayed following incubation at 37°C for 24h. All biofilm data were normalised according to the mean OD_{595nm} of control wells containing only THYG media. In all biofilm studies, strains were analysed in eight replicate wells and in three independent experiments. Data are presented as means +/- SE (One-way ANOVA, followed by Tukey's post-hoc test, $p<0.0001$). (B) Comparison of biofilm formation between *emm*32.2 and *emm*6.0, *emm*89.0 and *emm*1.0 (One-way ANOVA, followed by Tukey's post-hoc test, $p = <0.001$) (C) Relationship between capsule thickness and biofilm on a logarithmic scale graph. The Pearson correlation coefficient (r) after log_{10} transformation was -0.6408 ($r^2 = 0.42$; $p = 0.0007$). On comparison of the capsule thickness of the various *emm* types with their biofilm forming ability at 24h, thick capsule strains were found to be poor biofilm formers. N, non-invasive.* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$.

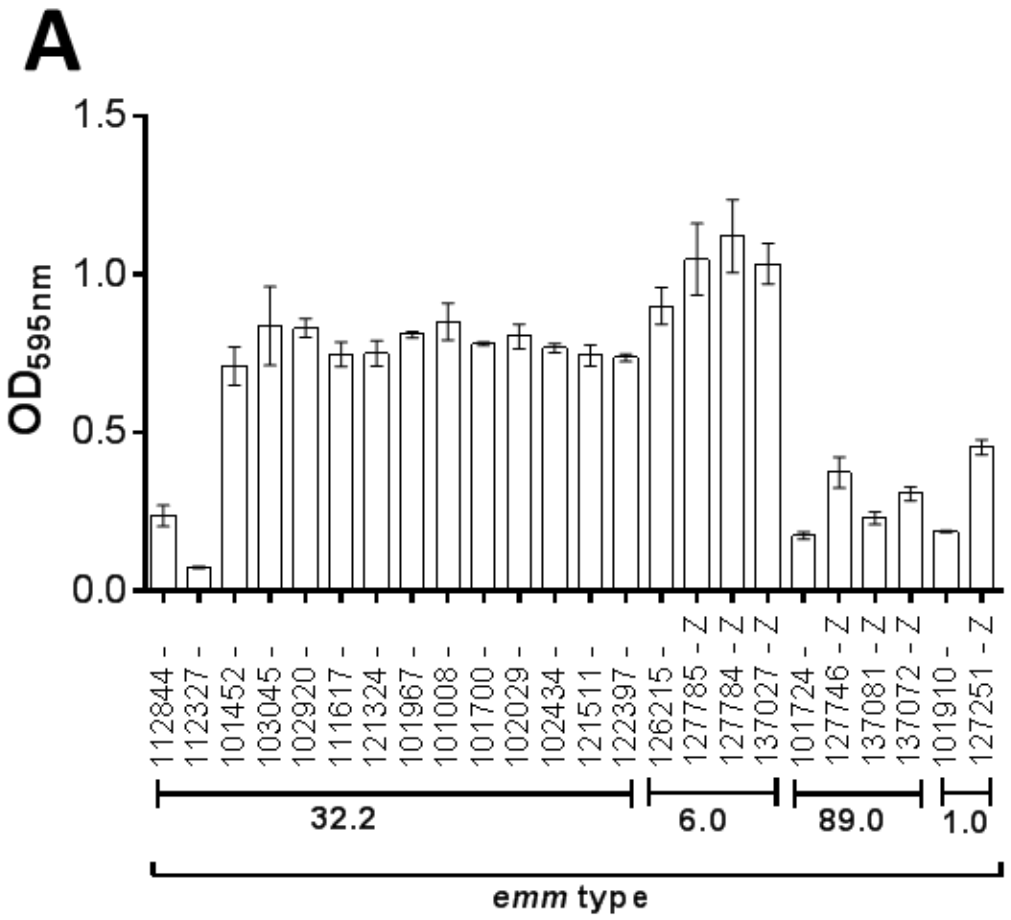
Table 3-5Pair wise comparison of isolates used in biofilm formation at 24h analysed using One-way ANOVA, followed by Tukey's post-hoc test. Significance indicated by adjusted p value. ns=not significant

	112844	112327	101452	103045	102920	111617	121324	101967	101008	101700	102029	102434	121511	122397	126215	127785	127784	137027	101724	127746	137081	137072	101910	127251
112844		ns	ns	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.035	ns	ns	ns	0.08
112327			ns	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.004	ns	ns	ns	ns	ns
101452				ns	0.016	0.035	0.005	0.044	0.027	ns	0.024	ns	0.003	0.013	<.001	<.001	<.001	<.001	ns	ns	ns	ns	ns	ns	ns	ns
103045					0.015	0.031	0.004	0.039	0.024	ns	0.021	ns	0.003	0.011	<.001	<.001	<.001	<.001	ns	ns	ns	ns	ns	ns	ns	ns
102920						ns	ns	ns	ns	ns	ns	ns	ns	ns	0.003	<.001	<.001	ns	<.001	ns	0.015	0.011	ns	ns	ns	ns
111617							ns	ns	ns	ns	ns	ns	ns	ns	0.011	<.001	<.001	<.001	ns	0.031	0.024	<.001	ns	ns	ns	ns
121324								ns	ns	ns	ns	ns	ns	ns	0.009	<.001	<.001	<.001	ns	0.004	0.003	<.001	ns	ns	ns	ns
101967									ns	ns	ns	ns	ns	ns	<.001	<.001	<.001	<.001	ns	0.039	0.031	<.001	ns	ns	ns	ns
101008										ns	ns	ns	ns	ns	0.001	<.001	<.001	<.001	ns	0.024	0.019	<.001	ns	ns	ns	ns
101700											ns	ns	ns	ns	<.001	<.001	<.001	<.001	ns	ns	0.049	<.001	ns	ns	ns	ns
102029												ns	ns	ns	0.002	<.001	<.001	<.001	ns	0.021	0.016	<.001	ns	ns	ns	ns
102434													ns	ns	ns	ns	<.001	<.001	ns	ns	0.044	<.001	ns	ns	ns	ns
121511														ns	0.013	<.001	<.001	<.001	ns	0.003	0.002	<.001	ns	ns	ns	ns
122397															0.003	<.001	<.001	<.001	ns	0.011	0.009	<.001	ns	ns	ns	ns
126215																0.015	ns	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
127785																	ns	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
127784																		ns	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
137027																			<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001
101724																				ns	ns	ns	ns	ns	ns	0.035
127746																					ns	ns	ns	ns	ns	ns
137081																						ns	ns	ns	ns	ns
137072																							ns	ns	ns	ns
101910																								ns	ns	ns
127251																										

3.2.10 Biofilm assay at 48h

At 48h (Figure 3-7A), for most of the *emm32.2* strains, a high level of biofilm formation was measured and variation was observed between strains at 48h ($p < 0.0001$) (Table 3-6). Hence, it was possible to divide *emm*32.2 at 48h into three groups on the basis of OD_{595nm} measurement: less than 0.2 (low: 112327) and more than 0.5 (high: 101452, 103045, 102920, 111617, 121324, 101967, 101008, 101700, 102029, 102434, 121511, and 122397).

No significant differences were determined among the *emm*6.0 strains, however, the *emm*6 type presented the highest degree of biofilm formation at 48h. Furthermore, our results showed the difference in biofilm formation between invasive (101910) and non-invasive (127251) strains of *emm*1.0 ($p=0.003$) after 48h of incubation. Adding to this, there was obvious diversity in biofilm formation amongst the *emm89.0* strains: the invasive (101724) strain exhibited a little amount of biofilm formation compared to non-invasive strains (127746, 137081 and 137072).


A
1.5
1.0
0.5
0.0
OD595nm
112844 -
112327 -
101452 -
103045 -
102920 -
111617 -
121324 -
101967 -
101008 -
101700 -
102029 -
102434 -
121511 -
122397 -
126215 -
127785 - Z
127784 - Z
137027 - Z
101724 -
127746 - Z
137081 - Z
137072 - Z
101910 -
127251 - Z
32.2
6.0
89.0
1.0
emm type

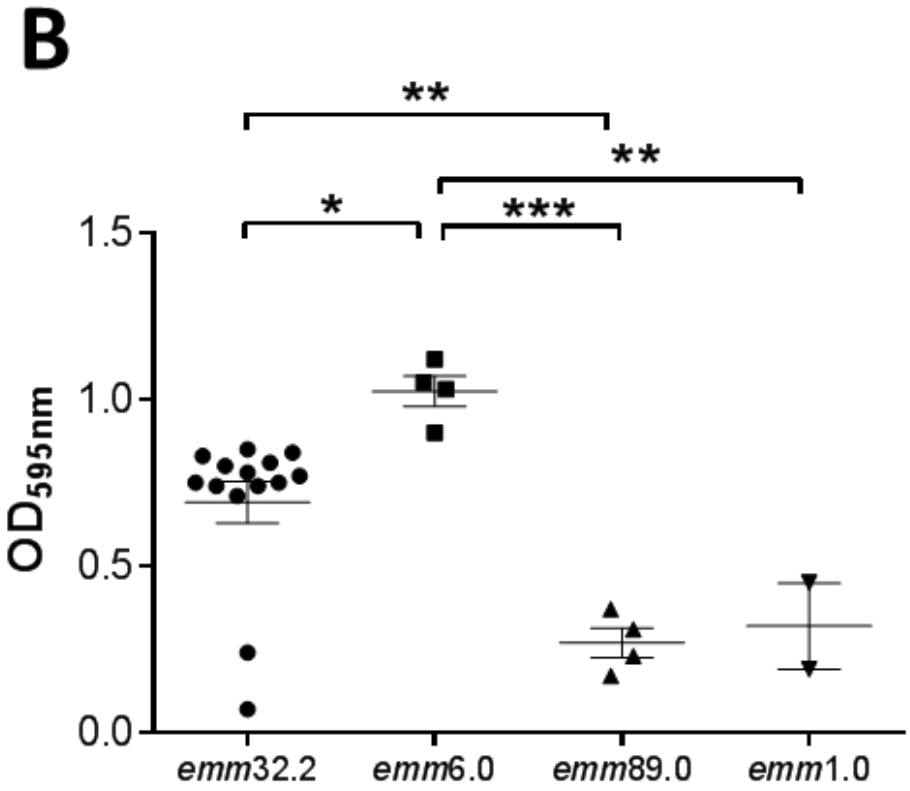


Figure 3-7 Biofilm formation of Group A Streptococcus at 48h

Biofilm formation of *emm*32.2, *emm*6.0, *emm*89.0 and *emm*1.0 was asset at 48h (One-way ANOVA, followed by Tukey's post-hoc test, p<0.0001). (B) Comparison of biofilm formation between *emm*32.2 and *emm*6.0, *emm*89.0 and *emm*1.0 using One-way ANOVA, followed by Tukey's post-hoc test (p=< 0.001). N, non-invasive. * P ≤ 0.05, ** P ≤ 0.01, *** P ≤ 0.001.

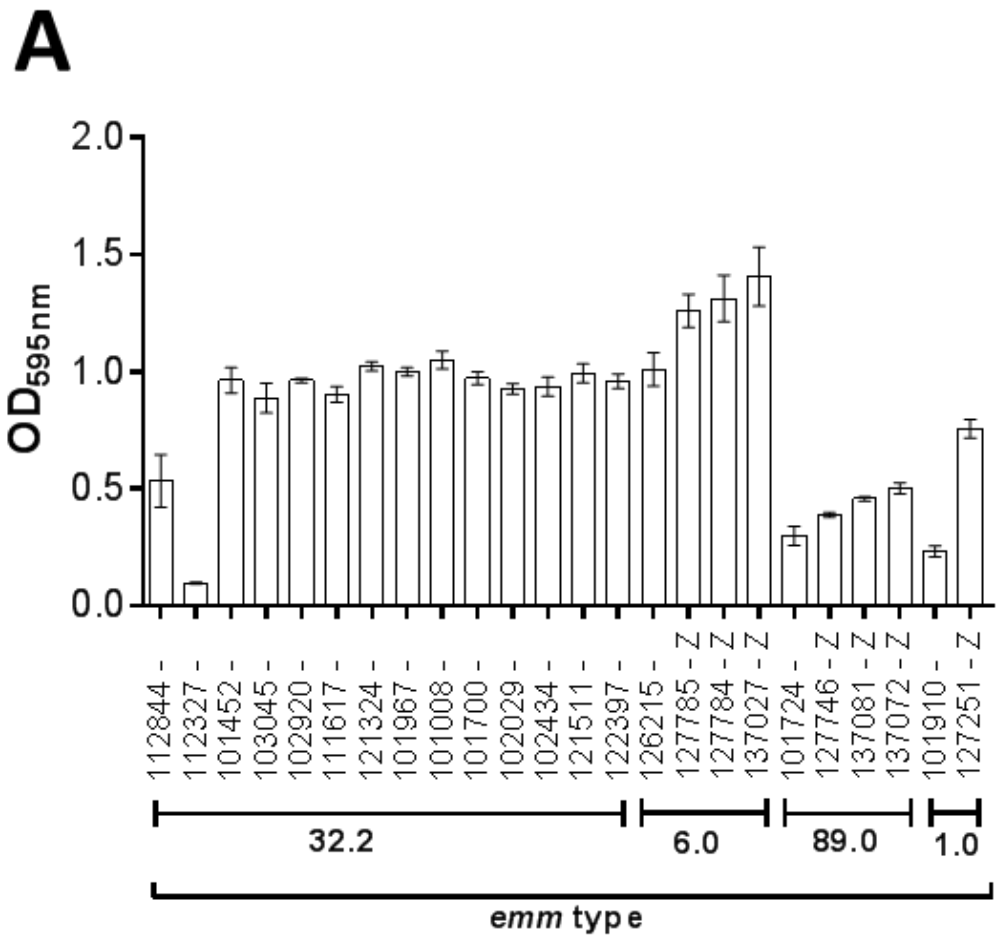
Table 3-6 Pair wise comparison of isolates used in biofilm formation at 48h analysed using One-way ANOVA, followed by Tukey's post-hoc test. Significance indicated by adjusted p value. ns=not significant

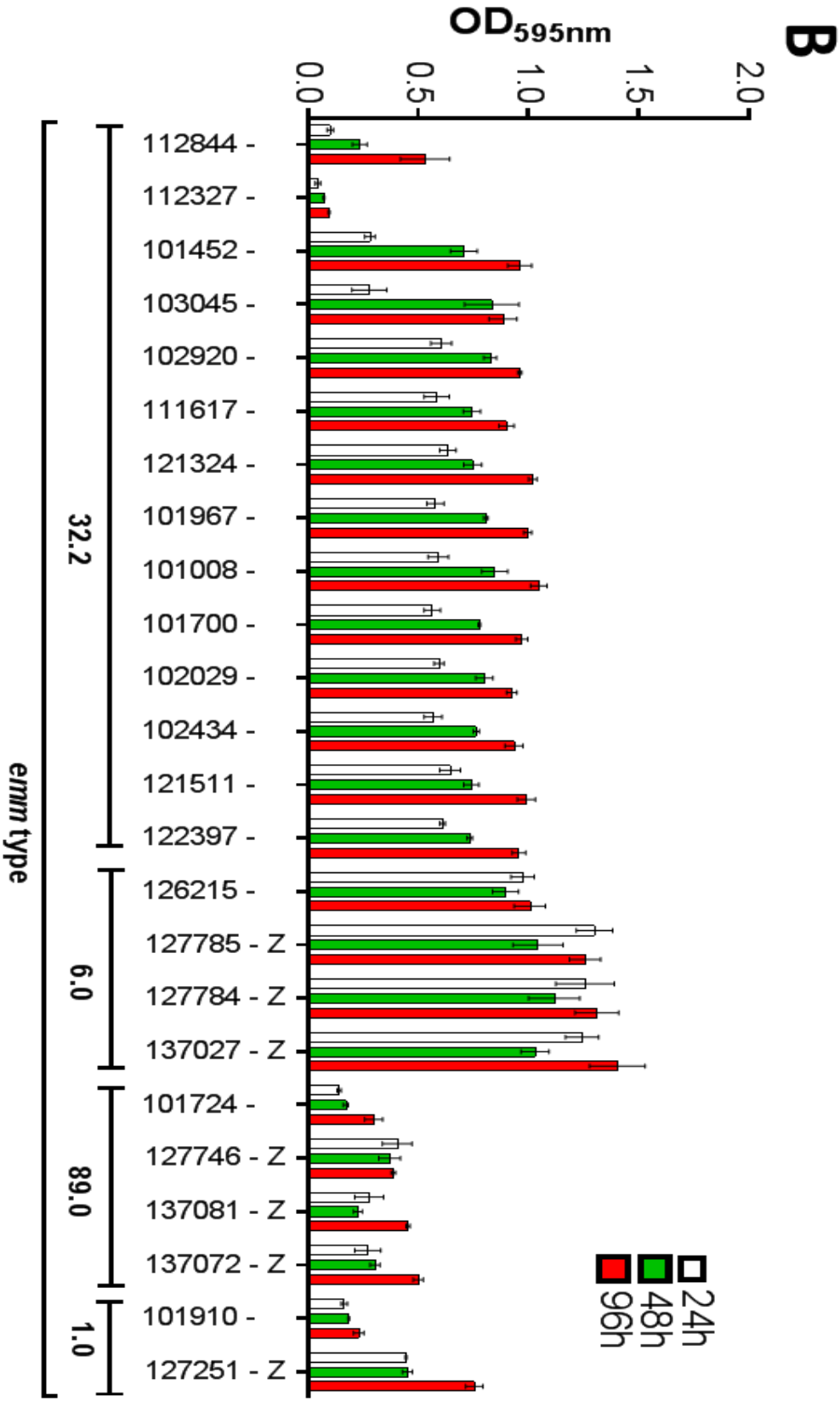
	112844	112327	101452	103045	102920	111617	121324	101967	101008	101700	102029	102434	121511	122397	126215	127785	127784	137027	101724	127746	137081	137072	101910	127251
112844		ns	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	ns	ns	ns	ns	ns
112327			<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	<.001	0.037	ns	ns	ns	0.002
101452				ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.009	<.001	0.015	<.001	0.009	<.001	<.001	ns	ns
103045					ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	<.001	<.001	<.001	<.001	<.001	0.001
102920						ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	<.001	<.001	<.001	<.001	<.001	0.002
111617							ns	ns	ns	ns	ns	ns	ns	ns	ns	0.037	0.002	ns	<.001	0.002	<.001	<.001	<.001	0.047
121324								ns	ns	ns	ns	ns	ns	ns	ns	0.042	0.002	ns	<.001	0.002	<.001	<.001	<.001	0.042
101967									ns	ns	ns	ns	ns	ns	ns	ns	0.026	ns	<.001	<.001	<.001	<.001	<.001	0.004
101008										ns	ns	ns	ns	ns	ns	ns	ns	ns	<.001	<.001	<.001	<.001	<.001	<.001
101700											ns	ns	ns	ns	ns	ns	0.008	ns	<.001	<.001	<.001	<.001	<.001	0.014
102029												ns	ns	ns	ns	ns	0.02	ns	<.001	<.001	<.001	<.001	<.001	0.005
102434													ns	ns	ns	ns	0.005	ns	<.001	<.001	<.001	<.001	<.001	0.023
121511														ns	ns	ns	0.033	0.002	<.001	0.002	<.001	<.001	<.001	ns
122397															ns	0.026	0.001	0.042	<.001	0.003	<.001	<.001	<.001	ns
126215																ns	ns	ns	<.001	<.001	<.001	<.001	<.001	<.001
127785																	ns	ns	<.001	<.001	<.001	<.001	<.001	<.001
127784																		ns	<.001	<.001	<.001	<.001	<.001	<.001
137027																			<.001	<.001	<.001	<.001	<.001	<.001
101724																				ns	ns	ns	ns	ns
127746																					ns	ns	ns	ns
137081																						ns	ns	ns
137072																							ns	ns
101910																								ns
127251																								

3.2.11 Biofilm assay at 96h.

Figure 3-8A is shown the absorbance reading obtained after 96h of incubation with the 24 clinical strains tested in this project (One-way ANOVA, followed by Tukey's post-hoc test $p<0.0001$) (Table 3-7). Amongst the *emm*32.2 strains, statistically significant differences were observed (One-way ANOVA, followed by Tukey's post-hoc test, $p<0.0001$). One strain (112327) presented the lowest amount of biofilm formation in comparison to the other *emm*32.2 strains. No substantial changes were among the strains in the biofilm formation of the *emm6.0* type. Similarly, there were statistically significant differences between *emm*89.0 strains (One-way ANOVA, followed by Tukey's post-hoc test, $p=0.0018$) and *emm*1.0 strains (Student t-test $p=0.0003$).

The biofilm formation of the four different *emm* types (*emm*32.2, *emm*0.6, *emm*89.0, and *emm*1.0) of GAS strains were assessed at different time intervals (24h–96h). In the quantitative assessment, the emm0.6 type produced substantial biofilms at 24h, 48h, and 96h. A gradual increase in biofilms formation was observed over time in the most of emm32.2 Strain. Most of the 4 *emm* types formed substantial amount of biofilms at 96h incubation (Figure 3-8B).


A
2.0
1.5
1.0
0.5
0.0
OD595nm
112844 -
112327 -
101452 -
103045 -
102920 -
111617 -
121324 -
101967 -
101008 -
101700 -
102029 -
102434 -
121511 -
122397 -
126215 -
127785 - Z
127784 - Z
137027 - Z
101724 -
127746 - Z
137081 - Z
137072 - Z
101910 -
127251 - Z
32.2
6.0
89.0
1.0
emm type


B
OD595nm
0.0
0.5
1.0
1.5
2.0
112844 -
112327 -
101452 -
103045 -
102920 -
111617 -
121324 -
101967 -
101008 -
101700 -
102029 -
102434 -
121511 -
122397 -
126215 -
127785 - Z
127784 - Z
137027 - Z
101724 -
127746 - Z
137081 - Z
137072 - Z
101910 -
127251 - Z
emm type
32.2
6.0
89.0
1.0
24h
48h
96h

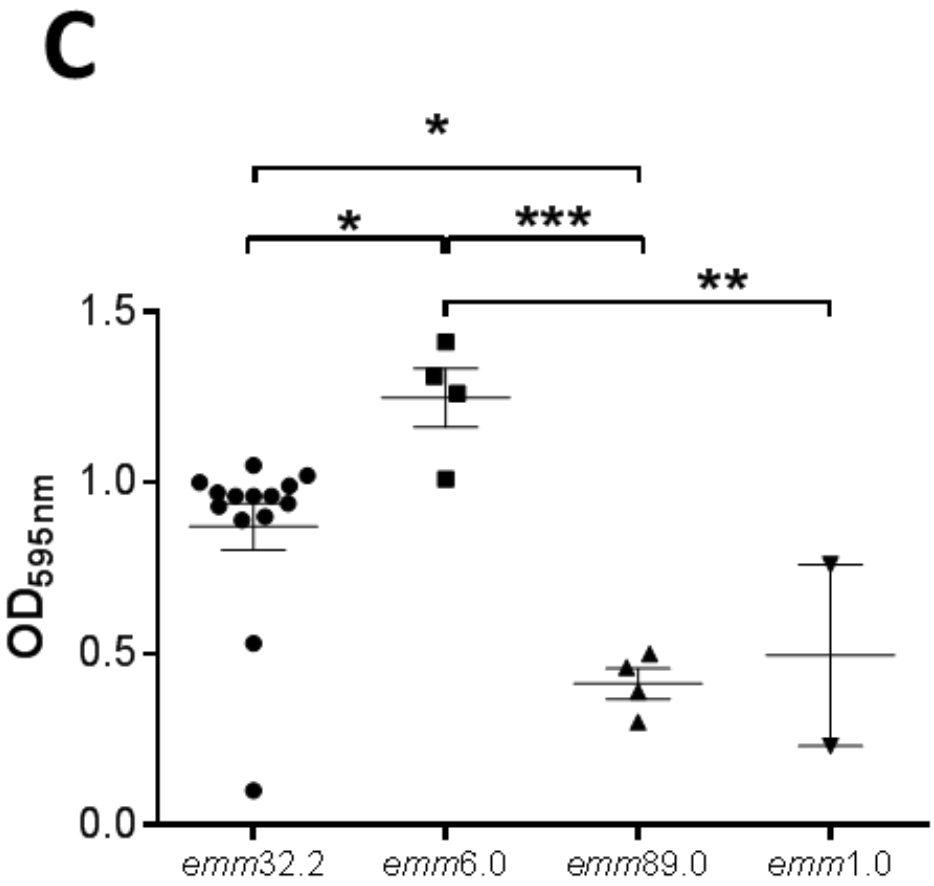


Figure 3-8 Biofilm formation of Group A Streptococcus at 96h

(A) Biofilm formation of *emm*32.2, *emm*6.0, *emm*89.0 and *emm*1.0 was determined at points 96h (One-way ANOVA, followed by Tukey's post-hoc test, $p<0.0001$).(B) Comparison of biofilm formation between *emm*32.2 and *emm*6.0, *emm*89.0 and *emm*1.0 using One-way ANOVA, followed by Tukey's post-hoc test at 96h ($p= < 0.01$). (C) Apart from 4 *emm*6 strains and one *emm*89.0 strains (127746), a gradual increase in biofilm formation was observed for all strains over time. N, non-invasive. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$.

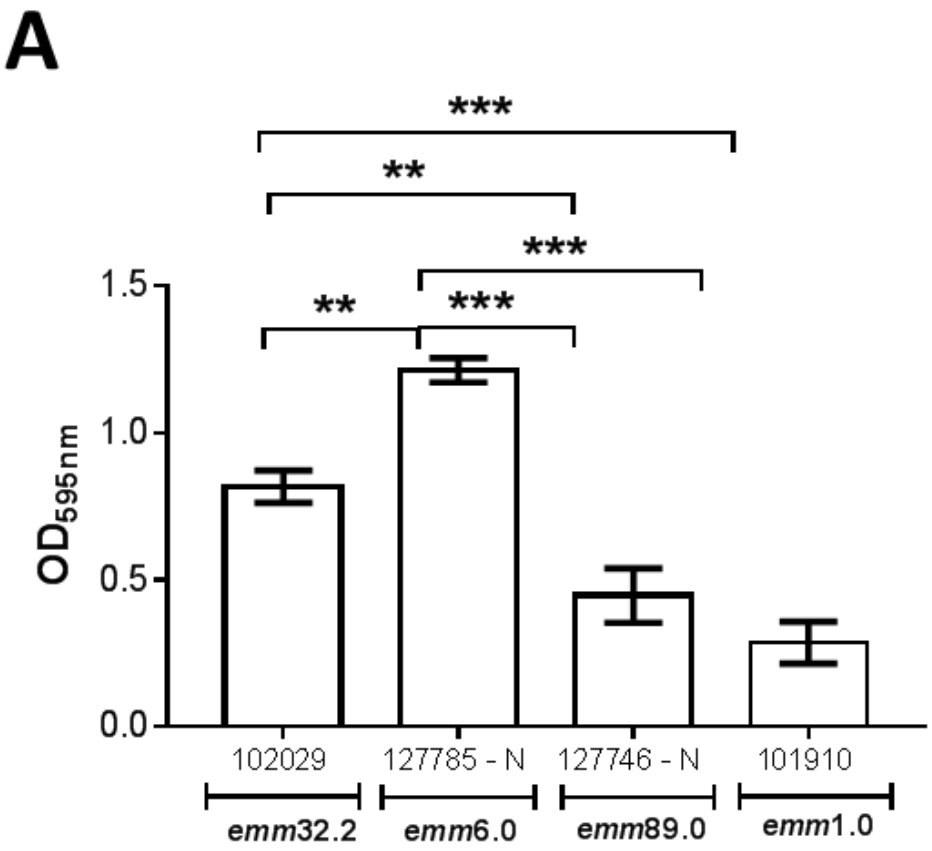
Table 3-7 Pair wise comparison of isolates used in biofilm formation at 96h analysed using One-way ANOVA, followed by Tukey's post-hoc test. Significance indicated by adjusted p value. ns=not significant

	112844	112327	101452	103045	102920	111617	121324	101967	101008	101700	102029	102434	121511	122397	126215	127785	127784	137027	101724	127746	137081	137072	101910	127251
112844		ns	0.033	ns	<.001	0.014	0.003	<.001	0.014	0.012	0.01	0.015	0.016	0.013	ns	0.029	ns	ns	ns	0.014	0.008	0.017	ns	0.029
112327	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
101452	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
103045	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	0.04	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns
102920	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.01	0.011	<.001	0.012	0.001	ns	
111617	ns	ns	ns	ns	ns		ns	0.002	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.017	ns	0.026		0.011	ns	
121324	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.041	0.003	0.013	0.049	0.016	ns	
101967	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	0.034	0.012	0.014	0.013	0.013	ns	
101008	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	0.007	0.038	0.017	0.04	0.009	ns	
101700	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	0.01	0.033	ns	0.036	ns	
102029	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	0.005	0.033	ns	0.036	ns	
102434	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	0.019	ns	ns	0.014	0.013	ns	
121511	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	0.02	ns	ns	ns	ns	
122397	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	0.018	0.024	ns	0.028	ns	
126215	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	0.026	ns	
127785	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	0.021	ns	0.045	ns	ns	ns	
127784	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	
137027	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.021	ns		ns	ns	ns	ns	ns	
101724	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	ns	ns	
127746	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		ns	ns	0.05	
137081	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		0.048	ns	
137072	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.048		ns	
101910	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns		
127251	ns	ns	ns	ns		ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	ns	0.05	ns	ns		

Biofilm assay at 168h / 7d

Following 168h / 7d of incubation, significant differences between the four *emm* type strains are observed, with some strains forming a low biofilm and others forming high biofilms (Figure 3-9A) (One-way ANOVA, followed by Tukey's post-hoc test, $p< 0.0001$). Most strains produced approximately the same amount of biofilm formation at 168h/ 7d as than that observed at 96h (Figure 3-9B).

In conclusion, the data show that the *emm* types (namely, *emm*32.2, *emm*6.0, *emm*89.0, and *emm*1.0) generated observable biofilms on uncoated plastic surfaces *in vitro*, and heterogeneity was identified within strain and from one strain to another. Therefore, it is reasonable to conclude that biofilm is a trait of the individual GAS strain, rather than of the defined GAS *emm* types. In addition, we observed variations in biofilm formation among the *emm*32.2 strains at various time points, thus, we could not establish any correlation between biofilm formation and invasive properties of *emm*32.2.


A
***
**
***
**
***
1.5
1.0
0.5
0.0
OD595nm
102029
127785 - N
127746 - N
101910
emm32.2
emm6.0
emm89.0
emm1.0

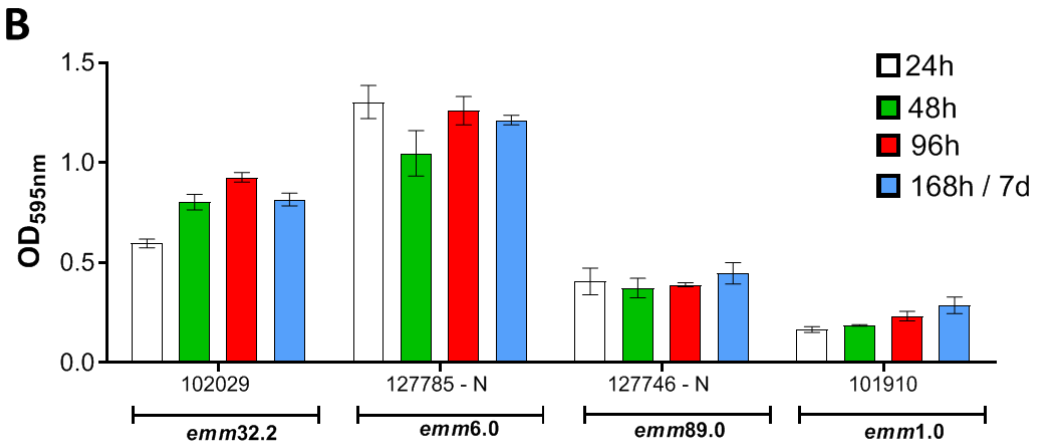


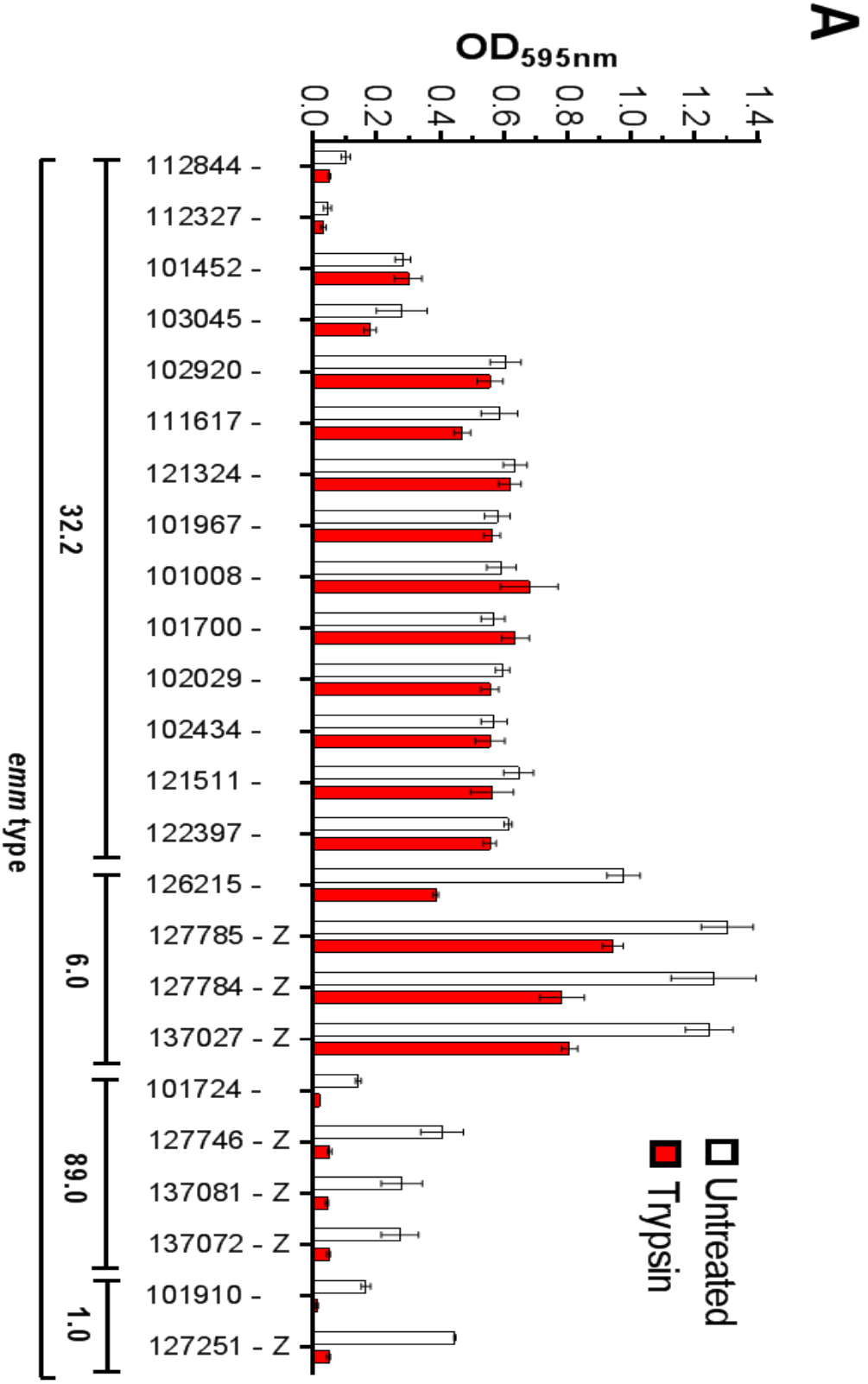
Figure 3-9 Biofilm formation of Group A Streptococcus at 168h / 7d

One representative strain from each respective *emm* type was grown at 37°C in THBY medium supplemented with glucose for 168h / 7d. (One-way ANOVA, followed by Tukey's post-hoc test, $p<0.0001$). Data are represented as mean (= sample OD – blank OD) +/- standard error of the mean (SEM). Results are derived from three independent experiments, each of which included eight replicates. (B) No significant differences were found between the amount of biofilm formed at 168h/ 7d compared to 96h. N, Non-invasive.* $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$.

3.2.12 GAS biofilm sensitivity to trypsin digestion

Trypsin is a pancreatic serine endoprotease that cleaves proteins or peptides on the carboxyl side of arginine (R) or lysine (K) residues [166]. To investigate the effects of trypsin enzymes on the degradation of biofilms of *emm*32.2 and *emm*1.0, *emm*89.0, and *emm*6.0 strains, trypsin was applied for 2h on 22h old biofilms.

Treatment of pre-formed biofilms with trypsin had a dispersal effect in all tested strains of *emm*0.6, *emm*89.0, and *emm*1.0 in comparison with non-treated controls (Figure 3-10A). Incubation with trypsin of the mature biofilms from strains *emm*6.0, *emm*89.0, and *emm*1.0 resulted in detachment in nearly all strains tested from 3 *emm* types. Most interestingly, however, *emm*32.2 strains were the most resistance to trypsin when compared to *emm*1.0, *emm*89.0, and *emm*6.0 strains (Figure 3-10B). Overall, as *emm*32.2 showed resistance to trypsin digestion, this feature could be one key contributing factor to *emm*32.2 GAS pathogenesis.


A
OD595nm
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
112844 -
112327 -
101452 -
103045 -
102920 -
111617 -
121324 -
101967 -
101008 -
101700 -
102029 -
102434 -
121511 -
122397 -
126215 -
127785 - Z
127784 - Z
137027 - Z
101724 -
127746 - Z
137081 - Z
137072 - Z
101910 -
127251 - Z
32.2
6.0
89.0
1.0
emm type
Untreated
Trypsin

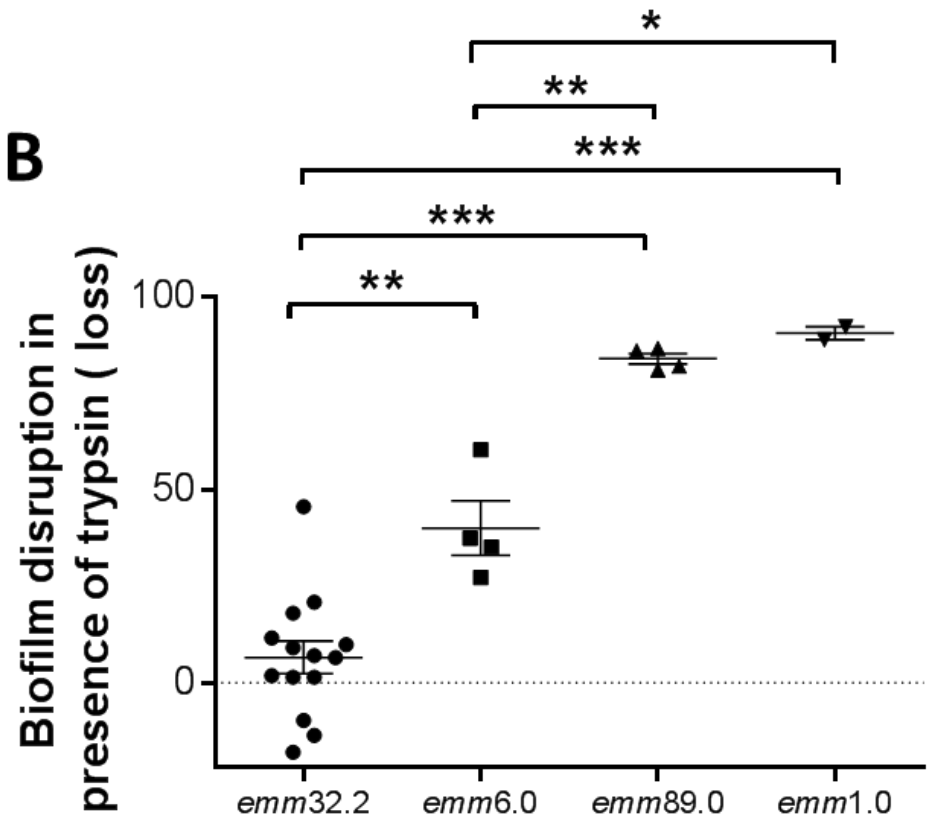


Figure 3-10 Effect of trypsin on pre-formed Group A Streptococcus biofilms

(A) Biofilm formation (OD595nm) was assessed in the absence/presence of trypsin for the final 2h of a 24h incubation period in 96 well microtitre plates. Values are presented as mean ± S.E.M. Data are represented as mean (= sample OD – blank OD) +/- standard error of the mean (SEM). Results are derived from three independent experiments, each of which included eight replicates. (B) Percent disruption of *emm*32.2, *emm*6.0, *emm*89.0 and, *emm*1.0 biofilms strains in presence of trypsin. (One-way ANOVA, followed by Tukey's post-hoc test, $p<0.0001$). Data are represented as mean +/- standard error of the mean (SEM). Results are derived from three independent experiments, each of which included eight replicates. N, non-invasive. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$.

3.3 Discussion

Table 3-8 Summary of phenotypic characteristics of Group A Streptococcus isolates

Strain	*emm* Type	Capsule Thickness[1]	Hyaluronic Acid [2]	Complement Deposition[3]	SLO Activity[4]	Biofilm at 24h[5]	NADase Activity[6]	Proteolytic Activity[7]	Trypsin Sensitivity[8]
112844	32.2i	H	H	L	H	L	N	Y	Y
112327	32.2i	H	H	L	H	L	N	N	N
101452	32.2i	M	M	L	H	M	N	Y	N
103045	32.2i	M	M	L	H	M	N	Y	N
102920	32.2i	L	L	L	M	H	N	Y	N
111617	32.2i	L	L	M	L	H	N	N	N
121324	32.2i	L	L	M	L	H	N	Y	N
101967	32.2i	L	L	L	L	H	N	N	N
101008	32.2i	L	L	M	L	H	N	N	N
101700	32.2i	L	L	M	L	H	N	N	N
102029	32.2i	L	L	H	L	H	N	N	N
102434	32.2i	L	L	M	L	H	N	N	N
121511	32.2i	L	L	M	L	H	N	Y	N
122397	32.2i	L	L	M	L	H	N	N	N
126215	6.0i	L	M	H	M	H	Y	Y	Y
127785	6.0n	L	L	H	L	H	Y	Y	Y
127784	6.0n	L	L	H	L	H	Y	Y	Y
137027	6.0n	L	L	M	M	H	Y	Y	Y
101724	89.0i	L	L	H	L	L	N	Y	Y
127746	89.0n	L	L	H	H	M	Y	Y	Y
137081	89.0n	L	L	H	H	M	Y	Y	Y
137072	89.0n	L	L	H	M	M	Y	Y	Y
101910	1.0i	L	L	H	L	L	Y	Y	Y
127251	1.0n	L	L	H	M	M	Y	Y	Y

i= invasive, n=non-invasive, L=Low, M=medium, H= High. Y=Yes, N=No

1. Capsule Thickness (pixels): (Low < 200, medium 200-400, high > 400)
2. Hyaluronic Acid (fg per 103 CFU): (low > 300, medium 300-600, high<600)
3. Complement Deposition (flincr): (low <4000, medium 4000 – 6000, high> 6000)
4. SLO activity (%): (Low <5, meduim5-10, high >10)
5. Biofilm information (OD: 595nm): (low<0.0.5, medium 0.05-.2, high 0.2)
6. NADase activity (fold change): (1/3 = Yes, 0/ 3 No)
7. Proteolytic Activity: (Clear zone=YES, No clear zone=No)
8. Sensitivity to Trypsin (%): (20 or more = Yes and 20 or less = No).
9. The shaded rows indicate the representative isolates which were used for transcriptomic analysis and proteomic analysis.

The aim of this project was to examine in detail the biological properties of a small group of group A streptococcal (GAS) isolates collected during a contained outbreak which occurred within the Merseyside region between January 2010 and September 2012. These isolates were determined to belong to a novel *emm* type (*emm*32.2) of GAS, which was compared against predominantly circulating GAS *emm* types using a range of in vitro assays. These circulating strains were used as comparators, in lieu of reference strains e.g. ATCC/NCTC sources.

Several *in vitro* experiments were conducted to compare phenotypic characteristics of *emm*32.2 and other GAS *emm* types. The experiments involved comparisons of important factors including capsule thickness, resistance to complement deposition, proteinase activity, SLO and NADase activity, and biofilm formation. The experimental results from the *in vitro* assays showed significantly different outcomes regarding the phenotypic attributes of *emm*32.2 and *emm*1.0, *emm*89.0, and *emm*6.0 strains (Table 3-8).

In the present study, our experimental findings showed a statistically significant difference between the *emm*32.2 and *emm*1.0, *emm*89.0, and *emm*6.0 isolates in relation to capsule thickness. For the *emm*6 strain, no significant difference was observed between invasive and non-invasive isolates. In addition, *emm*89.0 displayed significant differences between invasive and non-invasive isolates, and this was also the case for the *emm*1.0 strains. These findings are in line with literature, describing that, capsule variabilities are observed in M89 isolates [34]. The findings of the present study demonstrated the presence of capsule thickness variabilities not only between

emm strains, but also within the same *emm* type. Previous studies found that variance also exists in the synthesis of the hyaluronic acid capsule among GAS isolates of the same strain [167, 168]. One of the ways in which to account for variation between strains capsule production is polymorphisms in the promoter region upstream of the has operon. In the study conducted by Crater et al. (1995), showed that variabilities in capsule production from one strain to another are supported by transcriptional procedures [167]. One interesting observation from our data is that the variation of capsule thickness in the outbreak *emm*32.2 was unexpected since they share the same unique *emm* type. The ability of low and high capsule of *emm*32.2 strains (positive for the capsule locus *hasABC*) to cause invasive disease, along with the recognition that *emm*4 and *emm*22 strains are also acapsular [169], suggests that encapsulation is not as essential for pathogenesis by all outbreak strains (*emm*32.2) as previously believed.

Hyaluronic acid (HA) production was measured using an enzyme-linked immunosorbent assay (ELISA)-based assay specific for HA. Although there was variability in levels of HA among GAS strains, *emm*32.2 strains produced clearly detectable levels of HA. No HA was detected in strains that were members of the *emm*89.0 (127746, 137081, and 137072) on further investigation these isolates were negative for the HA capsule locus *hasABC* [146].

This study compared a validated ELISA-based method of assessing capsule thickness with a fluorescence microscopy-based method. The ELISA-based approach was confirmatory for the existing observation of that the *emm*32.2 outbreak cluster was phenotypically heterogeneous. Therefore, both approaches were a suitable way to

determine capsule thickness. Moreover, when comparing ELISA-based against a fluorescence microscopy-based method, 3 strains of *emm*89.0 were not closely aligned with the linear trend curve, which may be attributable to their lack of the *hasABC* operon encoding capsule biosynthesis enzyme [146]. Although capsular status is a fundamental virulence factor, it is not possible to distinguish *emm*32.2 from *emm*1.0, *emm*89.0, and *emm*6.0 isolates based on capsule thickness alone.

A previous study demonstrated a positive correlation between capsule thickness and protection against phagocytosis [73]. My project measured and compared C3b complement deposition levels on the surface of GAS isolates, and the findings indicated that *emm*32.2 had the greatest resistance to complement deposition. Nevertheless, important variations were observed regarding complement deposition sensitivity, both within and across *emm* strains. Therefore, it was reasonable to conclude that the differences in resistance to complement deposition was dependent on the strain and not on the serotype. The capsule thickness assay results demonstrate significant differences in capsule thickness between tested GAS strains. Capsule thickness has previously been reported to correlate with resistance to complement deposition on the bacterial surface [170, 171]. Our data indicate a clear correlation between increased capsule thickness and decreased complement deposition when comparing *emm*32.2 isolates *emm*32.2 and *emm*1.0, *emm*89.0, and *emm*6.0. Heterogeneity in the capsule could partly explain the differences in complement deposition observed between GAS isolates. However, the role other inhibitors of complement deposition such as complement-inhibitory proteins, including C4b

binding protein (C4BP) factor H and factor H-like protein 1 (FHL-1) or M protein need to be tested.

I next sought to examine the secreted cysteine proteinase (sepB) activity of my collection of GAS isolates. This enzymatic activity is documented for its contribution to GAS virulence, e.g., degradation of host extracellular matrix and impairment of immune responses [11, 95, 172]. The incidence of proteinase-positive GAS strains found in previous studies is as follows: 81% of 21 strains, [173] 78% of 47 strains, [174] strains [155]. These figures are largely consistent with my observation that 67% proteinase-positive strains out of the 24 strains (all tested isolates were *SepB*-positive [146]). Our results indicated that 6/14 *emm*32.2 isolates were proteinase producers, which is consistent with the literature reporting the absence of any correlation between *emm* GAS serotype and proteinase production [155, 175]. Our data demonstrated the presence of SpeB variabilities not only between *emm* strains, but also within the same *emm* type. These findings are in line with literature, describing that, significant strain-to-strain differences and among strains of the same M-protein serotype occur in *SpeB* expression levels [176, 177]. Previous study exhibited that variation in *SpeB* transcript level is not correlated with M-type [178]. Carroll and Musser [179] propose that the control of SpeB is highly complex, and is controlled by a number of environmental factors including pH and salt concentration, in addition to regulators of transcription. For instance, previous investigations on the regulation of transcription identified a 12 regulatory components including the protease B regulator (RopB) [180]. Although this study displayed a high incidence of proteinase activity amongst a

collection of 24 GAS strains, 8 out of the 14 *emm*32.2 isolates were determined to be protease negative. On the basis of this observation, I concluded that cysteine protease may not be the only factor responsible for the highly invasive properties of *emm*32.2 type and that other contributing factors remained to be identified.

Based on the NADase activity data collected in the present study, differences were observed across the 24 tested GAS isolates. Remarkably, while most of the *emm*1.0, *emm*89.0, and *emm*6.0 isolates had the capability of hydrolysing NAD^+, NADase activity was not observed for *emm*32.2 strains (all tested isolates are positive nga gene [146]). Consistent with previous studies [34], the experimental results indicated heterogenous NADase activity within *emm*89.0 strains. More importantly, our data indicated that the lack of NADase activity was shared by the invasive *emm*32.2 strain. This finding suggests *emm*32.2 is an inactive NADase subtype.

Data relating to SLO activity indicated variabilities in SLO activity for the examined GAS isolates (all tested isolates are positive SLO gene [146]). Noteworthy differences were observed across the 24 isolates, and it was also the case that significant differences were observed within the *emm*32.2 isolates. Both *emm*6.0 isolates and *emm*89.0 isolates did not display significant differences, while significant variations were found between the invasive and non-invasive *emm*1.0 isolates. My results using capsule thickness assay showed a statistically significant difference across GAS strains. Hence, it was reasonable to test a correlation between SLO haemolytic activity and capsule thickness. The findings showed a positive correlation between these 2 variables, thereby indicating that capsule thickness may play an important role in SLO

activity. My results failed to reveal an association between SLO haemolytic activity and NADase activity, which is consistent with data from the literature, indicating that SLO has no statistically significant impact on SPN production or NADase activity *in vitro* [110, 181]. Our data showed a high level of variation in SLO activity within the *emm*32.2 isolates. Hence, we were not able to identify via measurements of SLO haemolytic activity a relationship between SLO and *emm*32.2 strains. Further research is required to gain further insight into this relationship.

Current research has yet to describe conclusive findings on the relationship between human disease and GAS biofilms. Therefore, the next objective was to characterise the types of biofilms produced by GAS isolates. My study examined the biofilm formation of GAS isolates (including *emm*32.2, *emm*0.6, *emm*89.0, and *emm1.0*) at 24h, 48h, 96h and 168h/7d. The crystal violet staining technique was used to assess all the strains with respect to their capability to produce biofilms. In the quantitative analysis, our data showed significant variation among different *emm* types in their ability to form biofilms. However, *emm*6.0 strain generated high amount biofilms at 4 time points. Within *emm*32.2 strain, biofilm formation increased over time for most isolates. I selected representative isolates from each *emm* type and measured biofilm formation at 168h/7d, and it was noted that in every case, the amount of biofilm produced was approximately the same as that produced at 96h. This study revealed that the clinically important *emm* type (including *emm*32.2, *emm*6.0, *emm*89.0. and *emm*1.0) GAS strains produced a detectable amount of biofilm on uncoated plastic surfaces *in vitro* and revealed heterogeneity amongst the strains. These results infer that biofilm formation

is a trait of the individual GAS isolate, rather than serotypes. Our results were consistent with the literature. In particular, previous studies found that biofilm formation varied among different serotypes and even among strains of the same *emm* type, as well as also showing *emm*6.0 as a strong biofilm producer [67, 70, 156, 182]. Numerous transcriptional regulators were revealed to be involved in and key for the establishment and maintenance of biofilms [64, 67, 72].

The role of the hyaluronic acid capsule in GAS biofilms is not obvious yet [72]. There are inconsistent observations described by different groups [72]. Our data demonstrated that capsule reduction corresponded to the increased amount of biofilm, showing an inverse relationship between capsule expression and biofilm formation. These results suggest that capsule thickness could play an important role in the reduction of the amount of biofilm. Indirect results indicate that capsule production performs an inhibitory function for biofilm formation, given that the literature shows that covS deletion in numerous GAS strains results in higher capsule production, accompanied by reduced biofilm biomasses [183]. Comparably, capsule reduction has been found to correlate with higher biofilm formation among certain *streptococcus agalactiae* strains [184]. Taken together, the findings indicated that the examined strains showed the capability for biofilm formation. At the same time, variability within a single strain such as *emm*32.2 type were noted. Therefore, the role of biofilm formation in aid *emm*32.2 outbreak is not clear to help us to drawn precise conclusion. Importantly, it should be recognised that the present study is the first publication addressing the characterisation of biofilm formation by GAS *emm*32.2.

To investigate the impact that trypsin has on established biofilms, experiments were conducted to test the capability of Trypsin to disperse pre-formed mature biofilms. When examining the *emm*1.0, *emm*89.0, and *emm*6.0 strains, nearly all performed biofilms at the 24h time point were broken down. Nevertheless, regarding the protein contents of the *emm*32.2 biofilm matrix, this was intact following exposure to the trypsin enzyme which suggests that this could be due to differences in biofilm components or the presence of a protective gene. These findings are in line with findings from the genomic data, outlining that *emm*32.2 contains a unique putative trypsin-resistance surface protein T6 gene, which has previously been associated with virulence [146].

In summary, results have demonstrated that the *emm*32.2 strains is characterised by phenotypic heterogeneity and has variable characteristic regarding capsule thicknesses, proteolytic activity, SLO Activity, and biofilm formation. Furthermore, *emm*32.2 strains all present a lack of NADase activity. More important *emm*32.2 strain is less susceptible to complement deposition which leads to reduced phagocytosis by the host. As a result, this *emm* type would be more resistant to host immune defence, which may thus explain its virulence properties. Finally, it is also worth acknowledging that *emm*32.2 contains a putative trypsin-resistant surface protein T6, could be one of the factors contributing to induction of the *emm*32.2 outbreak. Ultimately, further investigations are necessary to confirm these suppositions.

Chapter 4 Comparative Transcriptomic Analysis of Group A Streptococcus

4.1 Introduction

The investigation of gene expression in response to certain stressors *in vitro*, has been useful to gain insight into the nature of the genes involved in the intracellular life of bacteria [157]. Transcriptomic studies can also help to identify novel genes and give us vital information about gene expression, which can be used to compare bacterial species. Measuring the expression of genes in different tissues, conditions, or time points provides information on how genes are regulated [185]. RNA sequencing (RNA-seq) technology has proven to be a highly effective approach for studying the bacterial transcription profile, but the cost of using this method is expensive. Therefore, few comprehensive expression profiling experiments of GAS and their virulence factors have been performed. Previously published work using RNA-seq and GAS, was achieved in *emm*1 strain MGAS2221 to identify genes where the expression was directly or indirectly regulated by Fatty Acid Synthesis (*fabT*). This study highlighted a number of genes as being differentially expressed including genes implicated in fatty acid synthesis and lipid metabolism which were significantly up-regulated in the *fabT* deletion mutant strain [157]. In another study, levels of RNA

expression of GAS biofilms were compared at multiple stages of growth in GAS strain 5448 (M1T1). It was shown that there were distinct expression profiles in the transcriptome between different time points of cells growing in suspension (planktonic) and those growing forming communities in a self-produced polymeric matrix (biofilm) [186].

The aim of this study was to compare the differentially expressed genes of *emm*32.2 (invasive outbreak isolates), relative to *emm*1.0 and *emm*89.0 isolates, under controlled growth conditions, to describe differences in physiological behaviour that can map to the properties defined in Chapter 3. In this project, RNA-seq was used in order to determine differences in gene expression levels of *emm*32.2 GAS strain (112327) compared to *emm*1.0 (101910 invasive) and *emm*89.0 (127746 non-invasive) during the early, mid, and late exponential phases of the growth curve.

4.2 Results

4.2.1 *In vitro* growth phases

Due to financial constraints, we proceeded to use 2 representative isolates of GAS serotypes found most to be the most prevalent globally, and in the UK compared against one *emm* 32.2 outbreak isolate. We selected *emm*32.2 isolate (112327) as a representative of the outbreak strain based on phenotypic data showing distinctive virulence features, i.e., high capsule and high SLO activity. The isolate (101910) was chosen as a representative of *emm*1 type which is one of the most studied lineages, and shown to be prevalent in several epidemic areas [28]. Isolate (127746) was chosen as a representative of *emm*89 type, which is among the top five leading *emm* types in the world [187] and a leading cause of disease in the United Kingdom and other industrialised countries, such as Canada [34, 188].

Bacterial growth rate may be indicative of invasive potential in the host [189], for this reason, the growth rates of *emm*32.2 (112327), *emm*1.0 (101910) and *emm*89.0 (127746) strains were compared *in vitro* Figure 4-1 and Figure 4-3.

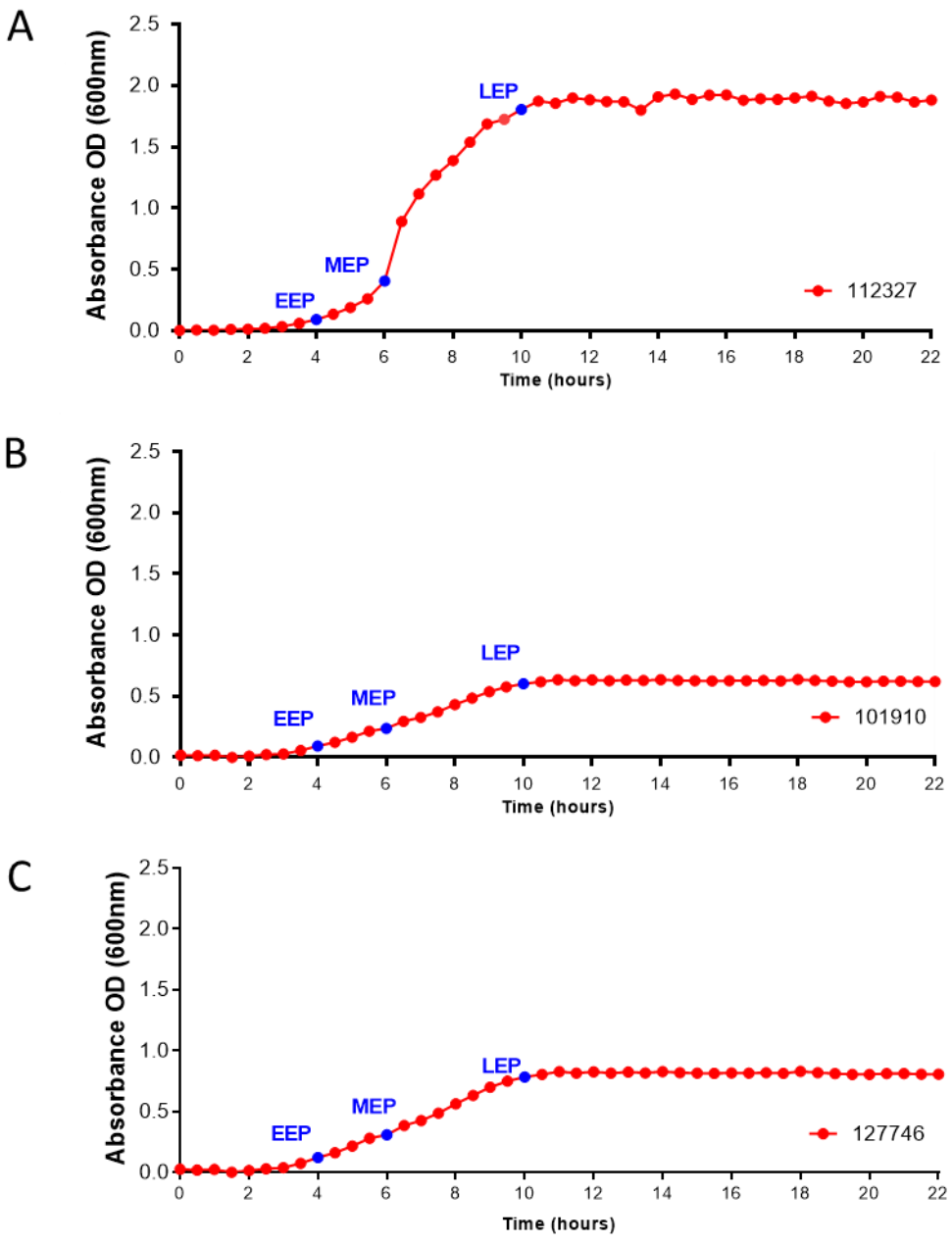


Figure 4-1 Absorbance vs. time growth profiles of the emm32.2 (A), emm1.0 (B) and emm89.0 (C).

For each strain, 10^5 of CFUs were inoculated in 200µl of THY media absorbance readings were determined at 600nm every 30 minutes for a total of 24 hours using an automated microplate reader (Fluorostar Omega, BMG Labtech) set up at a constant temperature of 37°C.

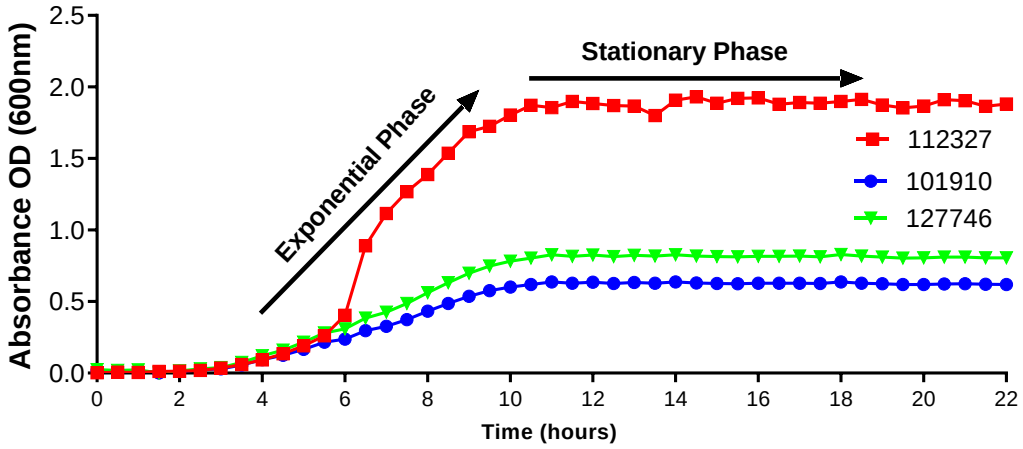


Figure 4-2 Absorbance vs. time growth profiles of the emm32.2 and emm1.0 and emm89.0.

For each strain, 10^5 of CFUs were inoculated in 200µl of THY media absorbance readings were determined at 600nm every 30 minutes for a total of 24 hours using an automated microplate reader (Fluorostar Omega, BMG Labtech) set up at a constant temperature of 37°C.

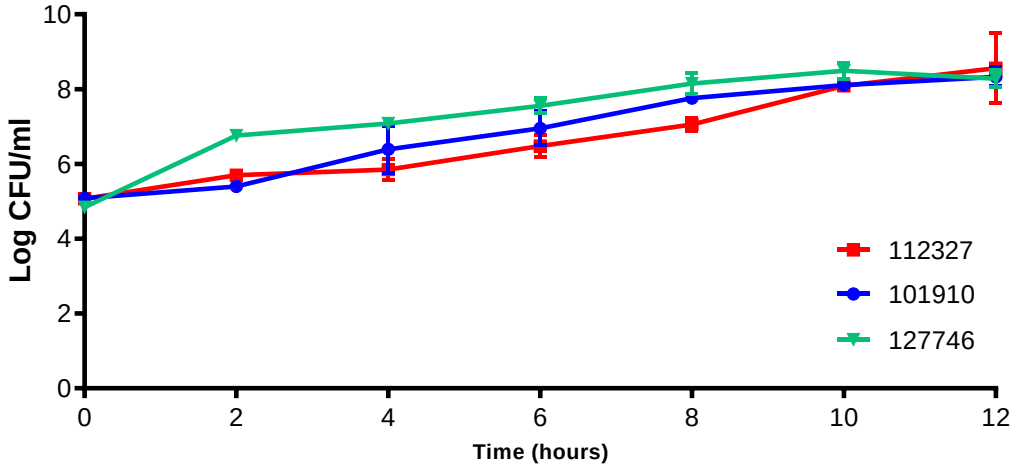


Figure 4-3 Growth curve of Group A Streptococcus strains.

The *emm*32.2 strain (112327) and *emm*1.0 and *emm*89.0 strains (101910 and 127746) were grown at 37C° in THY medium. Data are shown as Log mean CFU/ml (mean ± SEM) of bacterial cultures determined every 2 hours between 0 and 12 hours. The time points 4h, 6h, and 10h, are representative of the early, mid and late exponential phase, respectively, shown in Figure 4-1. The data are representative of 2 independent experiments. Statistical analysis was performed by two-way ANOVA.

Bacterial growth kinetics were investigated in order to identify the logarithmic (log) growth phase of each of the three strains of interest, i.e., *emm*32.2, *emm*1.0, and *emm*89.0 (Figure 4-1). Absorbance readings were recorded to plot the growth curve respective of each strain. Results showed that the exponential phase spans the period comprised between 4 and 10 hours; this 6-hour window covers the entire exponential phase of all three strains. Results shown in Figure 4-2 suggest that the growth rate of *emm*32.2 strains may substantially differ from the other 2 isolates. However, this may be simply due to the mucoid phenotype of the *emm*32.2 strain. To determine whether or not such difference may be reflected in the number of viable bacterial cells, viability counts were performed over time for all three strains (Figure 4-3). The CFU counts showed that the three strains did not significantly differ over their respective exponential growth phases. Hence under our experimental conditions, we state with confidence that the time points 4, 6 and 10 hours are representative of the early, mid and late exponential growth phase for all three strains.

4.2.2 Gene expression analysis of GAS

4.2.3 RNA-seq analysis

The expression profiles of virulence genes have been shown to be growth- phase specific in studies in GAS [190]. I examined the growth-phase specific gene expression of *emm*32.2 relatively to *emm*1.0 and *emm*89.0 isolates. I concentrated on the three phases of bacterial growth: early exponential phase (EEP), mid-exponential phase (MEP), and late exponential phase (LEP). RNA-seq analysis was performed to identify genes that were differentially expressed in *emm*32.2 (112327) relative to

*emm*1.0 and *emm*89.0 (101910 and 127746). RNA from each growth phase was isolated as described in Chapter 2 (Methods). GAS strains were cultured in Todd Hewitt broth-Glucose grown at 37°C and samples were collected after 4h, 6h, and 10h, corresponding to early, mid and late exponential phase, respectively.

Our results showed that the total number of sequencing reads was higher than 10 million for 3 tested isolates as shown in (Table 4-1).

Table 4-1 Number of sequencing reads used in the transcriptomic analysis

GAS *emm* type	Exponential growth phase	Total number of sequencing reads	Total number of filtered reads[1]	Total number of mapped reads for expression analysis[2]	Total number of mapped reads for operon prediction[3]
32.2	early	11785386	11651616	10560214	10521676
1.0	early	10997418	10739588	10316348	10337217
89.0	early	12989311	12851958	11987773	11957945
32.2	mid	11225267	11123317	10227353	10178903
1.0	mid	12405072	12208962	11646889	11674860
89.0	mid	12409153	12302231	11396134	11372263
32.2	late	12033061	11917074	10686688	10639072
1.0	late	11254678	11137167	10745052	10746209
89.0	late	12217600	12053742	6504477	6473947

1 .Filtering was performed with the TrimGalore software.
2. Mapping was carried out using the EDGE-pro pipeline.
3. Operon structure was predicted with the Rockhopper package.

4.2.4 Principal-component analysis (PCA)

To identify whether there is a pattern in the transcriptomic data that separates the samples from each other, principal-component analysis (PCA) was used which clusters genes into groups that represent the highest variation on the data. Our findings of the PCA displayed that the transcription profile of the *emm*32.2 strains and *emm*1.0 and *emm*89.0 samples at the different exponential growth phases cluster separately from each other into distinct groups as shown in Figure 4-4.

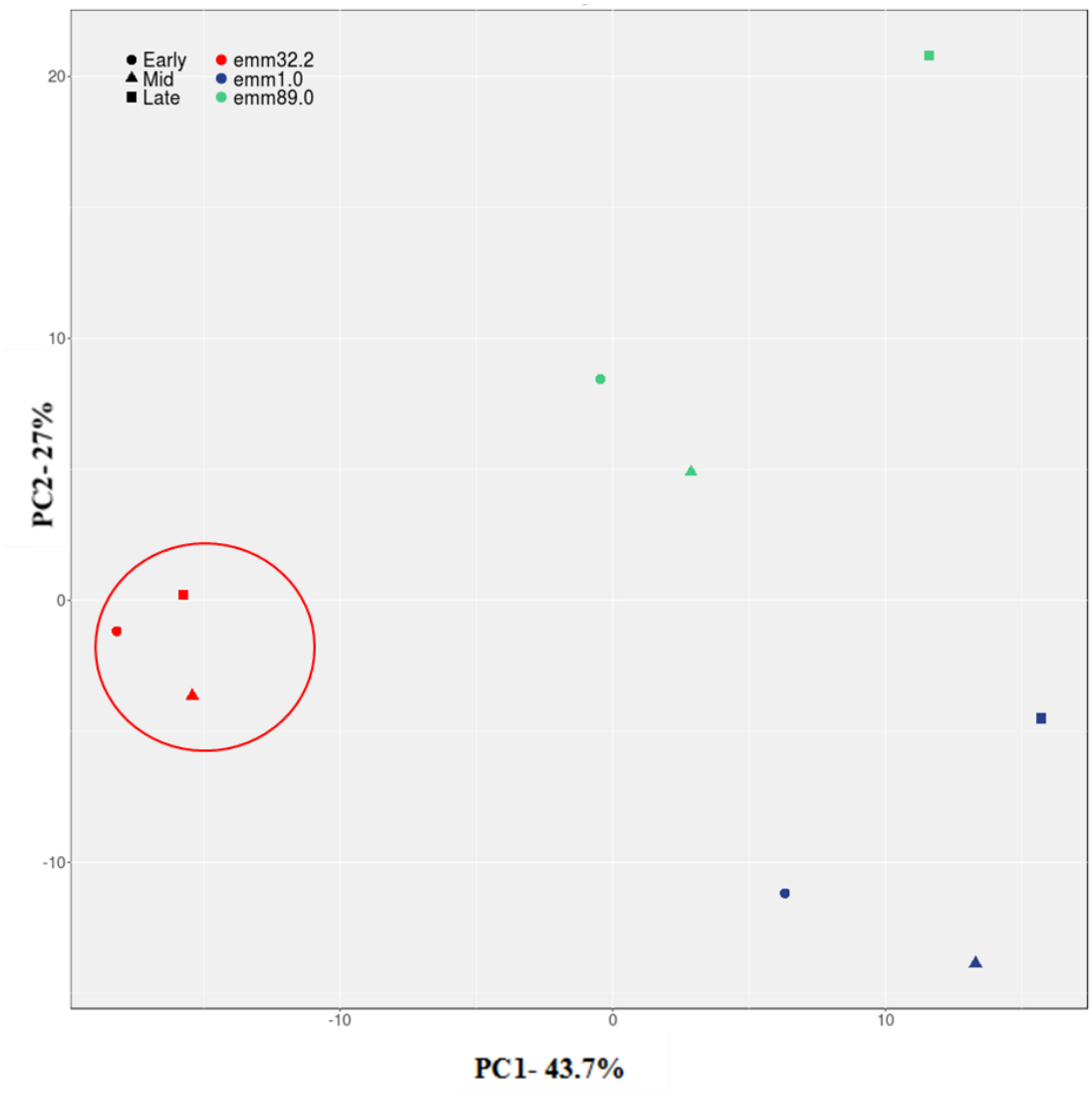


Figure 4-4 Clustering of emm32.2, emm1.0 and emm89.0 samples based on transcriptomic data.

The PCA plot represents regularised logarithm (rlog)-transformed expression values for 1,841 genes from each sample. X and Y axis show principal component 1 and principal component 2 that explain 43.7% and 27% of the total observed variance, respectively.

4.2.5 Clustering of sample-to-sample distance analysis

To identify a gene expression-based degree of similarity between *emm*32.2, *emm*1.0 and *emm*89.0 strains, hierarchical clustering of the transcriptomic data was performed as described in Chapter 2 (Materials and Methods). The rows and columns of the matrix are rearranged independently. As a result, genes with similar expression patterns are nearby. The computed dendrogram (tree) display the relationships among genes. The data exhibited that *emm*32.2*, emm*1.0 and *emm*89.0 strains are separated based on RNA expression (Figure *4-5*)**.** Although the *emm*1.0 and *emm*89.0 strains displayed expression profiles clearly divergent from that of the *emm*32.2sample in all the three time points, the latter was slightly more similar to that of the *emm*89.0 strain. All in all, the heatmap showed that *emm*32.2 strain, *emm*1.0 and *emm*89.0 were genetically distant from each other.

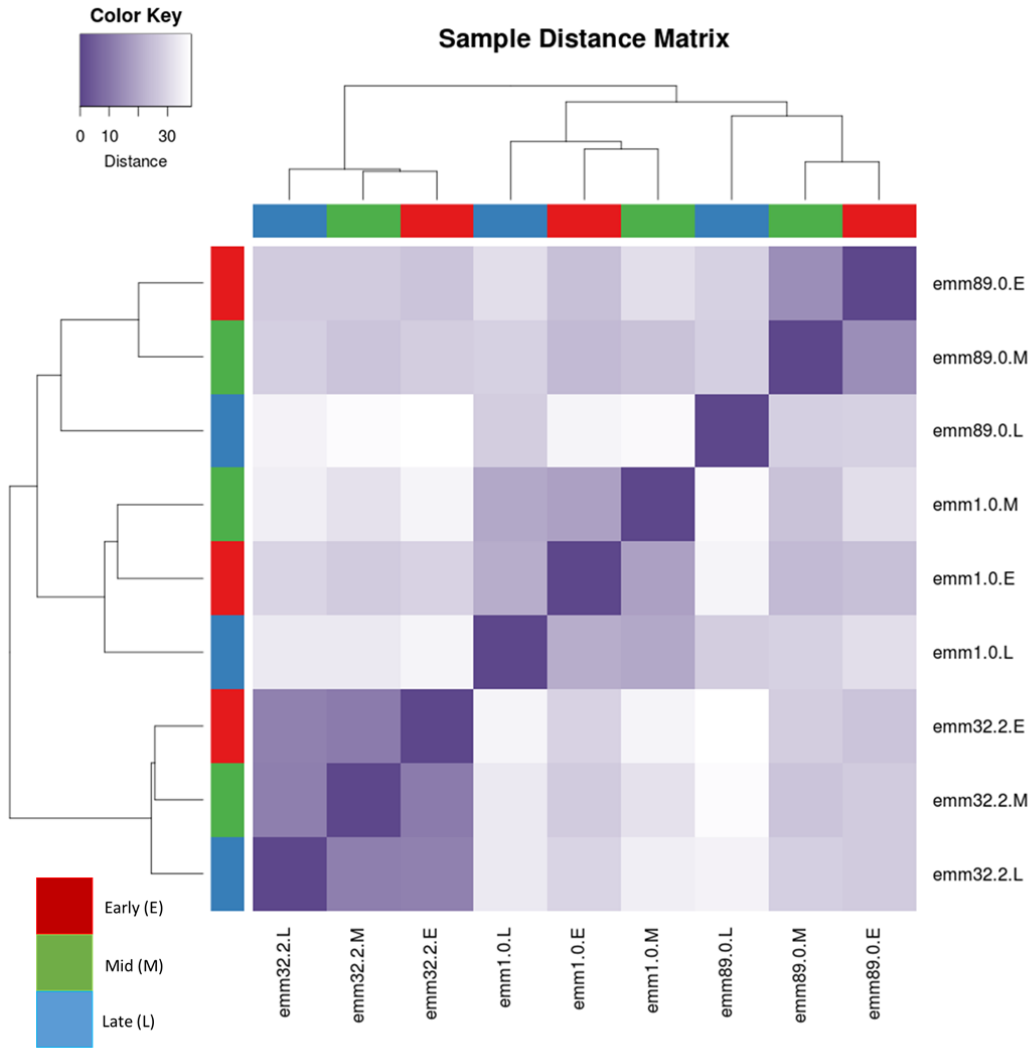


Figure 4-5 Heatmap showing the difference between emm32.2, emm1.0 and emm89.0 transcriptomic samples.

Each cell in the matrix represents relative expression of genes for a given strain based on rlog-transformed values. Colours of the cells reflect the similarity degree between the samples in terms of distance, as indicated in the colour key bar. Distance between the samples was determined using Euclidean measure whereas clustering was performed with the complete agglomeration method.

4.2.6 Pair-wise comparison of gene expression analysis

To summarise the fold change of gene expression alongside the corresponding p-value volcano plots were made, to give an overall view of the differences between *emm*32.2 strain and *emm*1.0 and *emm*89.0 strains (Figure 4-6A and

Figure 4-7A). This information together with biological information about each gene is valuable for identification of important virulence genes for further investigation.

4.2.7 Comparison of the expression of *emm*32.2 and *emm*1.0 genes

The genes differentially expressed according to growth phase, including genes differentially regulated in one or more growth phase, are shown in Figure 4-6A and Figure 4-7A. These genes were categorised based on their differential expression at one or more growth phases. We hypothesised that the *emm*32.2 genes would be differentially expressed under most of the conditions tested. Several of the genes meeting these criteria are discussed below.

As shown in Figure 4-6B, 4 genes were up-regulated during the early and mid-exponential phases of the *emm*32.2 strain compare to *emm*1.0. This number increased to 6 during the late exponential phase. Moreover, *hasABC* genes were shared between the early and late exponential phases (Table 4-2). Both *hasA* and *hasC* genes are involved in cell wall, membrane and envelope biogenesis (COG classification M). However, *hasB* gene is related to cell motility (COG classification: N). On the other hand, 15, 29, and 29 genes were down-regulated during the early, mid, and late exponential phases, respectively (Figure 4-6B.) 7 out of 24 down-regulated genes were

commonly shared between the early, mid, and late exponential phases of growth (Table 4-2). Overall, the number of down-regulated genes was higher than the number of up-regulated genes in *emm*32.2 compared to *emm*1.0, and the late exponential phase was associated with a greater number of differentially expressed genes.

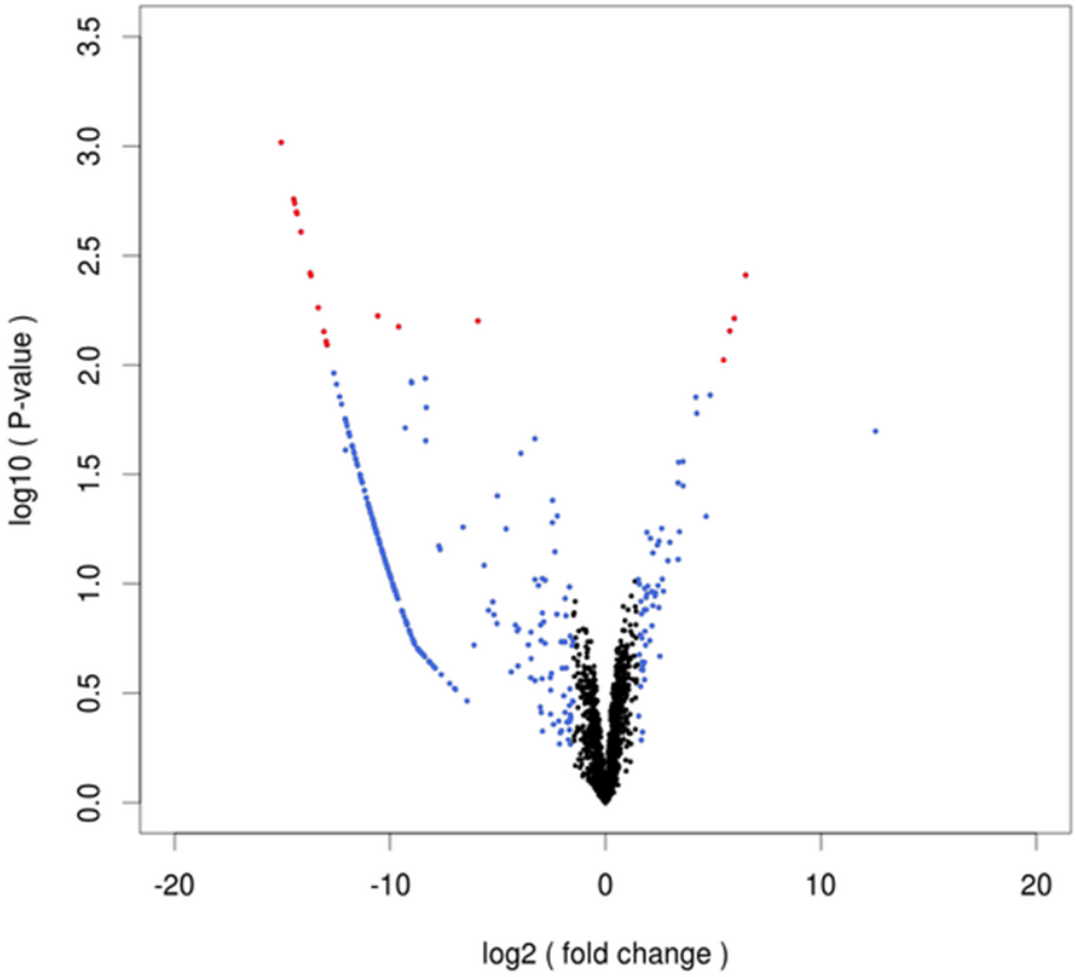
4.2.8 Comparison of the expression of *emm*32.2 and *emm*89.0 genes

Pairwise comparisons between the *emm*32.2 and *emm*89.0 strains at the early, mid, and late exponential growth phases, revealed a number of genes exhibiting either increased or decreased transcription levels. As shown in Figure 4-7B; a set of 8, 8, and 21 genes were up-regulated during the early, mid, and late exponential phase, respectively. A group of 6 up-regulated genes (spy0462, spy1192, spy0463, spy1222, spy0466, spy1193) were shared between the early, mid and late phases. Both Spy0462 and Spy0466 genes associate with translation, ribosomal structure and biogenesis (COG classification: J). Moreover, Spy0463 and Spy1222 genes link to Lipid transport and metabolism (COG classification: I) and in Replication, recombination and repair (COG classification: L), respectively. On the other hand, Spy0769 gene, and Spy0769, spy0164, and spy0165 genes were down-regulated in the *emm*32.2 strain during the early and mid-exponential phase, respectively. The spy0165 gene is related to cell cycle control, cell division, and chromosome partitioning (COG classification: D). The number of down-regulated genes increased to 11 during late phase (

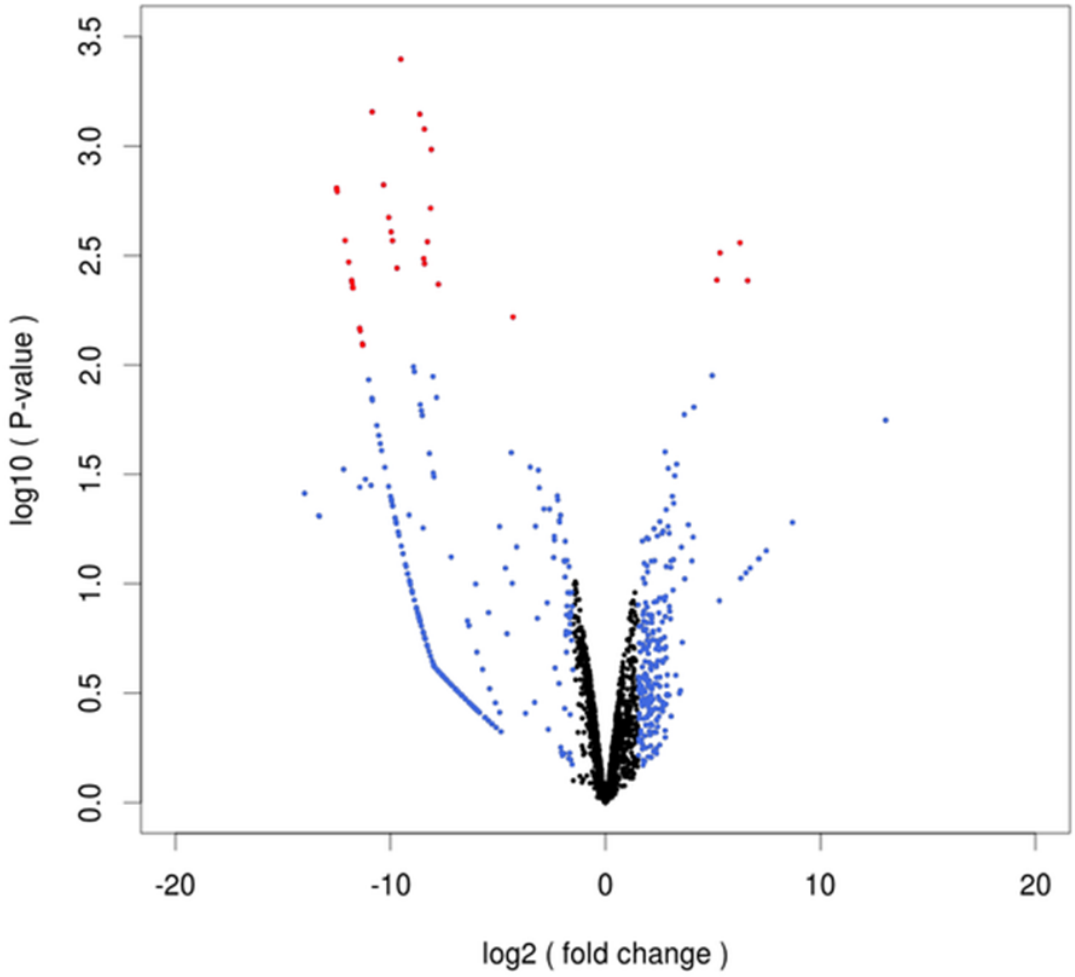
Figure 4-7B). Spy0769 genes were commonly shared between the early, mid, late exponential phases of growth. This gene is a putative cytosolic protein which implicates in defence mechanism (COG classification: V).

Figure 4-7B exhibited that the number of the up-regulated genes was higher than down-regulated genes in *emm*32.2 compared to *emm*89.0. Overall, the growth phase with the greatest number of differentially expressed genes was in late exponential phase.

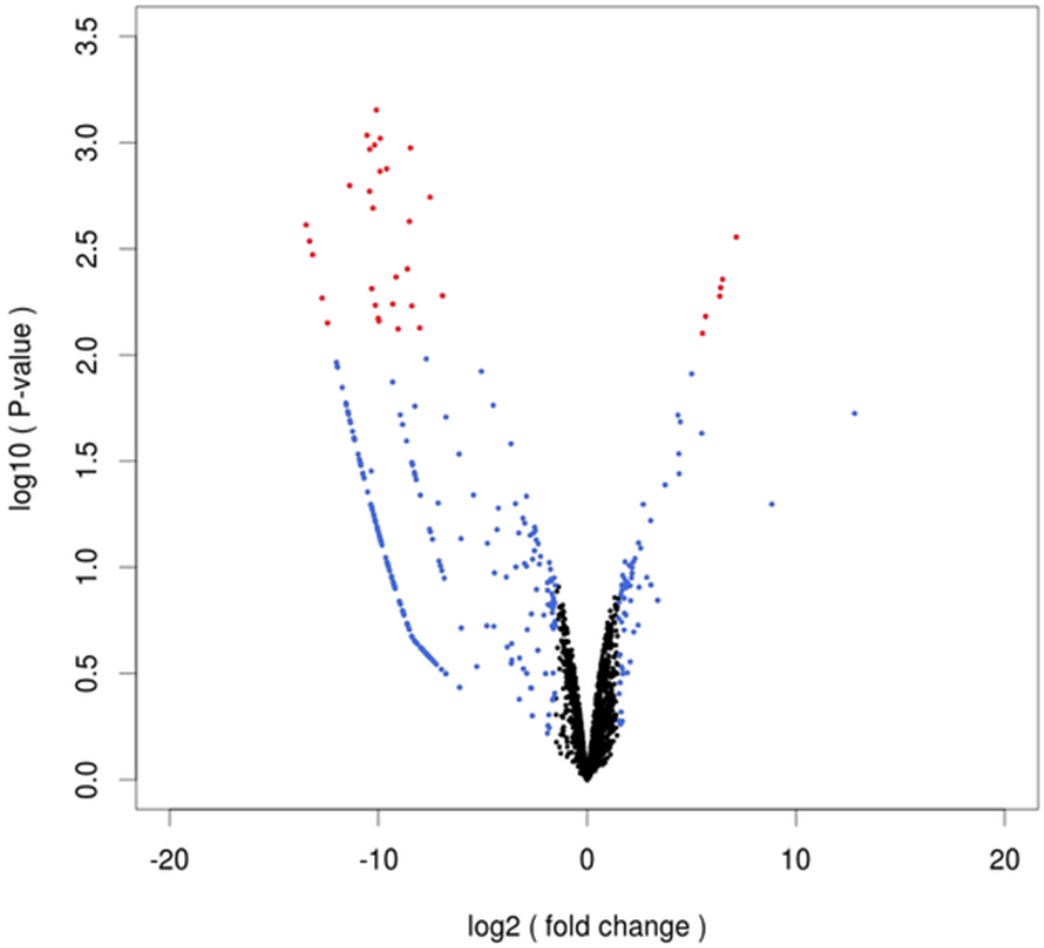
A-i


log10 (P-value)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
-20
-10
0
10
20
log2 (fold change)

A-ii


log10 (P-value)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
-20
-10
0
10
20
log2 (fold change)

A-iii


log10 (P-value)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
-20
-10
0
10
20
log2 (fold change)

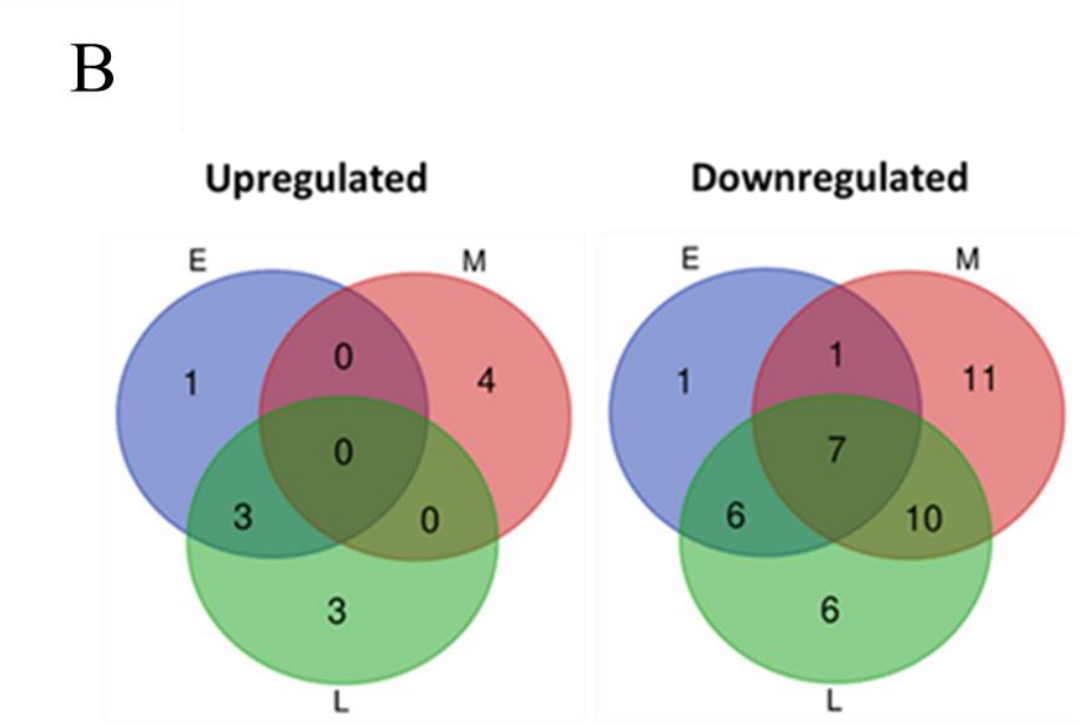
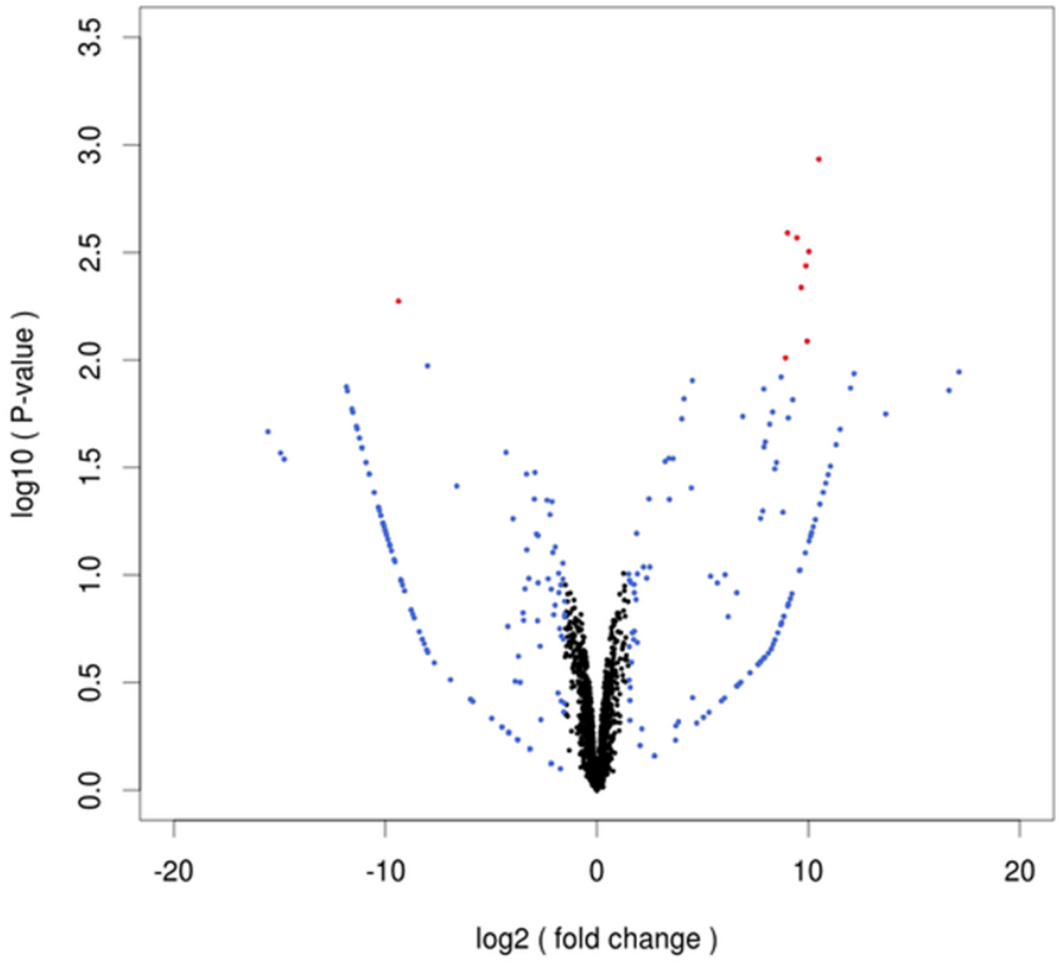


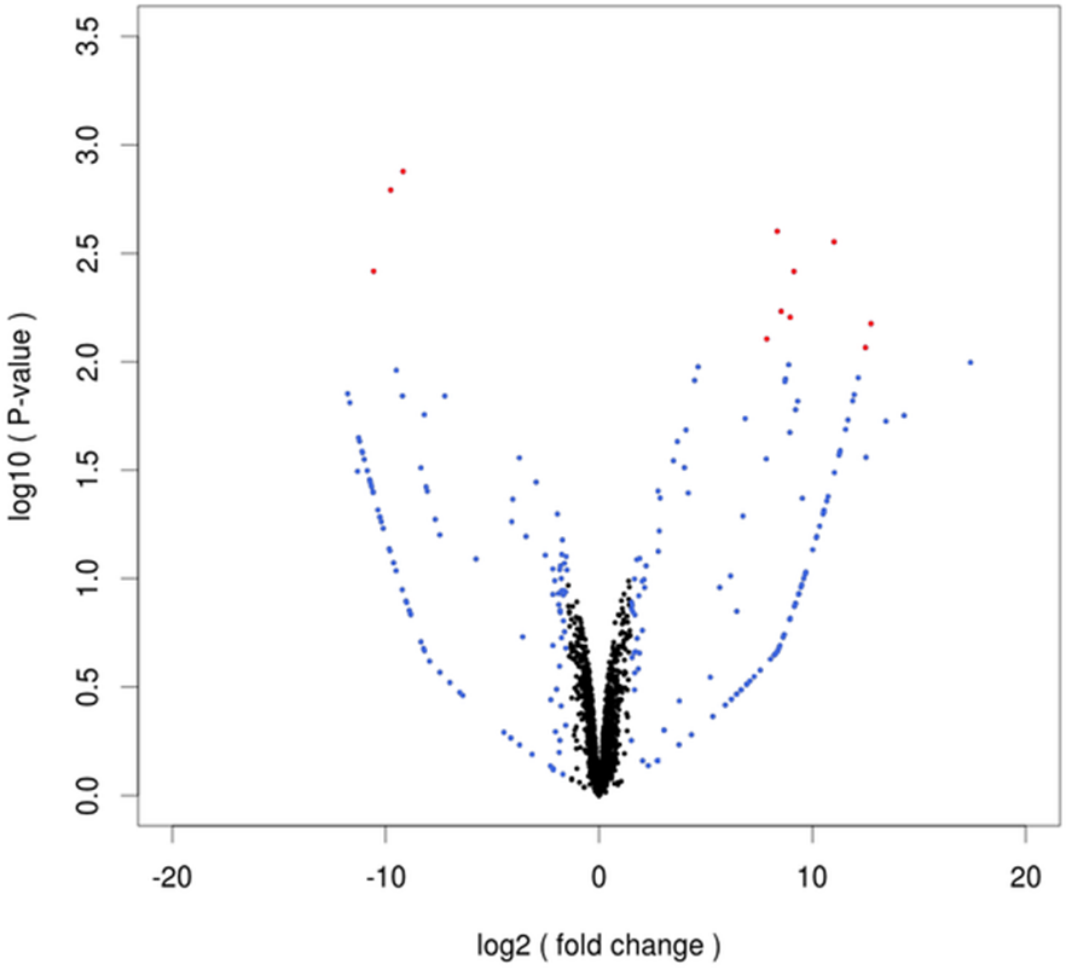
Figure 4-6 Analysis of differential gene expression in emm32.2 compared to emm1.0.

A) Volcano plots showing fold changes and statistically significant differences in expression of *emm*32.2 compared to *emm*1.0 at i) early exponential phase (EEP), ii) mid-exponential phase (MEP), and iii) late exponential phase (LEP). The negative $\log_{10}$ p-values (y-axis) are plotted against $\log_2$ fold changes (x-axis). Black and blue dots represent non-differentially expressed genes whereas red dots correspond to those considered as differentially expressed. Red dots display p-values <0.01 and $\log_2$-fold change (LFC) values > 4 or < - 4. In contrast, black points represent genes that did not pass both the P and LFC cut-off values whereas the blue dots correspond to genes that failed on meeting the cut-off p-value only. B) Venn diagrams representing the number of differentially expressed virulence genes between *emm*32.2 and *emm*1.0 at the early (E), mid (M) and late (L) exponential growth phases: (i) up-regulation and (ii) down-regulation.

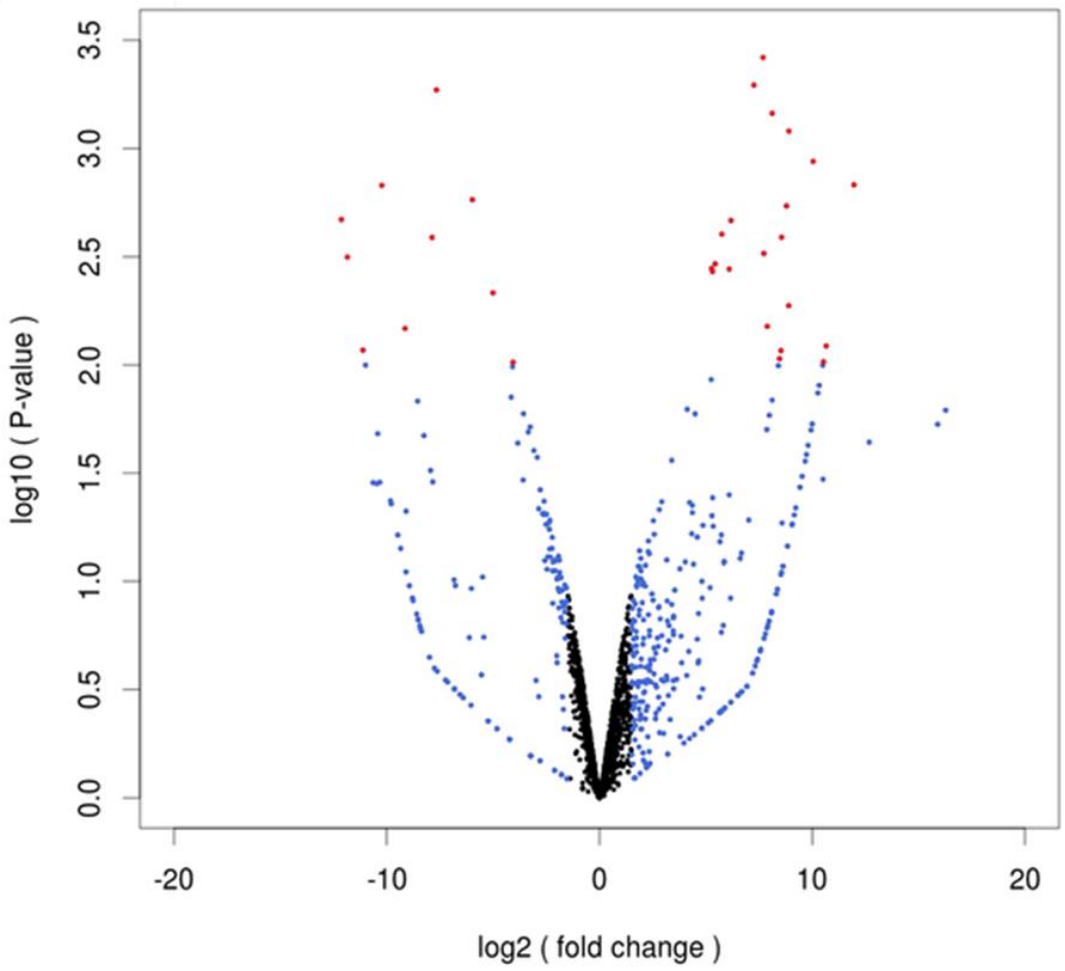
A-i


log10 (P-value)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
-20
-10
0
10
20
log2 (fold change)

A-ii


log10 (P-value)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
-20
-10
0
10
20
log2 (fold change)

A-iii


log10 (P-value)
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
-20
-10
0
10
20
log2 (fold change)

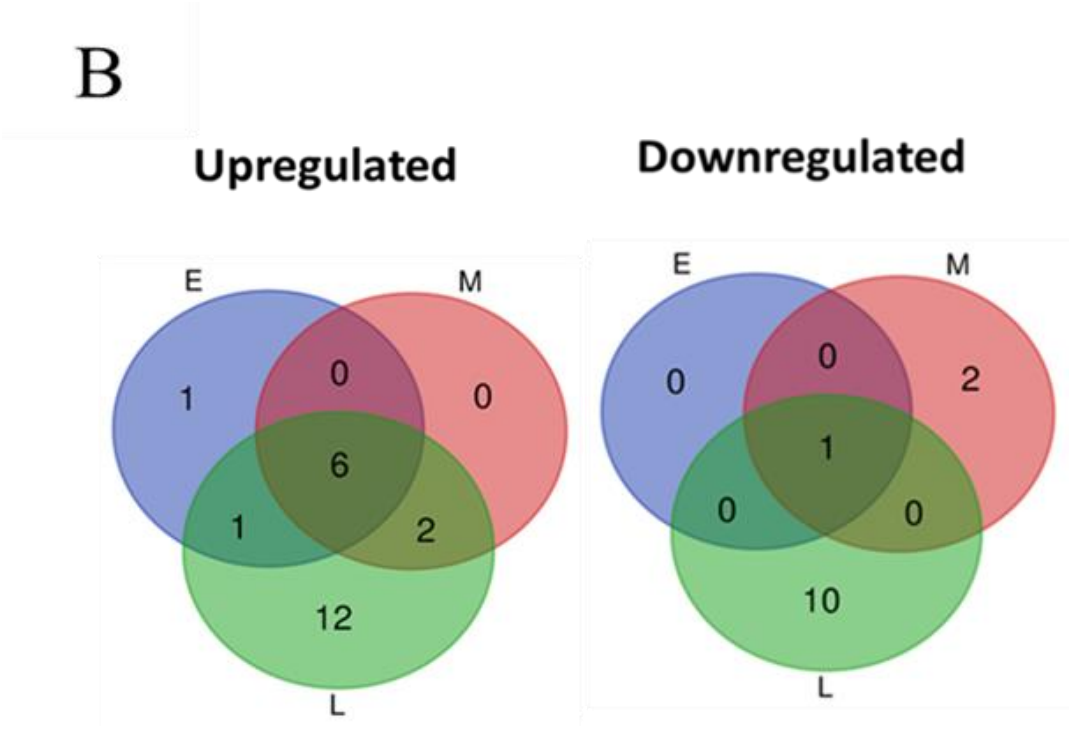


Figure 4-7Analysis of differential gene expression in emm32.2 compared to emm89.0.

A) Volcano plots showing fold change values and statistically significant differences in expression of *emm*32.2 compared to *emm*89.0 at the i) early exponential phase (EEP), ii) mid-exponential phase (MEP), and iii) late exponential phase (LEP). The negative log10 p-values (y-axis) are plotted against log2 fold changes (x-axis).). Black and blue dots represent non-differentially expressed genes whereas red dots correspond to those considered as differentially expressed. Red dots display p-values <0.01 and log_2-fold change (LFC) values > 4 or < - 4. In contrast, black points represent genes that did not pass both the P and LFC cut-off values whereas the blue dots correspond to genes that failed on meeting the cut-off p-value only. B) Venn diagrams representing differentially expressed virulence genes between *emm*32.2 and *emm*89.0 at the Early (E), mid (M) and late (L) exponential phases: (i) up-regulation and (ii) down-regulation.

Depending on the growth phase, several *emm*32.2 strain genes (excluding rRNA and tRNA genes) were differentially expressed. Many genes showed a significant difference between at least one-time point (Figure 4-8). To determine which genes were up- or Down-regulated in the *emm*32.2 strain (112327) relative to *emm*1.0 strain (101910) and *emm*89.0 (127746), we restricted the list of genes to only those that showed a significant difference in at least one growth phase (p-value =< 0.01; $\log_2$-fold change > 4 or<- 4).

4.2.9 Differential expression genes of *emm*32.2 strain compared to *emm*1.0 strain

This restriction generated a list of 11 genes with higher expression and 42 genes with lower expression in at least one growth phase when comparing *emm*32.2 to *emm*1.0 (Table 4-2 and Table 4-3). These genes were predicted to make up a total of 17 operons. Notably, the number of differentially expressed genes was greater during the mid and late exponential phases. The functional breakdown of these genes on the basis of their assigned Cluster of Orthologous Groups (COG) classification is shown in Table 4-2 and Table 4-3.

4.2.10 Differential expression genes of *emm*32.2 strain compare to *emm*89.0 strain

A group of 22 genes with higher expression and another group of 13 genes with significantly lower expression were found in at least one growth phase when comparing *emm*32.2 to *emm*89.0 strain (Table 4-2 and Table 4-3). These genes were

predicted to make up a total of 15 operons. The genes and their classification to COG functional groups are shown in detail in Table 4-2 and Table 4-3. The number of differentially expressed genes was greater during the late exponential phase.

4.2.11 Differential expression of genes related to metabolism.

Among Down-regulated transcripts genes in *emm*32.2 compared to *emm*1.0, the majority were involved in carbohydrate transport and metabolism (COG classification: G). As shown in Table 4-2 Table 4-3, through comparison of transcription levels and genome location Spy1064 (*malC*) and Spy1066 (*amyB*), Spy1062 (*malA*), Spy1067 (*malX*) shared the same operon which means these genes display similar transcription levels and they located next to each other. *malC* was significantly Down-regulated during mid-exponential phases, however, *amyB, malA,* and *malX* were significantly down-regulated during mid and late exponential phases.

Overall pairwise comparison between the *emm*32.2 and *emm*1.0 and *emm*89.0 during exponential growth phases, revealed relatively 76 genes that showed a significant difference between at least one growth phase.

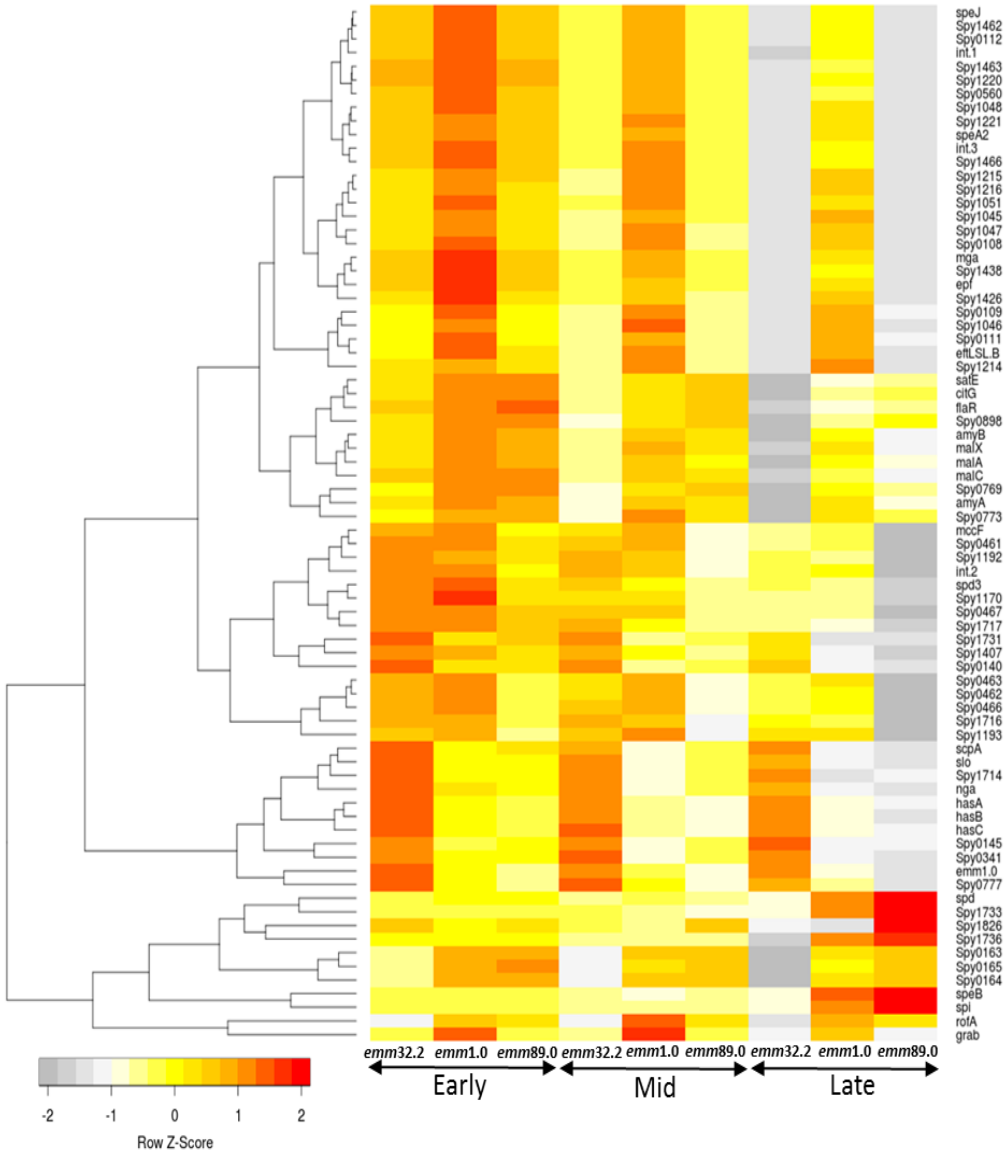


Figure 4-8 Expression profiles of significant Group A Streptococcus genes during the exponential growth phase.

Each row in the heatmap represents a gene, and each column represents one individual GAS strain in a specific growth phase. Variance stabilizing transformation (VST) values, corresponding to the relative expression of 76 GAS genes, were used to construct the heatmap. The colour scale at the bottom left is relative to the average expression value of a gene for the given strains in a row, with grey and red colours indicating the lowest and highest difference from the average. Differences are represented as z-scored values. Distance between the genes, represented as a tree in the left side of the figure, was computed using Euclidean measure whereas clustering was carried out with the complete agglomeration method.

Table 4-2 Up-regulated differentially expressed genes in the emm32.2 strain compared to emm1.0 or emm89.0.

M5005 locus	Gene [I]	Gene product [II]	Log fold change [III, IV]			COG[V]	Operon[VI]
			Early	Mid	Late		
***emm*1.0**							
Spy0145	–	hypothetical protein	6.51	–	–	–	–
Spy1853	*hasC*	UTP--glucose-1-phosphate uridylyltransferase	5.98	–	7.14	M	–
Spy1851	*hasA*	hyaluronan synthase	5.77	–	6.48	N	b
Spy1852	*hasB*	UDP-glucose 6-dehydrogenase	5.48	–	6.39	M	b
Spy1731	–	putative cytosolic protein	–	6.62	–	–	–
Spy0140	*ifs*	Protein immunity factor for Nga	–	6.26	–	–	c
Spy0141	*slo*	streptolysin O	–	5.33	–	–	c
Spy1715	*scpA*	C5A peptidase precursor	–	5.18	–	O	d
Spy1714	–	cell surface protein	–	–	6.35	–	d
Spy1719	*emm*1.0	M protein	–	–	5.67	D	–
Spy0341	–	Lactocepin	–	–	5.52	O	–
***emm*89.0**							
Spy1716	–	Transposase	10.50	–	8.91	X	r
Spy0462	–	hypothetical protein	10.03	7.86	7.73	J	s
Spy1192	–	phage protein	9.94	8.96	7.88	–	t
Spy0463	–	putative cytosolic protein	9.88	8.54	8.79	I	s
Spy1222	*int.2*	Integrase	9.66	11.01	8.56	L	–
Spy0466	–	hypothetical protein	9.46	9.13	10.04	J	–
Spy1193	–	phage protein	9.02	8.35	8.12	–	t
Spy0461	–	putative cytosolic protein	8.92	–	–	K	s
Spy1407	–	Esterase	–	12.74	11.96	I	–
Spy1717	–	Transposase	–	12.48	10.66	–	r
Spy0467	–	Transposase	–	–	10.53	X	u
Spy1169	*spd3*	Streptodornase	–	–	8.90	–	–
Spy1170	-	putative membrane associated protein	–	–	8.53	–	–
Spy0464	*mccF*	microcin C7 self-immunity protein	–	–	8.46	–	v
Spy0341	–	Lactocepin	–	–	7.69	O	–
Spy0145	–	hypothetical protein	–	–	7.26	–	–
Spy0139	*Nga*	NAD glycohydrolase	–	–	6.18	–	c
Spy0140	*ifs*	Protein immunity factor for Nga	–	–	6.10	–	c
Spy1715	*scpA*	C5A peptidase precursor	–	–	5.75	O	d
Spy0141	*slo*	streptolysin O	–	–	5.44	–	c
Spy1714	-	cell surface protein	–	–	5.31	–	d
Spy0777	*sclB or scl2*	collagen-like surface protein	–	–	5.26	–	–

I. Gene name. II. Gene descriptions derived from the gene bank. III. Numbers represent log2 fold change between the *emm*32.2 and *emm*1.0 and *emm*89.0 strains. IV. Missing numbers indicate that the gene does not satisfy the cut-off p-value (p= <0.01). V. Capital Letters refer to the functional categories of the assigned Cluster of Orthologous Group (COG) http://www.ncbi.nlm.nih.gov/COG :[D] Cell cycle control, cell division, chromosome partitioning, [M] Cell wall/membrane/envelope biogenesis,[N] Cell motility, [O] Post-translational modification, protein turnover, and chaperones,[X] Mobilome :prophages, transposons , [J] Translation, ribosomal structure and biogenesis, [K] Transcription, [L] Replication, recombination and repair,[I] Lipid transport and metabolism. VI. Identical lower case letters indicate genes predicted to be in the same operon according to an analysis by the program Rockhopper.

Table 4-3 Down-regulated differentially expressed genes in emm32.2 strain compared to emm1.0 or emm89.0

M5005 locus	Gene [I]	Gene product [II]	Log fold change [III, IV]			COG[V]	Operon[VI]
			Early	Mid	Late		
***emm*1.0**							
Spy1106	*grab*	protein G-related alpha 2M-binding protein	-5.92	–	–	G	–
Spy0769	–	putative cytosolic protein	-9.60	-9.97	-13.16	V	e
Spy1720	*mga*	M protein *trans*-acting positive regulator	-10.57	-8.42	–	–	–
Spy1047	–	phage protein	-12.93	–	-10.42	–	–
Spy1438	–	phage protein	-12.96	-12.50	-9.06	–	f
Spy0561	*Epf*	putative extracellular matrix binding protein	-13.07	–	-10.14	–	–
Spy0108	–	signal peptidase I	-13.33	-10.85	-11.37	–	g
Spy1046	–	phage protein	-13.67	–	-10.55	–	–
Spy1426	–	phage protein	-13.70	-9.91	-13.46	X	h
Spy0110	*eftLSLB*	putative exported protein	-14.13	-8.42	-10.41	S	g
Spy0165	–	Transposase	-14.32	-9.52	-9.92	–	q
Spy0163	–	3'-phosphoadenosine 5'-phosphosulfate	-14.34	–	-8.51	R	i
Spy0109	–	fibronectin-binding protein	-14.43	–	-9.92	–	g
Spy0111	–	hypothetical protein	-14.47	-8.10	-9.61	–	g
Spy0164	–	parB-like nuclease	-15.05	–	-10.10	D	i
Spy0106	*rofA*	transcriptional regulator	–	-4.30	–	–	–
Spy1466	–	phage protein	–	-7.77	–	–	j
Spy1051	–	phage protein	–	-8.13	-8.02	–	k
Spy1467	*int.3*	Integrase	–	-8.28	–	L	–
Spy1048	–	phage protein	–	-8.45	-12.43	–	–
Spy0773	–	hypothetical protein	–	-8.64	-8.61	–	–
Spy0996	*speA2*	Enterotoxin	–	-9.70	-10.01	–	–
Spy1221	–	phage protein	–	-10.08	-12.70	–	l
Spy1045	–	transcriptional regulator	–	-10.32	-10.26	K	–
Spy1220	–	phage protein	–	-11.29	–	–	l
Spy1064	*malC*	maltose transport system permease protein	–	-11.30	–	G	m
Spy0560	–	transcriptional regulator	–	-11.41	–	–	n
Spy1463	–	phage protein	–	-11.43	–	–	–
Spy1462	–	phage protein	–	-11.74	–	–	–
Spy0112	–	Transposase	–	-11.75	–	X	–
Spy1052	*int.1*	Integrase	–	-11.78	–	L	–
Spy0356	*speJ*	exotoxin type J precursor	–	-11.80	–	–	–
Spy1062	*malA*	maltodextrose utilization protein	–	-11.94	-9.98	G	m
Spy1066	*amyB*	neopullulanase/cyclomaltodextrinase/maltogenic	–	-12.11	-10.32	G	m
Spy1065	*amyA*	alpha-amylase	–	-12.47	-13.30	G	m
Spy1067	*malX*	maltose/maltodextrin-binding protein	–	-12.49	-9.31	G	–
Spy1736	–	hypothetical protein	–	–	-6.93	–	–
Spy1735	*SpeB*	streptococcal pyrogenic exotoxin B	–	–	-7.53	–	p
Spy1216	–	phage protein	–	–	-8.40	–	a
Spy1734	*Spi*	streptopain protease inhibitor	–	–	-8.47	–	o,p
Spy1215	–	phage protein	–	–	-9.16	–	–
Spy1214	–	phage protein	–	–	-10.18	–	–
***emm*89.0**							
Spy0769	–	putative cytosolic protein	-9.38	-10.56	-12.14	V	e
Spy0164	–	parB-like nuclease	–	-9.18	–	D	i
Spy0165	–	Transposase	–	-9.76	–	–	q
Spy1738	*Spd*	phage-associated deoxyribonuclease	–	–	-4.07	–	–
Spy1826	–	putative membrane associated protein	–	–	-5.01	S	w
Spy1733	–	hypothetical protein	–	–	-5.98	–	x
Spy1736	–	hypothetical protein	–	–	-7.66	–	–
Spy0773	–	hypothetical protein	–	–	-7.87	–	–
Spy0163	–	3'-phosphoadenosine 5'-phosphosulfate	–	–	-9.03	R	i
Spy0892	*satE*	SatE	–	–	-9.14	–	y
Spy0898	–	2-(5"-triphosphoribosyl)-3'-dephosphocoenzyme	–	–	-10.23	H	z
Spy1701	*flaR*	DNA topology modulation protein	–	–	-11.12	F	aa
Spy0899	*citG*	transcriptional regulator, GntR family	–	–	-11.85	K	z

I. Gene name. II. Gene descriptions derived from the gene bank. Numbers represent log2 fold change between the *emm*32.2 and *emm*1.0 and *emm*89.0 strain. IV. Missing numbers indicate that the gene does not satisfy the cut-off p-value (p= <0.01). V. Capital Letters refer to the functional categories of the assigned Cluster of Orthologous Group (COG) http://www.ncbi.nlm.nih.gov/COG : [F] Nucleotide transport and metabolism,[G] Carbohydrate transport and metabolism,[H] Coenzyme transport and metabolism, [X] Mobilome : prophages, transposons ,[R] General function prediction only,[S] Function unknown,[V] Defence mechanism,[D] Cell cycle control, cell division, chromosome partitioning,[K] Transcription,[L] Replication, recombination and repair. VI. Identical lower-case letters indicate genes predicted to be in the same operon according to an analysis by the program Rockhopper.

4.2.12 Differential expression of genes related to virulence

Our RNA-seq analysis gave a comprehensive snapshot of *emm*32.2 gene expression relative to *emm*1.0 and *emm*89.0. A number of genes have been previously identified as GAS virulence factors [157, 186]. I will discuss only the significant virulence factor determinants, as per listed in Table 4-2 and Table 4-3. The transcriptome expression profiles for these genes are shown in the heatmap in Figure 4-9**.** A number of the virulence genes showed distinct patterns of differential expression. These genes are identified among the up- or down-regulated genes identified in our data.

Our data presented an equal number (n=6) of significantly up-regulated and down-regulated genes when comparing *emm*32.2 to *emm*1.0. Capsule genes (hyaluronic acid synthesis *hasABC* genes) were significantly up-regulated during both early and late-exponential phases. Noteworthy, the data exhibited that *hasA* and *hasB* share the same operon. Moreover, C5a peptidase (*scpA*) and Streptolysin O (*slo*) were significantly up-regulated during the mid-exponential phases. The highest number of up-regulated genes showed greater expression during the late exponential growth phase.

Conversely, protein-G related-α2-macroglobulin-binding protein (*grab*) and M protein trans-acting positive regulator (*mga*) were significantly down-regulated during the early, and early and mid-exponential phases, respectively. Streptococcal pyrogenic erythrogenic toxin B (*sepB*) also was significantly down-regulated during the late exponential phases. In addition, Alpha-amylase (*amyA*) was significantly down-regulated during the mid and late exponential. Streptococcal pyrogenic exotoxins (*speJ*), and streptococcal pyrogenic exotoxins (*speA*) were significantly down-

regulated during the mid and mid and late exponential, respectively. Most of the down-regulated genes showed greater expression during late exponential growth phase.

When comparing *emm*32.2 to *emm*89.0, our results revealed 5 significantly up-regulated genes. Adhesin gene such as spy0777 showed greater expression during late exponential growth phase, along with a *scpA*, *slo*, and nga which involved in GAS immune evasion mechanisms and cytolysis respectively. Nevertheless, *spd* were significantly down-regulated during late-exponential phases.

Overall, our data displayed a list (n=16) of well-known virulence factors which showed different patterns of differential expression in *emm*32.2 strain (112327) when compared to *emm*1.0 and *emm*89.0 strains.

Figure 4-9Expression profiles of Group A Streptococcus virulence factors during exponential growth phase.

Each row in the heatmap represents a gene, and each column represents one individual GAS strain. VST values, corresponding to the relative expression of 47 GAS genes, were used to construct the heatmap. Each cell in the matrix represents the relative expression of a gene for a given strain in VST values.

4.3 Discussion

In the previous chapter, I have described the differential phenotypic properties of *emm*32.2 strains compare to *emm*1.0 and *emm*89.0 strains. In this chapter. We have been investigating the genomic characteristics of this subtype to explain the clinical phenotype. We have presented transcriptomic data that gives an overview of differentially expressed genes of *emm*32.2 strain. PCA, heatmap data and volcano figures have revealed that *emm*32.2 strain had significantly altered expression of a wide range of genes.

Environmental stimuli and growth phases are all believed to influence the expression of virulence factors [191]. Virulence genes were shown to present growth-phase specific expression profiles in bacterial species such as GAS [190]. Therefore, in my project, I aimed to investigate the absence/presence and variations in gene expression in accordance with the metabolic changes occurring during GAS growth in liquid broth. This was deemed an appropriate approach as depletion of nutrients and decreasing pH conditions occurring during the bacterial growth mimic the environment to which the pathogen may be exposed during infection [192].

In my project, I focused on the exponential phase (early-mid-late) of bacterial growth. The exponential phase is believed to coincide with the stage of the infection when bacterial numbers reach a peak in density *in vivo*. It corresponds to a period of active cell division and is characterised by the highest overall number of transcripts [193]. For instance, several key genes, such as M protein, C5a peptidase, capsule operon, and

mga which regulates a number of virulence genes are maximally expressed during the exponential phase [168, 190].

First, we tested the hypothesis that the *emm*32.2 strain differed in growth characteristics from *emm*1.0 and *emm*89.0 strains over 12 hours period (Figure 4-3). Our data showed that the growth rates of *emm*32.2 strain and the *emm*1.0 and *emm*89.0 strains were similar during the exponential growth phase under our experimental conditions indicating differential growth rate will not affect our results. Our results suggest that differential growth rate during the exponential phase is not responsible for highly invasive clinical phenotypes as caused by *emm*32.2.

Pairwise comparisons between the *emm*32.2 strain and *emm*1.0, as well as between *emm*32.2 strain and *emm*89.0 strain revealed that 76 genes exhibited either increased or decreased transcriptional levels. There was a difference in the expression levels between the early vs. mid vs. late exponential growth phases. However, this difference was most obvious between the early and late exponential phases. Most interestingly, we observed that the late exponential phase was associated with the highest number of differentially expressed genes. These data suggest that *emm*32.2 strain adapts to changes in the availability of metabolic substrates during the late exponential phase by extensive changes in gene expression-hence further understanding of the changes occurring during this phase may help reveal some key aspects of *emm*32.2 virulence.

We sought to categorise the differentially expressed genes on the basis of their biological functions. Among the down-regulated genes, the majority of genes was found to be involved in carbohydrate transport and metabolism (COG classification:

G). Interestingly, *malC, malA,* and *malX*, which are located within the same operon, were significantly down-regulated in *emm*32.2 compared to *emm*1.0. These genes encode enzymes involved in the entry of carbon sources into the glycolytic pathway [157]. The function of these genes is Maltodextrose utilisation protein, Maltose/maltodextrin transporter membrane protein, and Maltose/maltodextrin transporter subunit, respectively. Based on this observation, we might expect to observe a difference in the growth phase between *emm*32.2 isolate and *emm*1.0 isolate when maltose is the only carbon source in the liquid broth. Therefore, further investigation is needed to assist further whether the observation from the transcriptomic analysis can be linked to phenotypic analysis.

Despite these differences in genes involved in metabolism, the most significant differences between *emm*32.2, *emm*1.0 and *emm*89.0 strains were in the expression of virulence factors. The analysis was carried out on the differential expression of 16 well-studied virulence factors and indicated that 9 virulence factors associated with invasive potential (*hasABC*, *slo*, *emm*1, *scpA*, *nga*, spy0777, *spd3*) are up-regulated in *emm*32.2 compared to *emm*1.0 and *emm*89.0. In contrast, 7 genes including factors, *mga, speB, speJ, speA2, amyA, grab, spd*, were shown to be down-regulated in *emm*32.2 when compared to *emm*1.0 and *emm*89.0 of strain.

A number of *in vivo* studies reported that the expression of GAS capsule was a requirement for full virulence [79, 84]. Our data displayed that *hasABC* genes where up-regulated in *emm*32.2 strain. These genes responsible for capsule production which are regulated by a highly conserved three-gene namely: *hasA,* encode the hyaluronate

synthase, *hasB* encoding UDP-glucose dehydrogenase, and *hasC* encodes UDP-glucose pyrophosphorylase [74, 75]. These genes are essential virulence factors for GAS which prevents opsonised mediated phagocytosis [73, 82], obstructs the antibody's access to surface epitopes [81], helps to inhibit complement deposition [73]. These results are consistent with phenotypic data presented in Chapter 3 (Figure 3-1) showing that capsule of *emm*32.2 strain was significantly higher than *emm*1.0 capsule.

Slo encodes SLO and *nga* encodes streptococcal NAD+ glycohydrolase (SPN) they are crucial virulence factors of GAS which facilitate pathogeneses, including escaping from the endosome, resistance to phagocytic killing and promotion of intracellular bacterial survival [111, 194, 195] as well as inhibition of neutrophil chemotaxis [196]. *In vivo* studies carried out in assessing the impacts of *slo* as causative agents for mice infections, as well as the resistance to bactericidal impacts of human polymorphonuclear leukocytes (PMNs), finding that both *slo* and nga are vital determinants of GAS pathogenesis [181]. Importantly, our data showed nga and *slo* share the same operon. This supports the literature review where the nga gene is co-expressed with *slo* [11]. Our data showed high expression of both *slo* and nga and there were similar findings recently described by Turner et al. [34] who found alteration in early M89 strains to the modern M89 strains by acquiring a NADase and streptolysin O locus. These results are in line with phenotypic data presented in Chapter 3 (Figure 3-5) showing that *slo* production of *emm*32.2 strain was significantly higher than *emm*1.0.

M protein encoded the *emm* gene, is a significant virulence factor [86]. As a result of ability M protein to bind to multiple human proteins, including C4b-binding protein (C4BP) [88, 89], does not M proteins only promote bacterial aggregation and adherence to host cells but it also avoids complement and phagocytosis [90]. A former study has displayed the essential role of the M protein in pathogenesis by showing that the lack of M protein GAS strain is susceptible to phagocytosis [197]. Furthermore, The streptococcal C5a peptidase gene (*scpA*) is a proteolytic enzyme [198]. The function of *scpA* is to split the human serum chemotaxin C5a at the polymorphonuclear leukocyte (PMNL) binding site which leads to the obstruction of phagocyte recruitment to the site of infection [11, 199]. *In vivo* studies, C5a has been shown to induce a potent serum antibody response that enhances bacterial clearance from the nasopharynx [200]. Streptococcal collagen-like scl2 is known as *sclB* which encode collagen-like proteins. It is an adhesion protein belonging to the collagen superfamily [201, 202]. Even though the specific function of this gene in pathogenesis is not fully understood, the data suggests that *scl2* interacts with T cells and cause hyper-activation of the immune response [203]. *Spd3* encodes putative secreted DNases is prophage which enhances evasion of the innate immune system [204].

In comparison to this, our data displayed low expression of 7 virulence factor genes in *emm*32.2 strain, for example, *grab* and *mga*. The *grab* gene encode GRAB, a surface anchored protein that binds 2-macroglobulin, a protects virulence factors from degradation by both host and bacterial proteases [205]. Mga, encoded within the *emm* cluster, is a critical positive transcriptional regulator of virulence genes found in all

GAS strains. The role of *mga* in GAS virulence has been demonstrated in numerous *in vivo* models of GAS disease, including several invasive disease models [206].

SpeB reduces phagocytic activity and degrades both GAS factors (including M1, SLO, and secreted inhibitor of complement) and host molecules (including the cytokines, chemokines, and immunoglobulins) [95, 179]. *In vivo* studies [97] propose that SepB modulates the host inflammatory response through a number of mechanisms including degradation of complement factors [97, 98]. It was suggested that this virulence factor degradation is controlled by altered *speB* expression during *in vivo* infection [97]. Although SpeB protein sequence is highly conserved among GAS isolates, the expression of the *speB* gene is not [207]. In line with our transcriptomic data, phenotypic analysis data from chapter 3 (Table 3-3) which showed that streptococcal cysteine proteinase activity of *emm*32.2 strain (112327) was undetectable. Spy1065 (*amyA*), which encodes a freely secreted putative cyclomaltodextrin α-glucanotransferase. *AmyA* gene is an extracellular α-glucan degrading enzyme involved in virulence[208]. Previous *in vivo* study sought to define the molecular mechanisms by which GAS degrades α-glucans and to determine if α-glucan degradation contributes to GAS host-pathogen interaction. This study demonstrates that *amyA*-mediated α-glucan digestion influences GAS transepithelial migration and that GAS has adapted to use a host enzyme to initiate α-glucan catabolism. The same study showed that strains lacking *amyA* strain were less virulent following mouse mucosal challenge [208]. Our results are consistent with a previous study showing that in comparison to pharyngitis strain, invasive M1 strains have a high level of *amyA* transcript. *SpeA* is associated with bacteriophages.The *speA2* variant evolved from

SpeA1 and is typically found in M1 isolates [209], and many studies have shown an association between *speA* and invasive disease. [210-212]. However, *speJ* is chromosomally encoded and is absent in a number of GAS isolates [213, 214].

We identified a series of genes that were down-regulated (Table 4-3) in *emm*32.2 isolate (112327) compared to *emm*1.0: 101910 and *emm*89.0: 127746 (from 0 to 5 aligned reads), however, it is still not clear whether such down-regulation is due to gene regulation or to the absence of these genes from the *emm*32.2 genome. To assess this possibility further, we used blast to compare the sequences of those down-regulated genes from the reference genome which used for the transcriptomic analysis to the ones reported by Cornick et al. [146] for the *emm*32.2 strain (112327) analysed in this work. We found that 38 of the 42 putatively down-regulated genes, identified when comparing *emm*32.2 to the *emm*1.0, strain are absent from the collection of genes reported by Cornick et al. Likewise, 10 of the 13 genes identified when comparing *emm*32.2 to *emm*89.0 were not found in the Cornick et al. dataset. However, the remaining 4 genes from the down-regulated list of *emm*32.2 relative to *emm*1.0 (*rofA, grab, spi, speB*) and 3 genes from the down-regulated list of *emm*32.2 relative to *emm*89.0 (Spy_1733, Spy_1826 and *spd*) were present in the *emm*32.2 strain, suggesting that their expression is repressed at the time point where they were identified for reasons that remain to be explored.

Transcriptional regulatory networks are playing a pivotal role in bacterial pathogenesis by modifying gene expression in reaction to environmental stimuli [215]. However, the understanding of how bacterial transcriptional regulatory networks function during

host-pathogen interaction is an area of increasing research interest [215]. Earlier studies reported that these regulators were typically expressed in a growth phase-dependent manner [193]. Identification of these regulatory networks may help towards the design of therapeutic strategies aimed at inhibiting the expression of key virulence factor [193]. My results showed the differential expression of transcription factors that were shown to modulate the expression of key virulence factors, namely *slo, nga and speB*.

CovR/S is one of the most important transcription regulators and is known to modulate the expression of up to 15% of the GAS transcriptome, mainly in the late exponential [216]. *CovR/S* regulates the expression of the *speB* and downregulates multiple virulence factors, including hyaluronic acid capsule and *slo* [11]. My data failed to show any significant change in the expression of *covR/S*. Since the data was unable to exhibit any significant difference in the expression the *covR/S*, it is possible that the shift in capsule genes expression was due to effects from other regulators. We also investigated the expression of *rocA* genes as a previous study showed that *rocA* negatively regulates capsule production [78]. Our data failed to illustrate any significant change in *rocA* gene expression. We also took an interest in *slo*R, which was shown to downregulate the expression of *slo* [217]. Here again, our data did not display any significant changes in the expression of *slo*R.

Multiple GAS transcription factors have been implicated in regulating the expression of *speB*. Although many factors have been shown to influence expression of *speB*, only a limited number have been shown to be directly involved [179, 218, 219]. *RopB*

(also known as *Rgg*) is the best-characterised transcriptional regulator of *speB* and is a positive regulator binding directly at the promoter region of the *speB* gene [179, 218, 219]. Previous authors showed that *speB* expression was strongly down-regulated in *RopB/Rgg* mutants [11]. Even though *RopB/Rgg* failed to meet our transcriptomic analysis criteria (i.e., p-values <0.01 and $\log_2$-fold change (LFC) values > 4 or < - 4), our data showed that *RopB/Rgg* is down-regulated in *emm*32.2 strain during early-mid-late exponential phase ($\log_2$-fold change (LFC) values < - 2).

One study showed that transcripts encoding virulence factors including SPN and SLO were more abundant in the RopB/Rgg-deficient strain, which was accompanied by increased levels of SPN and SLO activity during the exponential growth phase. These data suggested that RopB/Rgg represses the transcription of *nga* and *slo* genes during the exponential growth phase, and also controls the degradation of SpeB secreted proteins [220].

I also examined another regulator such as the catabolite control protein A (*CcpA*) which is a stand-alone transcription regulator that modulates genes involved in carbon utilisation. Previous studies have shown that *ccpA* is a positive regulator of SpeB expression [221, 222]. Our data did not show any difference in the expression of *ccpA*.

In summary, our data showed no variations in the gene expression of key transcription regulators except that of *Rgg/RopB*, which showed a lower expression in the *emm*32.2 isolates. Our data hence suggest that *Rgg/RopB* may be a key factor player in the differential expression in important virulence factors such as nga, *slo* and, *speB*,

particularly during the exponential growth phase. We thus propose that *Rgg/RopB* is one key factor involved in the change in gene expression of *slo, nga* and *speB*.

My study is the first to characterise the *in vitro* transcriptome of the *emm*32.2 GAS strain and gives new insights into *emm*32.2-specific gene expression. Differential regulation of some well-known virulence factors occurs during the exponential growth phase. I propose this set of genes contributes to *emm*32.2 invasiveness. Alternatively, my results comparing *emm*32.2 isolate and *emm*1 isolate (invasive) suggests that there may be multiple routes to pathogenesis in GAS. This might be one reason why GAS is such a successful bacteria in its relationship with a human host. In support of this, I have also shown that some well-known virulence factors were up-regulated compared to invasive isolate (*emm*1.0: 101910) in the *emm*32.2 and secreted in larger quantities and could contribute to adaptation during infection. Regardless, in view of our findings that a number of well-studied virulence factor were up-regulated in *emm*32.2 compare to *emm*1.0 and *emm*89.0 suggest that these set of genes largely contributes to *emm*32.2 invasiveness, I believe further studies to elucidate the full mechanism of invasive pathogenesis are necessary such as generating gene expression deficient-strains and investigating their phenotype both *in vitro* and *in vivo*.

Chapter 5 Comparative Proteomic Analysis of Group A Streptococcus Secretome

5.1 Introduction

Proteomics is the large-scale study of proteomes, defined as the entire set of proteins that a given cell or organism expresses at a specific time or under certain conditions. Proteomics analysis is concerned with gaining insights into complex cellular mechanisms through qualitative and quantitative comparisons of proteomes [223]. Not only does proteomics deliver quantitative information, which helps to assess protein levels and expression, but it also provides qualitative information, which helps to verify the occurrence of protein isoforms and post-translation modification. Remarkably, protein abundance and functions are often influenced by the developmental and physiological status of cells. Hence, proteomics represents a very valuable tool to better understand the molecular processes underlying the differences between such states.

Due to the advance gained in GAS genome sequencing, and the progress made in mass spectrometry (MS) analysis, characterisation of the GAS proteome has been improved. Although GAS has four sub-proteomes, including cytoplasmic proteins, cell

membrane proteins, cell wall-associated proteins, and culture supernatant proteins (CSPs), it is known that key GAS virulence factors are secreted proteins [224].

In our study, we aimed to investigate the absence/presence and variations in protein abundance in accordance with the metabolic changes occurring during GAS growth in liquid broth. This was deemed an appropriate approach as depletion of nutrients and decreasing pH conditions mimic the environment to which the pathogen is exposed during infection [192].

Although my previous chapter revealed important differences in gene expression between the *emm*32.2, *emm*1.0 and *emm*89.0 isolates during the exponential phases of growth (early-mid-late), it was essential to evaluate whether these changes were reflected at the protein level, which are the effectors of cellular functions. Thus, the goal of this chapter was to identify and characterise the set of proteins found in the culture supernatants of GAS isolates grown *in vitro* ,and identify possible proteins contributing to virulence of *emm*32.2 isolate. Here I compared the secretome of the *emm*32.2 (112327 invasive outbreak isolates) to *emm*1.0 (101910 invasive) and *emm*89.0 (127746 non-invasive) isolates during the exponential phase of growth (early-mid –late). The information obtained from the proteomic analysis will help us to further understand the secretome features of *emm*32.2 isolate which might explain the unique clinical and epidemiological attributes of this strain, and aid future vaccine development.

5.2 Results

5.2.1 Proteomic Analysis of GAS

As extracellular GAS proteins play critical roles in the pathogenesis of GAS [198], I studied proteins present in the culture supernatant of *emm*32.2 , *emm*1.0 and *emm*89.0 isolate. I focused on three key growth phases: early exponential phase (EEP), mid-exponential phase (MEP), and late exponential phase (LEP). Briefly, culture supernatant proteins (CSPs) from each growth phase were isolated as described in the Chapter 2 Materials and Methods. Each isolate was grown at 37ºC and samples were collected after 4h, 6h, and 10h, corresponding to early, mid and late exponential phase, respectively. During the protein identification process of the mass spectrometry analysis, in order to eliminate biologically irrelevant variations, only proteins with 2 or more uniquely identified peptides were included in the analysis [225-227]. In total, 284 proteins were detected.

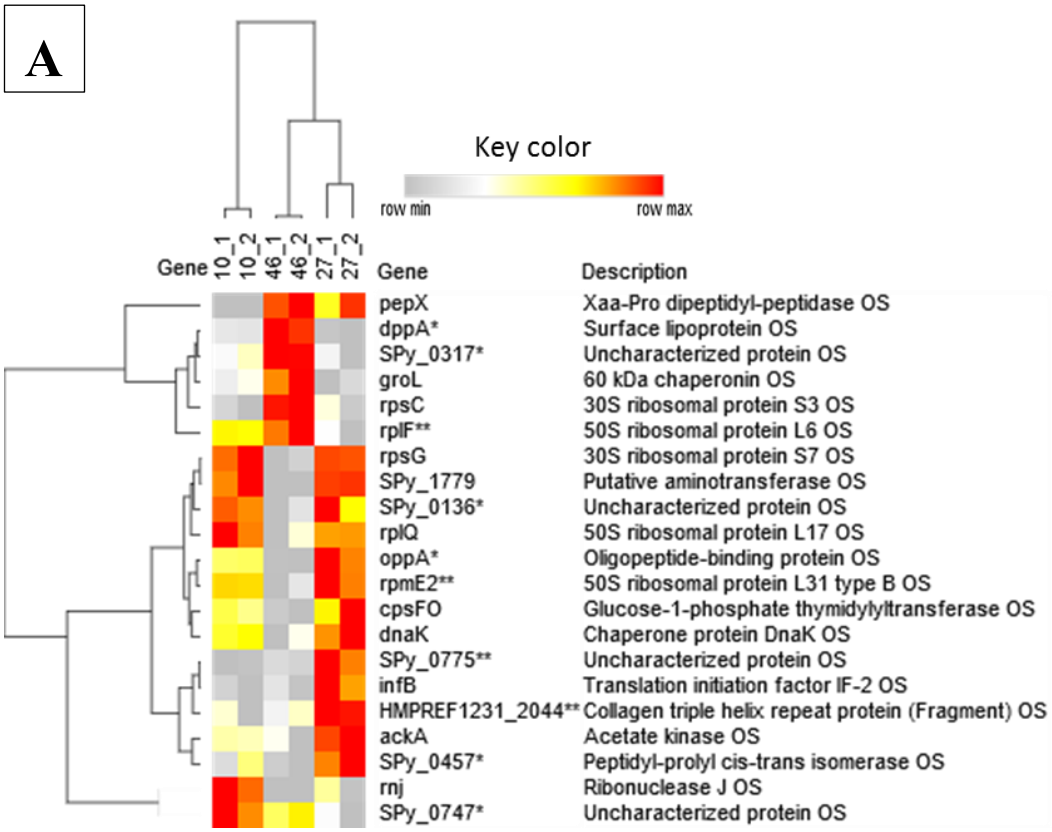
5.2.2 Differential Secretomics of GAS

To identify differences between the secretomes of *emm*32.2, *emm*1.0 and *emm*89.0 isolates, hierarchical clustering of the detected proteins was performed, which involves creating clusters based on the degree of similarity of the proteins abundance levels. Several extracellular proteins were differentially expressed according to the exponential growth phase (early, mid, late). These proteins displayed a significant difference ($p<0.05$) between at least one of the time points.

The Figure 5-1 shows the heatmap for the proteins expressed by *emm*32.2, *emm*1.0 and *emm*89.0 isolates during the early (Figure 5-1A), mid (Figure 5-1B) and late exponential growth phases. Here, the heatmap was used to provide with an overview of the relative abundance of the identified proteins, and represent their expression as groups sharing similar patterns according to each one of the growth phases. Thus, the data exposed a comparative overview of the secretomes of the *emm*32.2, *emm*1.0 and *emm*89.0 strains during the early, mid and late exponential growth.

The data showed that the *emm*32.2, *emm*1.0 and *emm*89.0 are separated based on their protein expression. The *emm*32.2 isolate is more similar to *emm*89.0 isolate at both the early and mid-exponential phase. On the other hand, the data showed a larger dissimilarity between the *emm*32.2, *emm*1.0 and *emm*89.0 isolates during the late time point.

Relative abundance of 21 (Figure 5-1A) and 10 (Figure 5-1B) proteins showed a significant difference between GAS isolates ($p < 0.05$) during the early and mid-exponential growth phase, respectively. However, this number of significant proteins increased to 91 proteins during late exponential phase. 10 out 21, 3 out 10, 40 out of 91 proteins detected during early, mid, and late exponential growth phase, respectively, are predicted to be the cultural secreted proteins.


A
Key color
row min
row max
Gene
10_1
10_2
46_1
46_2
27_1
27_2
Gene
Description
pepX
Xaa-Pro dipeptidyl-peptidase OS
dppA*
Surface lipoprotein OS
SPy_0317*
Uncharacterized protein OS
groL
60 kDa chaperonin OS
rpsC
30S ribosomal protein S3 OS
rplF**
50S ribosomal protein L6 OS
rpsG
30S ribosomal protein S7 OS
SPy_1779
Putative aminotransferase OS
SPy_0136*
Uncharacterized protein OS
rplQ
50S ribosomal protein L17 OS
oppA*
Oligopeptide-binding protein OS
rpmE2**
50S ribosomal protein L31 type B OS
cpsFO
Glucose-1-phosphate thymidylyltransferase OS
dnaK
Chaperone protein DnaK OS
SPy_0775**
Uncharacterized protein OS
infB
Translation initiation factor IF-2 OS
HMPREF1231_2044**
Collagen triple helix repeat protein (Fragment) OS
ackA
Acetate kinase OS
SPy_0457*
Peptidyl-prolyl cis-trans isomerase OS
rnj
Ribonuclease J OS
SPy_0747*
Uncharacterized protein OS

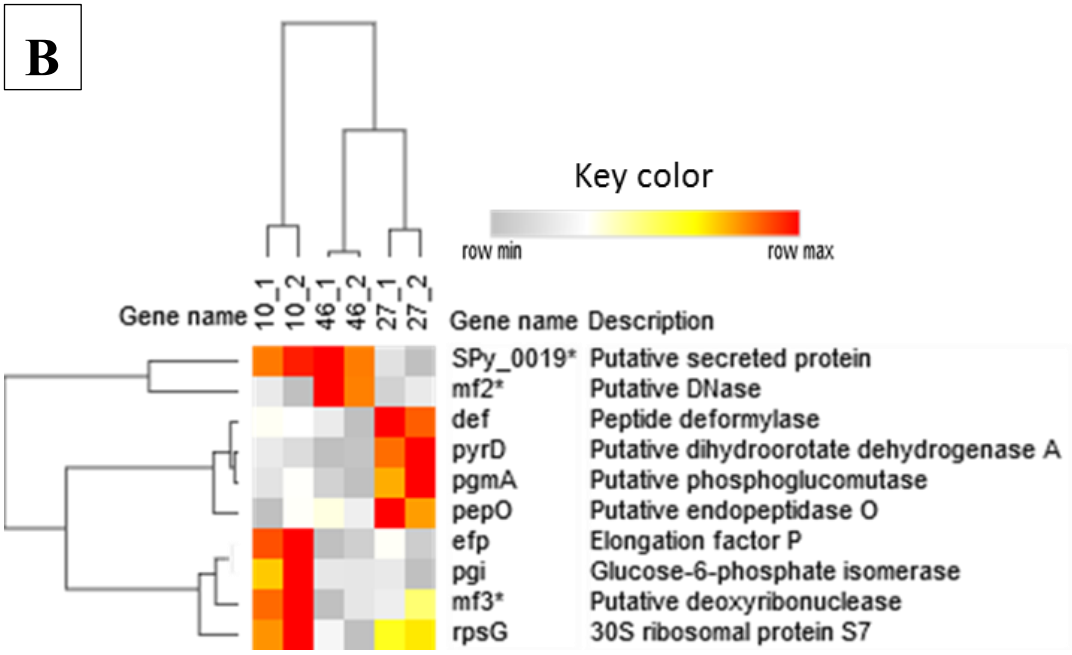


Figure 5-1 Heatmap representation of secreted Group A Streptococcus proteins

Protein groups presented in the figure were selected based on their normalized peptide intensities as described in Chapter 2 (Methods) during A) early and B) mid exponential phase. Each isolate was analysed in two independent experiments. Rows represent genes and columns represent individual GAS isolates. Each cell in the heatmap represents relative expression values. The colours represent the relative abundance of the expressed proteins, i.e., a protein with the highest intensity (relative to the other samples) is coloured red whereas a protein with lowest Intensity is grey. Sample Identifiers: biological replicates for strain emm32.2 (112327) were coined 27_1 and 27_2, 10_1 and 10_2 for strain *emm*1.0 (101910), and 46_1 and 46_2 for *emm*.89.0 (127746). *Proteins predicted to be secreted by signal P (http://www.cbs.dtu.dk/services/SignalP) or phobius (http://phobius.sbc.su.se/). ** Proteins predicted to be secreted by secretome P (http://www.cbs.dtu.dk/services/SecretomeP/).

5.2.3 In-Depth analysis of GAS secretome during late exponential phase

My data showed the differential expression of a group of 10 to 91 important proteins during the exponential growth phase. To increase the confidence on the detection of differentially expressed proteins, we restricted the original list of differentially expressed proteins to those that showed both a significant difference in corrected p-value (q) ≤ 0.05 and fold change values ≥ 2. This narrowed down the list of proteins to 35, which were either up-regulated or down-regulated during the late exponential phase of growth (Figure 5-2). On the basis of their q values, the proteins identified during the early and mid-exponential growth stages were found excluded from further analysis.

In Table 5-1 are lists the names of the 35 significant proteins that were identified as differentially expressed during the late exponential phase. Therefore, the proteins presented in Table 5-1 correspond to those that showed the highest relative abundance in the supernatant of *emm*32.2, *emm*1.0 and *emm*89.0 isolates, which may contribute to differential clinical phenotype of GAS. These proteins were divided into 3 groups; up-regulated in *emm*32.2 (Red in Table 5-1), up-regulated in *emm*1.0 (Green in Table 5-1), up-regulated in *emm*89.0 (Blue in Table 5-1). In total, 17 proteins presented higher expression in the *emm*32.2 strain compared to the *emm*1.0 and *emm*89.0 isolates, and among these 17 proteins, 6 of them are well-known virulence factors such as the C5a peptidase (ScpA) and Streptolysin O (SLO). From 17 proteins, 11 are predicted to be the cultural secreted proteins (Table 5-1). Protein G-related alpha 2M-

binding protein (GRAB), which is a well-known virulence factor, and spy2176, corresponding to an uncharacterised protein, were both detected in high abundance in *emm*1.0. Both spy2176 and GRAB are predicted to be the cultural secreted proteins. Mitogenic factor (MF2), which is a major DNase, was predicted to be cultural secreted proteins in *emm*89.0. Overall, key virulence factors such as the ScpA and SLO were found to be in high abundance in the supernatant of the *emm*32.2 isolate.

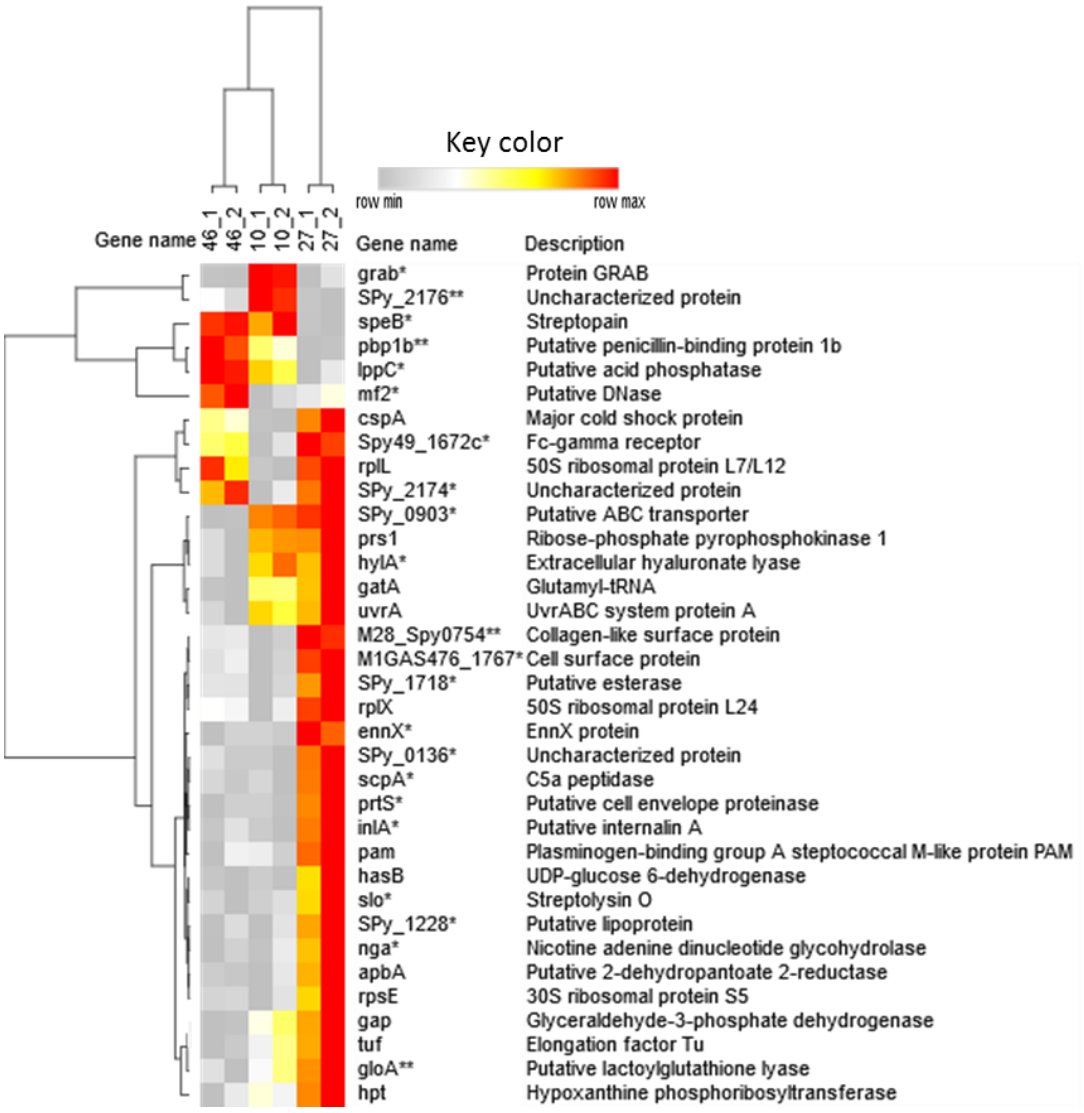


Figure 5-2 Heatmap representation of secreted Group A Streptococcus proteins

Protein groups displayed in the figure were selected based on their normalized peptide intensities as described in Chapter 2 (Methods) during late exponential phase. Each isolate was analysed in two independent experiments. Rows represent genes and columns represent the individual GAS isolates. Each cell in the heatmap represents relative expression values. Sample Identifiers: biological replicates for strain emm32.2 (112327) were coined 27_1 and 27_2, 10_1 and 10_2 for strain *emm*1.0 (101910), and 46_1 and 46_2 for *emm*.89.0 (127746).*Proteins predicted to be secreted by signal P (http://www.cbs.dtu.dk/services/SignalP) or phobius (http://phobius.sbc.su.se/). ** Proteins predicted to be secreted by secretome P (http://www.cbs.dtu.dk/services/SecretomeP/).

Table 5-1 Significant Group A Streptococcus proteins during late exponential phase.

			Average Log 2 fold change[II]		
Protein ID	**Gene name**	**Description [I]**	***emm*89.0 (127746)**	***emm*1.0 (101910)**	***emm*32.2 (112327)**
Q48TT9	*M28_Spy0754***	Collagen-likesurface protein	20.235	18.985	25.855
J7MBD1	*M1GAS476_1767**	Cell surface protein	15.37	13.94	22.21
Q99YG3	*SPy_1718**	Putative esterase	18.195	17.48	22.18
Q9A1W3	*rplX*	50S ribosomal protein L24	23.64	23.055	25.68
A0A0H3C1S6	*ennX**	EnnX protein	17.89	18.085	24.15
Q9A1R6	*spy_0136**	Uncharacterised protein	16.265	15.835	20.685
U2U073	*pam*	Plasminogen-binding group A streptococcal	17.575	17.855	24.89
Q9A180	*prtS**	Putative cell envelope proteinase	18.625	18.67	25.055
Q99Z76	*inlA**	Putative internalin A	17.275	16.825	22.415
P58099	*scpA**	C5a peptidase	22.2	22.08	26.97
P0C0F5	*hasB*	UDP-glucose 6-dehydrogenase	18.495	18.645	24.355
P0C0I3	*slo**	Streptolysin O	23.94	24.275	27.965
Q99ZH4	*SPy_1228**	Putative lipoprotein	20.955	21.02	24.37
Q7DAN2	*nga**	Nicotine adenine dinucleotide glycohydrolase	21.28	21.7	25.755
P65666	*apbA*	Putative 2-dehydropantoate 2-reductase	19.86	20.045	22.845
P66583	*rpsE*	30S ribosomal protein S5	21.94	21.86	24.48
M4YW00	*Hpt*	Hypoxanthine phosphoribosyltransferase	15.895	16.99	19.95
P65243	*prs1*	Ribose-phosphate pyrophosphokinase 1	10.78	17.465	18.655
P66063	*rplL*	50S ribosomal protein L7/L12	23.355	21.955	23.605
Q9A074	*SPy_0903**	Putative ABC transporter	17.955	20.36	20.7
Q99Y38	*lppC**	Putative acid phosphatase	20.695	19.55	17.745
Q99YC0	*gatA*	Glutamyl-tRNA	16.14	18.435	20.095
P0C0G7	*Gap*	Glyceraldehyde-3-phosphate dehydrogenase	26.12	28.61	31.145
P0C0F1	*cspA*	Major cold shock protein	21.945	20.61	23.44
A0A0H3BZ39	*Spy49_1672c**	Fc-gamma receptor	16.33	12.59	19.685
Q99Y84	*uvrA*	UvrABC system protein A	13.425	16.9	18.26
P0C0J1	*SpeB**	Streptopain	29.165	28.81	24.14
P69952	*Tuf*	Elongation factor Tu	26.305	28.585	31.895
Q99ZX4	*HylA**	Extracellular hyaluronate lyase	16.885	20.12	20.525
Q99XJ7	*SPy_2174**	Uncharacterised protein	21.78	17.96	22.155
Q9A121	*gloA***	Putative lactoylglutathione lyase	19.225	20.445	22.41
Q9A1U2	*pbp1b***	Putative penicillin-binding protein 1b	19.69	17.88	16.28
Q7DAL7	*grab**	Protein GRAB	14.79	21.66	15.31
Q99XJ6	*SPy_2176***	Uncharacterised protein	19.3	21.635	18.62
Q9A0M1	*mf2**	Putative DNase	26.3	21.315	22.6

I. Gene descriptions derived from the UniProt (http://www.uniprot.org/).
II. Average Log 2 fold change is calculated from two biological replicates.
*Proteins predicted to be secreted by signal P (http://www.cbs.dtu.dk/services/SignalP/) or phobius (http://phobius.sbc.su.se/).
** Proteins predicted to be secreted by secretome P (http://www.cbs.dtu.dk/services/SecretomeP/).

5.2.4 Principal-component analysis (PCA)

In order to gain further evidence on the differential expression between *emm*32.2, *emm*1.0 and *emm*89.0 secretome proteins, and to determine whether there is a concealed pattern in the proteomic data during late phase growth, we carried out a principal-component analysis (PCA) of our data (Figure 5-3***).*** Supportive of our heatmap observations (Figure 5-2), the results clearly showed that the 3 *emm* strains form distinct clusters (Figure 5-3***),*** and most interestingly, these differences are mainly due to the differential expression of the 35 proteins identified above.

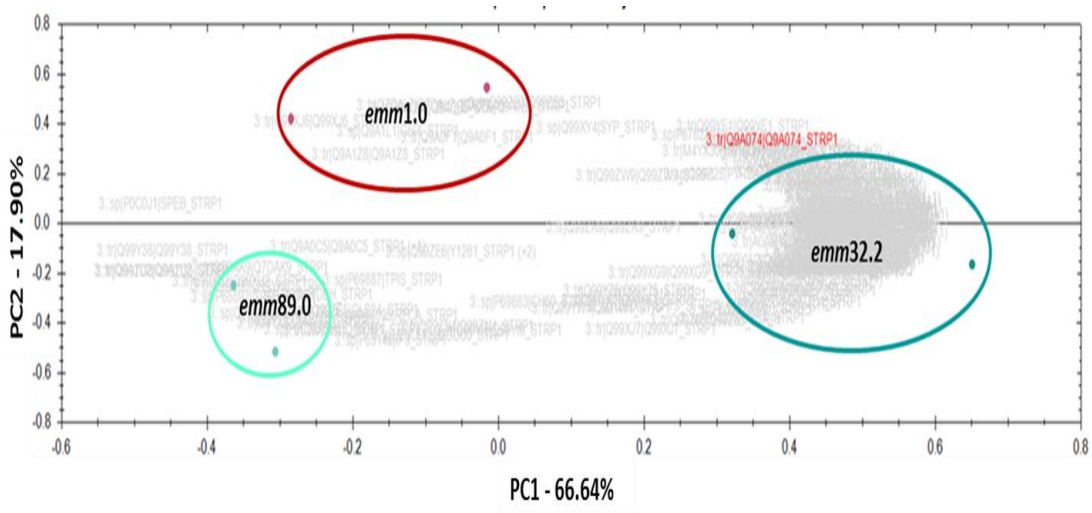
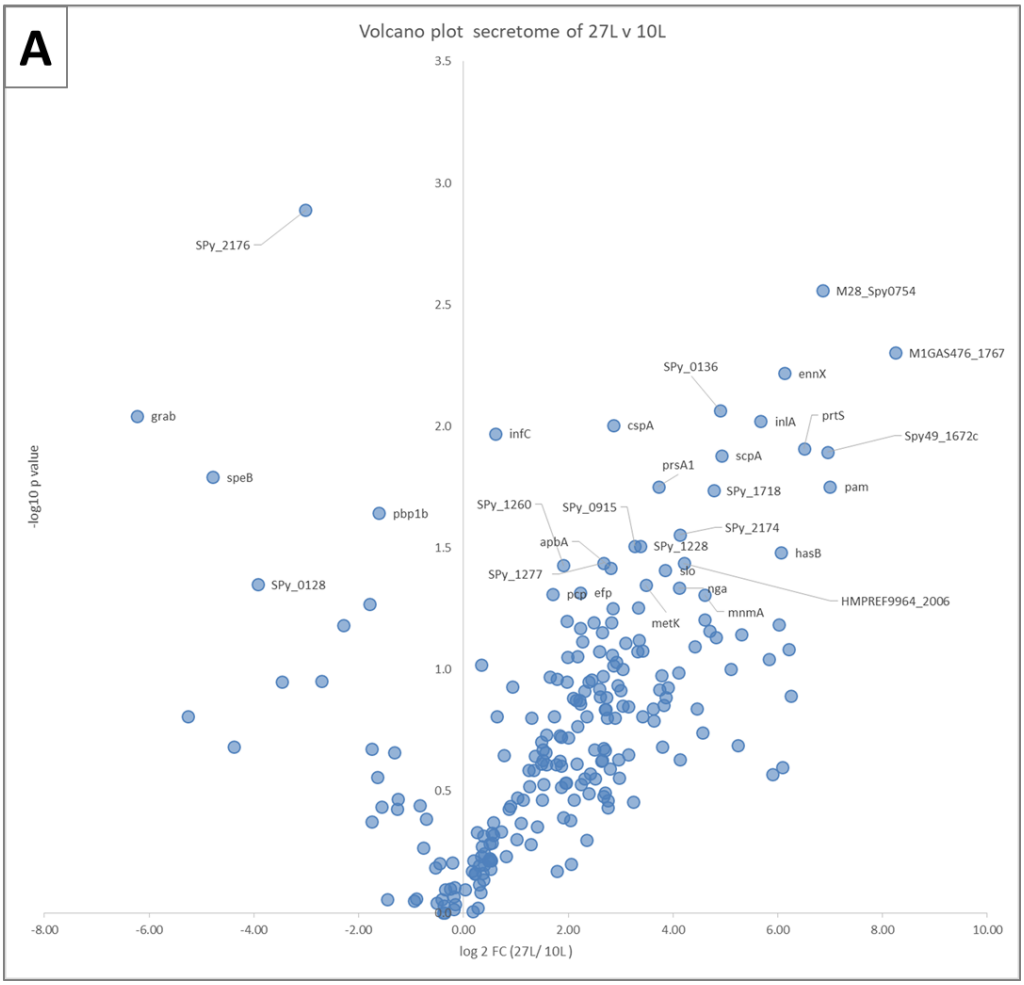


Figure 5-3 Principal component analysis of emm32.2, emm1.0 and emm89.0 isolates.

Proteomics data are showing secreted proteins with significant differential abundance between the samples. Two biological replicates were performed for each isolate. X- and Y-axis show principal component 1 and principal component 2 which explain 66.64% and 17.90% of the total variation in the dataset, respectively.

5.2.5 Pairwise comparison of GAS secretome

To gain a more detailed understanding of each comparison, we used volcano plots, which summarise fold change (x-axis) and statistical differences (y-axis = p-value) of the analysed data. Volcano plots in Figure 5-4A and Figure 5-4B provides an overview of the differences observed between *emm*32.2 and *emm*1.0 (Figure 5-4A), and between *emm*32.2 and *emm*89.0 (Figure 5-4B) during the late exponential phase of growth. The volcano plot represents a useful way to display significant differences in relative to protein abundance. These data, together with biological function about each protein, is beneficial for identifying key proteins for further investigation. Overall, pairwise comparisons revealed a variety of proteins that showed a significant difference between the *emm*32.2, *emm*1.0 and *emm*89.0 during the late exponential phase.


A
Volcano plot secretome of 27L v 10L
-log10 p value
log 2 FC (27L/ 10L)
3.5
3.0
2.5
2.0
1.5
1.0
0.5
-8.00
-6.00
-4.00
-2.00
0.00
2.00
4.00
6.00
8.00
10.00
SPy_2176
M28_Spy0754
M1GAS476_1767
SPy_0136
ennX
grab
inlA
prtS
infC
cspA
Spy49_1672c
scpA
speB
prsA1
SPy_1718
pam
pbp1b
SPy_1260
SPy_0915
SPy_2174
apbA
SPy_1228
hasB
SPy_1277
SPy_0128
HMPREF9964_2006
mnmA
metK

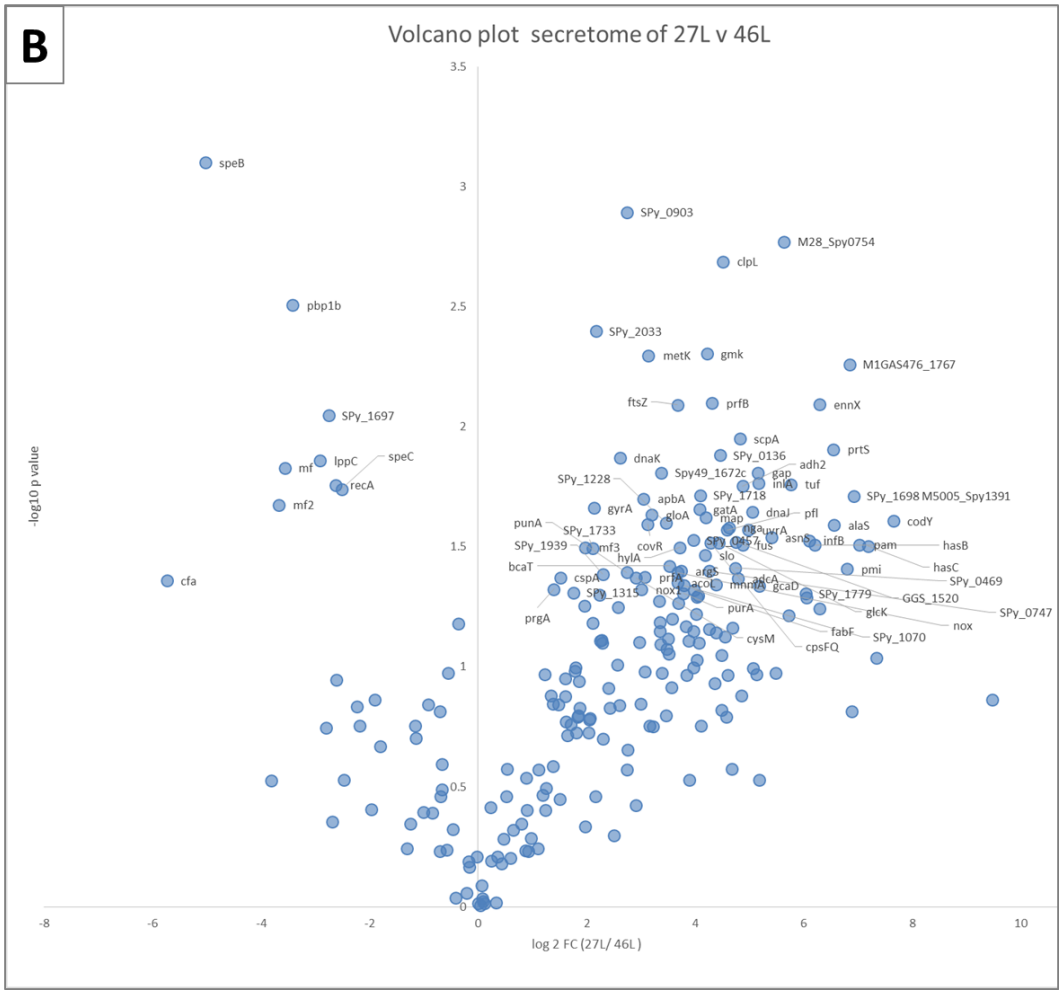


Figure 5-4 Volcano plots of the emm32.2 vs. emm1.0 and emm89.0 isolates' secretome data during the late exponential phase of growth.

Volcano plots showing log2 fold changes and statistically significant differences in the proteins' relative abundance of *emm*32.2 compared to A) *emm*1.0 and B) *emm*89.0. The negative $\log_{10}$ p-values (y-axis) are plotted against $\log_2$ fold changes (x-axis). Each blue dot represents the difference in the relative abundance of a given protein. Proteins meeting a p-value below 0.05 are labelled with their corresponding names in the figure.

5.2.6 Functional characterisation of identified proteins

To highlight the cellular and molecular functions of the identified differentially expressed gene products, we performed functional annotations using DAVID for all the Gene Ontology (GO) categories and functional enrichment based on the Kyoto Encyclopedia of Genes and Genomes (KEGG). Analysis of significantly up-regulated and down-regulated proteins during the late exponential phase of growth revealed significantly enriched genes belonging to two groups: extracellular region, which includes the C5a peptidase (scpA), Hyaluronate lyase (hyIA), GAS cell envelope protease (PrtS), Streptococcal pyrogenic erythrogenic toxin B (SpeB), and Streptolysin O (SLO). Interestingly, ScpA , SpeB, SLO were documented as essential GAS virulence factors. This is in line with the literature that describes a variety of extracellular products, many of which are secreted proteins, often considered to be play an important role in streptococcal pathogenesis [11] ($p < 0.05$, Table 5-2).

Table 5-2 GO and KEGG enrichment analysis of significantly expressed Group A Streptococcus proteins.

ID	Gene Ontology terms	Genes	p-value
GO:0005576	extracellular region	*scpA ,hylA, prtS, speB, slo*	7.7E-4
GO:0009405	Pathogenesis	*scpA , speB, slo*	2.4E-3

5.2.7 Proteins associated with GAS virulence factors

Our proteomic analysis highlighted a subset of well-known virulence factors that showed high or low expression patterns relative to the other compared *emm* isolates (Table 5-1). A set of 6 proteins with documented functions in pathogenesis, including UDP-glucose 6 Dehydrogenase, collagen-like surface protein, C5a peptidase, Streptolysin O (SLO), streptococcal NAD+ glycohydrolase (SPN), and Putative cell envelope proteinase, were significantly more abundant in *emm*3.2 than in the *emm*1.0 and *emm*89.0 isolates. On the other hand, the protein-G related-α2-macroglobulin-binding protein and uncharacterised proteins were significantly less abundant in *emm*32.2 isolate.

5.2.8 Other notable proteins

Our comparative secretome analysis highlighted the importance of well-known proteins that have previously been reported as promising vaccine candidates, such as SLO, ScpA, putative esterase (Spy1718) and putative lipoprotein (spy1228). These proteins were significantly more abundant in *emm*32.2 than in *emm*1.0 and *emm*89.0. Overall, my data identified very important protein antigens, which have not reached human clinical trials as vaccine candidates.

5.2.9 Correlation between transcriptome and proteome

While both the transcriptomic and proteomic analysis revealed differential expression of a number of virulence factors, particularly those involved in carbohydrate transport and metabolism, the overlap between gene expression and secreted proteins were small during late exponential phase, with only 8 out 61 (13%) genes/products were shared by both datasets. Of the 61 genes found to be either up- or down-regulated in the transcriptomic analysis (Table 4-2 and Table 4-3) only 8 genes (*hasBC, slo,nga, SpeB*, Putative esterase, *scpA*, Collagen-like surface protein (*sclB*) of them had a corresponding protein in the secretome. Although *hasC* gene did not meet the criteria of proteomic analysis (corrected p-value (q) $q \leq 0.05$), all 8 proteins showed a trend in their expression that matched those of the corresponding transcript.

5.3 Discussion

In the previous chapter, we showed the differences in gene expression levels of *emm*32.2 GAS isolates (112327) compare to *emm*1.0 (101910) and *emm*89.0 (122746) during exponential phase (early-mid-late.) However, in order to identify the differentially expressed proteins that maybe relevant to the invasiveness of *emm*32.2 isolate. The secretome of the *emm*32.2 isolate was compared to those of *emm*1.0 and *emm*89.0 isolates during the exponential phases of growth, i.e., early, mid and late exponential. I presented proteomic data that give an overview of up- and down-regulated proteins during the exponential phase. Heatmaps data and volcano figures have exposed that, during the late time point, *emm*32.2 had significantly altered expression of a wide range of proteins associated with GAS virulence.

In this project, we conducted an initial analysis of proteins found in relative abundance in the culture supernatants of 3 GAS strains grown *in vitro* at 37°C. The study focused on *emm*32.2 (112723 outbreak invasive isolate), *emm*1.0 (101910 invasive), and *emm*89.0 (127746 non-invasive). This analysis included any proteins with 2 or more uniquely identified peptide. Accordingly, a set of 284 proteins was identified proteins during each time point in the exponential growth phase.

To recognise the significantly differentially expressed of these proteins, only proteins with a p-value = <0.05 were included in the analysis. Consequently, less than 25 significant proteins were identified during early and mid-exponential phase but significant changes in the relative abundance protein was seen during the late time point. This data suggests that the late exponential phase may correspond to the phase

of growth when virulence associated proteins are more readily detected- Hence further understanding of the changes occurring during this phase may help expose some crucial aspects of *emm*32.2 virulence and metabolism. This result is consistent with the transcriptomic data presented in Chapter 4.

Our data showed a list of significant proteins during growth phase (early-mid-late exponential phase). To enhance the confidence of differentially expressed prediction, we decided to use both the corrected p-value (q) < 0.05 and relative fold change ≥ 2. This lead to exclude both early and mid-point from further analysis and limited the list of 91 proteins to 35 proteins during late exponential growth phase, a total of n = 14 out of the 35 proteins was identified in previous secretome studies on GAS [228-230].

To learn about variation in proteins expression according to growth phase, I performed gene functional classification based on all GO categories and KEGG. This analysis allowed us to identify 2 significant groups. The first group is connected to extracellular region and it has 5 proteins and the second group which has 3 proteins is associated with pathogenesis. The presence of extracellular region as a significant group in the GO enrichment analysis suggested that the method that is used in this study is appropriate. Moreover, the presence of a pathogenesis group in our data is in line with recent literature reviews [224], and these results support the idea that crucial extracellular GAS proteins have been known to have a role in the GAS pathogenesis.

Among 35 changes in the abundance of CSPs between *emm*32.2 isolate and *emm*1.0 and *emm*89.0, set of 8 proteins have been shown to be associated with virulence factors. This list includes ScpA, HasB, SLO, SPN, M28 spy0754 and Prts, which were

more abundant in the *emm*32.2 isolate, and GRAB, and SpeB, which were less produced by *emm*32.2 isolate.

Proteins involved in invasive GAS disease which they were high in *emm*32.2 isolate including *hasB* gene which encodes UDP-glucose dehydrogenase and is one of the capsule genes [74, 75]. The bacterial capsule is an essential virulence factor for GAS which prevents opsonised phagocytosis [73, 82]. Not only does GAS capsule prevent opsonised phagocytosis [73, 82] but also it obstructs the antibody's access to surface epitopes [81], and helps to inhibit complement deposition [73]. Our data show high expression of *hasB* gene product consistent with other research about the production of hyaluronic acid taking place at the exponential phase of growth [167, 168].

GAS elaborate a variety of potent cytotoxins, another vital mechanism for innate immune resistance. SLO is a multifunctional virulence factor, which contributes to GAS pathogenesis through interaction, and cooperates with other virulence factors such as nga. These virulence factors are fundamental virulence factors of GAS which promote pathogenesis through mechanisms such as endosome evasion, resistance to phagocytic killing and promotion of intracellular bacterial survival [111, 194, 195] as well as inhibition of neutrophil chemotaxis [196]. In previous *in vivo* studies, authors have shown that both SLO and nga are vital determinants of GAS pathogenesis [181]. Our data show high expression of *nga* gene and *slo* gene product in *emm*32.2 compared to *emm*1.0 and *emm*89.0.These results are in line with phenotypic data presented in Chapter 3 (Figure 3-5) showing that SLO production of *emm*32.2 strain was significantly higher than *emm*1.0.

Our data show high expression of both putative cell envelope protease (PrtS) and C5a peptidase in *emm*32.2 compare to *emm*1.0 and *emm*89.0. PrtS also is known as SpyCEP, is a cell wall-associated peptidase that can inactivate a number of pro-inflammatory chemokines [231]. PrtS activity increases GAS resistance to neutrophil-mediated killing [232]. The ScpA is usually cell-wall anchored and expressed on the surface of all serotypes of GAS where it specifically destroys C5a [233, 234]. In *in vivo* studies, C5a has been shown to induce a potent serum antibody response that enhances bacterial clearance from the nasopharynx [200]. GAS has evolved mechanisms to interfere with two of the most potent molecules promoting neutrophil recruitment, the CXC chemokine interleukin-8 (IL-8) and the complement-derived anaphylatoxin C5a. In this way, the kinetics of the innate immune response to GAS infection is delayed, favouring *emm*32.2 isolate survival.

Streptococcal collagen-like Scl2 is known as SclB, which encode collagen-like proteins. It is an adhesion protein belonging to the collagen superfamily [201, 202]. Even though the specific function of this gene in pathogenesis is less understood, previous reports suggest that Scl2 interacts with T cells and cause hyper-activation of the immune response [203].

Furthermore, 2 important proteins SepB and GRAB found to be in high abundance in the supernatant of the *emm*1 isolate but less abundance in the supernatant of *emm*32.2. SpeB is anchorless adhesin, with enzymatic function as cysteine protease. This protein has multiple functions including adhesion and has the ability to change the phenotype of the producing strain, not only by indirectly affecting transcription ,but also through

various post-translational events, such as releasing proteins from bacterial surfaces or degrading them to terminate their activity [95, 235]. Previous work has indicated a crucial role for the active SpeB which can degrade many of cultural supernatant GAS proteins such as numerous well-studied virulence factors [97]. In line with our proteomic data, phenotypic analysis data from chapter 3 (Table 3-3) which showed that streptococcal cysteine proteinase activity of *emm*32.2 strain (112327) was undetectable. My data suggest the decrease in the SepB gene could contribute to invasiveness of the *emm*32.2 isolate

Protection from degradation by either host and bacterial proteases is accomplished by GRAB which is a surface protein and key protease inhibitor present in human plasma [205, 236]. The GRAB gene is present in most GAS isolates and is well conserved. The SpeB can also be retained on the bacterial surface by being trapped by alpha 2-M-binding protein bound to the GRAB protein [201, 202]. My data suggest the increase in the GRAB gene could be the reason to protect *emm*1.0 from degradation and contribute to invasiveness of the *emm*1.0 isolate.

In support of our transcriptomic data, our comparative secretomics data provided crucial information on proteins secreted by the *emm*32.2 isolate. A group of 7 proteins, namely ScpA, UDP-glucose 6 Dehydrogenase, SLO, SPN, M28 spy0754, GRAB, and SepB were found in the supernatant of proteome samples. The subset of virulence factors found in the proteome samples showed expression patterns similar to those described in the transcriptomic analysis.

Furthermore, many of the proteins presented significantly up- or down-regulated expression when comparing *emm*32.2 and *emm*1.0 and *emm*89.0 isolates, supporting the observation that *emm*32.2 has a unique biological phenotype. Interestingly, several of the differentially expressed proteins were shown to be important GAS virulence factors. My data suggest the increase in these virulence factors apparently promoted invasiveness of the *emm*32.2 isolate.

A number of non-M-protein antigens including SLO, ScpA, putative esterase and lipoprotein proteins were found to be more abundant in the *emm*32.2 compared to *emm*1.0 and *emm*89.0 strains. These proteins have also been characterised as GAS vaccine candidates. Notably, two of these proteins, SLO and ScpA, have already been considered as vaccine candidates [237].

Likewise, putative esterase (Spy1718) is a non-M-protein antigen that was previously reported in GAS culture supernatant [238]. This is consistent with our data. Putative esterase is an extracellular product that appears to have a role in virulence and pathogenesis, being essential for invasive infections and systemic dissemination. Although the role of putative esterase has yet to be fully determined, *In vivo* studies suggest that putative esterase appears to have a role in Invasive skin Infection and systemic dissemination [239].

In addition, putative extracellular lipoproteins GAS (spy1228) is a putatively conserved lipoprotein [240]. Even though, the function of the spy1228 has not been explained yet, previous work has shown that the spy1228 provokes an antibody response in both mice and humans [241]. The Spy1228 is protective GAS vaccine

antigens recognised by both B and T cells in human adults and children. *In vivo* study has revealed that protection against GAS infection was presented by the spy 1228. [242]. In line with current literature, my result showed that the spy1228 is an extracellular secreted product [243]. My data have led to the identification of verified proteins that may represent good candidates for vaccines.

There was little overlap between the transcriptome and proteome datasets under the experimental conditions used in this project. Important virulence factor genes such as *hasA*, *emm*1, were not found in the culture supernatant of GAS. Previous studies had also documented that the correlation between bacterial transcriptomes and proteomes may not be consistent [244]. One possible reason may be that the half-life of some proteins varies depending on the experimental conditions, e.g., growth media [245]. Post-transcriptional and post-translational changes can also significantly impact protein abundance [224]. For instance, SpeB is a key post-translational regulator of the extracellular proteome in GAS [224]. SpeB-mediated remodelling of the streptococcal surface proteome was shown significantly influence host-pathogen interactions and has been associated with pathogen dissemination [179]. The activation of SpeB results in results in proteolysis of both the enzymatic degradation of streptococcal proteins that mediate adherence and internalization and the degradation of host extracellular matrix proteins [246, 247].

Interestingly, our data showed less similarities between the transcriptome and proteome datasets in the late exponential phase. During this growth phase, cellular waste accumulation is expected to be high which may lead to inconsistencies between

the transcriptome and proteome. This is in line with previous work that showed that as growth progresses, the correlation between transcriptome and proteome decreases. The author suggested that production of the pervasive proteins and cellular waste accumulating during the late exponential phase may lead to differences between the transcriptome and the proteome [186].

In conclusion, our results indicated that several recognised virulence factor of GAS were more abundant in *emm*32.2 compare to *emm*1.0 and *emm*89.0 isolates including UDP-glucose 6 Dehydrogenase, collagen-like surface protein, C5a peptidase, Streptolysin O, SPN, and putative cell envelope proteinase which has been shown to play important roles in streptococcal pathogenesis. These data offer a prospective explanation for *emm*32.2 invasiveness. Interestingly, the relative abundance of 4 non-M-protein antigens including SLO, ScpA, putative esterase and putative lipoprotein protein were high in *emm*32.2. Three of these proteins are likely to be involved in GAS pathogenesis. As this was the first study to characterise the proteome of *emm*32.2 isolate *in vitro*, my work gives new insight into relative abundance proteins including non-M-protein antigens, which might assist in the development of potential GAS vaccine candidates.

Chapter 6 Conclusion

6.1.1 Overall summary

Group A Streptococcus (GAS), also known as *Streptococcus pyogenes*, is an important cause of human diseases ranging from minor conditions such as pharyngitis, to severe invasive infections such as necrotising fasciitis [17]. A half a million of GAS-associated deaths occur annually as a result of high incidence of diseases, including rheumatic heart disease (RHD) [20]. GAS colonises skin and mucosal membranes, including the nasopharynx epithelium and vagina. Over 100 distinct GAS serotypes have been described and over 130 M serotypes have been identified to date [9, 10].

We have studied invasive GAS involving *emm*32.2 strains derived from an outbreak in Merseyside between 2010 and 2012. This outbreak was unusual because it had a 29% mortality rate, was confined to adults, and had a propensity to occur in IV drug users, the homeless or alcoholics. We previously reported that, in total, 14 *emm*32.2 strains had been isolated and sequenced. Furthermore, the comparative genomic analysis revealed that 19 genes were unique to the outbreak isolates [146].

6.1.2 Phenotypic analysis of *emm*32.2 isolates compared to *emm*1.0, *emm*89.0, and *emm*6.0 isolates.

The overarching aim of the research presented in this thesis was to conduct *in vitro* examinations of the *emm*32.2 outbreak strain to identify whether there are differences in bacterial virulence that can be linked to the genomic data and may explain the epidemiological features of this *emm* subtype.

A range of virulence factors were assessed as possible causal factors of GAS infections, thereby aiding in the identification of the virulent phenotype of the outbreak strains (*emm*32.2). A large part of my research project focused on carrying out a comparative examination of invasive *emm*32.2 isolates (n=14), and invasive and mucosally isolates, i.e., *emm*6.0, *emm*89.0, and *emm*1.0, with respect to their capsule thickness, resistance to complement deposition, proteinase, NADase and SLO activity, and biofilm formation.

Our results demonstrated that phenotypic heterogeneity exists within the *emm*32.2 isolates as evidenced by variability in capsule thicknesses, proteolytic activity, SLO Activity, and biofilm formation. Interestingly, NADase activity was found to be below detection in all *emm*32.2 strains. More importantly, the *emm*32.2 isolates are less susceptible to complement deposition which may be translated to reduced phagocytosis in the host. These results suggest that this *emm* type is more resistant to host immune defence, which may thus explain some of its virulence properties. Finally, it is also worth acknowledging that *emm*32.2 contains a putative trypsin-resistant surface protein T6, which plays a crucial role in tissue adherence, biofilm

formation and extracellular translocation [248]. It could be one of the key factors contributing to the *emm*32.2 invasiveness.

6.1.3 Characterisation and comparative analysis of differentially-expressed genes of *emm*32.2, *emm*1.0 and *emm*89.0 GAS isolates in function of *in vitro* growth phases.

Standard research approaches were used for the purpose of illuminating phenotypic differences on distinct *emm* strains. Heterogeneity was identified in the outbreak strains, as well as across the *emm*32.2, *emm*1.0 and *emm*89.0 strains. The resultant observational phenotypic data encouraged me to characterise and comparatively examine the differentially-expressed genes of the *emm*32.2, *emm*1.0 and *emm*89.0 isolates with respect to different points of the exponential growth phase, as this may aid in identifying novel transcriptomic features specific to the outbreak isolates.

This study is the first to characterise the *in vitro* transcriptome of *emm*32.2 GAS strain. Our results give new insights into the gene expression of GAS strains. Our data showed that differential regulation of some well-known virulence factors including *hasABC*, *slo*, *emm*, and *nga*, are up-regulated in the *emm*32.2 type and secreted in larger quantities. Our results suggest these virulence factors are responsible for highly invasive clinical phenotypes as those caused by *emm*32.2 and could contribute to adaptation during infection.

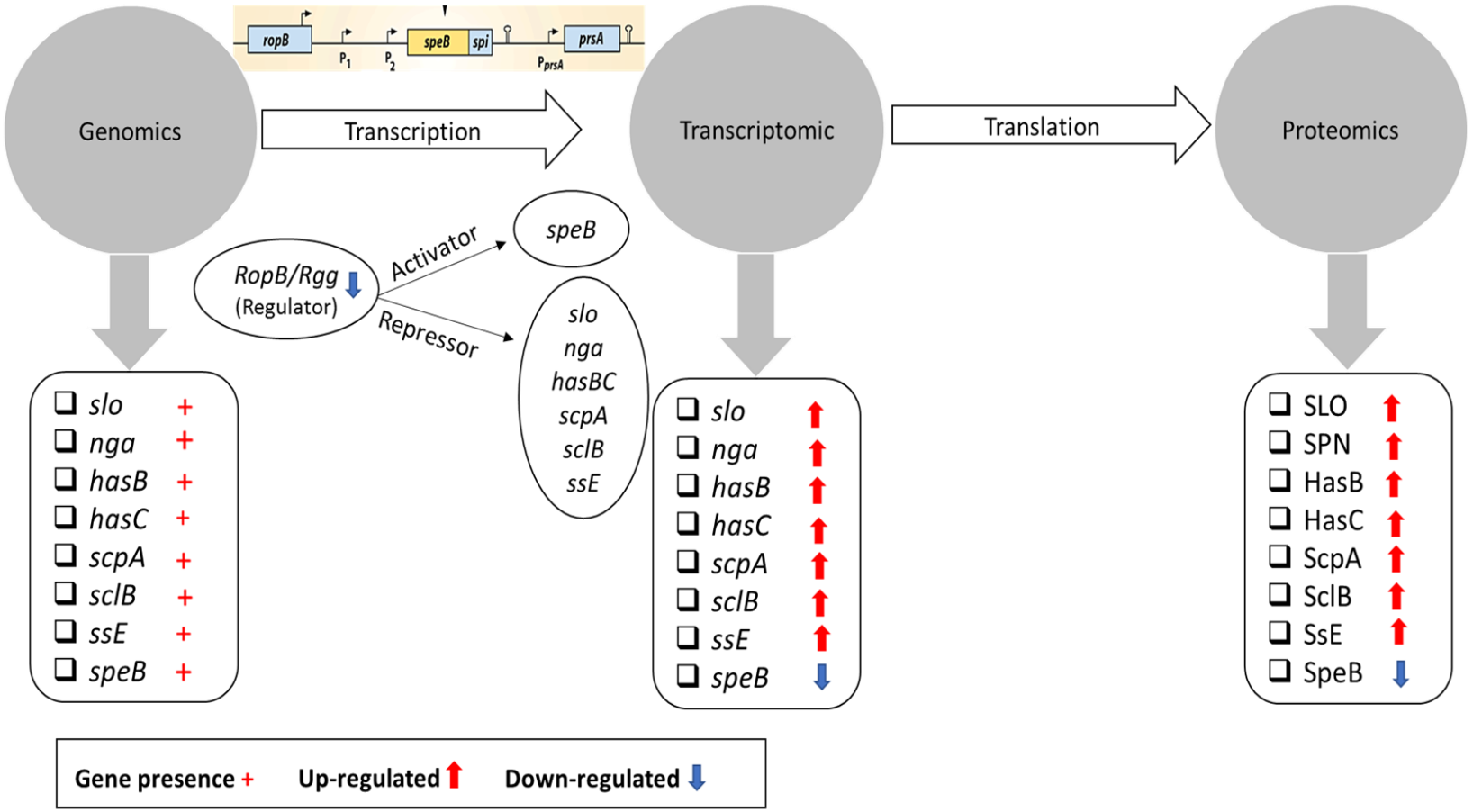
6.1.4 Characterisation and comparative analysis of the secretomes of the *emm*32.2, *emm*1.0 and *emm*89.0 isolates.

Analysis of the transcriptome of *emm*32.2, *emm*1.0, and *emm*89.0 isolates were performed, and genes differentially regulated between the isolates at three distinct growth phases were analysed. To determine the novel secretomic features that are potentially specific to the *emm*32.2 type, I decided to examine the set of proteins that are expressed and secreted to the cultures supernatant in function of the growth phase. In the future, it might also be worth examining the rates of production and degradation of the proteins of interest.

Our data exhibited that several recognised GAS virulence factors were more abundant in *emm*32.2 compared to *emm*1.0 and *emm*89.0 isolates, including UDP-glucose 6 Dehydrogenase, collagen-like surface protein, C5a peptidase, Streptolysin O, SPN, and putative cell envelope proteinase which has been shown to play essential roles in streptococcal pathogenesis [11]. These data offer a prospective explanation supportive of *emm*32.2 invasiveness. Interestingly, relative abundance of 4 non-M-protein antigens including SLO, ScpA, putative esterase and putative lipoprotein protein was high in the *emm*32.2 type. Three of these proteins are likely to be involved in GAS pathogenesis [11, 239]. As this is the first study to characterise the proteomics of an *emm*32.2 isolate *in vitro*, my work gives new insights into the relative abundance of GAS secreted proteins, including non-M-protein antigens, which may assist in the development of potential GAS vaccine strategies.

Overall, my project provides evidence that virulence factors such as SLO and SPN are essential for the virulence of *emm*32.2 GAS. Characterisation of group A streptococcus outbreak isolates using high-throughput phenotypic, transcriptomic and proteomic analysis reveal genes associated with highly invasive strains (Figure 6-1) and gives new insights into the development of potential GAS vaccine strategies.

Figure 6-1 Schematic representation of hypothetical regulatory mechanisms in Group A Streptococcus.



Representation of Omic data performed in this project for the *emm*32.2 isolate (112327) relative to *emm*1.0 strain (101910) and *emm*89.0 (127746), showing the variations seen in the speB gene . P1and P2 indicate the two speB promoters (adapted from [179]).

Altogether, my omics analyses converge towards the hypothesis that the Rop/Rgg system is at the centre of the differential clinical phenotype observed between *emm*32.2 and *emm*1.0 and *emm*89.0 GAS strains (Figure 6-1). The *Rop/Rgg* is adjacent to *speB* and inactivation of *Rop/Rgg* abrogates *speB* expression [249]. *Rop/Rgg* binds to the *speB* promoter regions to activate its transcription [219]. Rop/Rgg influences the expression of a variety of secreted virulence factors, such as *speB*, but also *slo* and *nga* [249, 250]. Our data suggest that lower expression levels of *rop/rgg* may be associated with increased expression levels of *slo* and *nga* concomitantly with decreased expression of *speB* in the *emm*32.2 strain (112327) compared to *emm*1.0 (101910) and *emm*89.0 (127746). In the exponential phase of growth, a similar pattern was seen in the proteomics analyses.

In summary, these results re-emphasizing the importance of regulatory networks of GAS such as those mediated by the *Rop/Rgg* system in the control of virulence factors gene expression, such as *slo* and *nga.* Investigations aiming to develop therapeutic or preventive strategies against GAS should be designed to disrupt the Rgg system.

6.2 Limitation

The transcriptomic analysis of *emm*32.2, *emm*1.0 and *emm*89.0 showed that key virulence factors were differentially expressed *in vitro*. However, there are some limitations. The *emm*1 type MGAS5005 publicly available genome was selected as a reference because it has been previously used for the same purpose in two recent transcriptomic studies of GAS strains [157, 186], thus aiding to the comparison of our results with similar data from other GAS isolates. The fact that to date a complete *emm*32.2 genome has not been reported added to the decision of using the MGAS5005 genome as our reference. However, based on our previous findings about the genomic diversity of the *emm*32.2 strains [146], we are aware that there are *emm*32.2-specific genes (including its prophage content) that are not represented in the MGAS5005 genome ,therefore, the information about their expression is missing from our analyses. Future analyses should address the expression of such strain-specific genes in order to investigate their implications in the clinical phenotype of *emm*32.2 isolates. An additional limitation in our transcriptomic approach resides in the total absence of biological replicas in the analysis which might limit our identification of differentially expressed genes.

6.3 Future prospects

The data generated during my project raise some key unanswered questions. For example, our data described phenotype heterogeneity between the *emm*32.2 strains and *emm*1.0 *emm*89.0, and *emm*6.0 strains. Future investigations into the regulation of the *has* operon promoter should provide additional insights into the transcriptional control of capsule gene expression and its importance in the adaptation of the organism to the various environmental niches encountered during the process of infection. In addition, both transcriptomic and proteomic datasets showed that there were several differences between *emm*32.2, *emm*1.0 and *emm*89.0 in virulence factors particularly in the expression of *slo* and nga. As *nga* and *slo* encode for secreted toxin virulence factors which are known to be involved in GAS pathogenesis, further investigations will be directed towards understanding the molecular mechanisms underlying upregulation of these two genes [11]. We could, for example, interrogate the *emm*32.2 (112327) genomes for polymorphisms. We could also search for evidence of acquisition of a positive regulator gene. These types of genetic alterations may account for the increased expression of *nga* and *slo* genes. We next could analyse *emm*32.2 isolates (112327) genome sequences for polymorphisms in the upstream promoter region of the *nga* gene that could alter transcription. The literature has demonstrated that polymorphisms within the *nga* upstream region contribute to increased production of *nga* and *slo*. [251]. We could also examine regulatory genes known to influence virulence, such as *covRS, RopB/Rgg, and mga* for polymorphisms. The outcome of

these investigations may have implications for future research e.g., in GAS human vaccine development.

In addition, the next step would be to investigate the differential gene expression of *emm*32.2 isolates *in vivo* using a GAS animal model. This approach has been limited by the ability to separate relatively small quantities of bacterial RNA from the host RNA. It also would be of interest to perform the analysis of gene expression of *emm*32.2 *in vivo* at different time points and in different tissue niches during the progression of GAS infection, e.g., nasopharynx, nasal-associated lymphoid tissue (NALT), which function analogously to human tonsils [252, 253]. NALT tissue is considered as a significant site of GAS colonisation and demonstrated to be important in the evaluation vaccine candidates *in vivo* [254, 255]. Studies aimed to examine *in vivo* bacterial gene expression will be critical for a fuller understanding of GAS bacterial pathogenesis, host-pathogen interactions, and rational vaccine design.

Moving forward, based on our work which showed that *emm*32.2 isolates had different phenotypic characteristics from *emm*1.0, *emm*89.0, and *emm*6.0 isolates, it would be useful to determine the contribution of SLO and SPN in the invasiveness of GAS (e.g., murine models). The relative contribution of each of these two toxins to pathogenesis of *emm*32.2 strains remains unexplored hence would help to link *in vitro* data with *in vivo* work. Our proteomic data also have demonstrated significant differences between *emm*32.2, *emm*1.0 and *emm*89.0 isolates. Further fractionation of our bacterial cultures including cell membranes, would allow us to increase our proteome coverage. The cell membrane is involved in many crucial processes, such as transport of waste products

or nutrients, and cell-cell communication, all of which are essential for bacterial survival. Therefore, It also would be of interest to target the bacterial membrane and its associated proteins as putative vaccine candidates [256]. These proteins are likely to be underrepresented in the supernatant analysis undertaken in this work. Although my proteomic data has demonstrated that several recognised GAS virulence factors were more abundant in *emm*32.2 compared to *emm*1.0 and *emm*89.0 isolates, including SLO and SPN, the molecular events that allow for widespread dissemination of this disease are not fully understood and requires further investigation. Also, exploring the contribution of other putative GAS virulence factors such as *RopB/Rgg* and *ennX* encodes an M-like protein may shed further light on the virulence of *emm*32.2 [11]. One powerful approach to further assess the impact of the virulence factors we identified in this project as potentially relevant for the *emm*32.2 pathogenic phenotype, such as *slo and nga,* is by generating knock out mutants of these genes in the bacterial chromosome. These mutants can then be subject of characterisation both *in vitro* and *in vivo* models to evaluate their role in septic arthritis and invasive disease. This type of approaches would also be beneficial in finding new drug targets for the development of antimicrobial therapies.

6.4 Closing remarks

The development of a robust vaccine has become essential, as GAS infections are associated with a variety of autoimmune sequelae, such as acute rheumatic fever, rheumatic heart disease, and acute glomerulonephritis [198]. Lack of effective treatment of such infections may result in higher morbidity and mortality rates, eventually having a significant impact on human public health [257]. The burden of GAS disease also remains a global challenge because of the capacity of GAS to cause outbreaks in a number of developed countries [11]. The issue of emerging antibiotic resistance in GAS (excluding penicillin and cephalosporin) is also an increasing concern which poses a serious threat to the treatment of infection. More specifically, resistance to clindamycin is of particular interest, as it is often used in combination with penicillin to treat GAS infections [258]. Due to the emergence of antimicrobial resistance, the development of new antimicrobial agents has become a priority [257].

GAS has remarkable ability to adapt to a variety of human tissues and numerous virulence factors have been described that allow this single species of bacterium to produce a wide range of disease [198]. It is not yet known how different streptococcal virulence factors interact with the host to produce these diverse diseases. GAS systemic disease reflects its ability to resist clearance by the immune defence mechanisms. GAS is not only able to delay phagocyte recruitment but also escape neutrophil extracellular traps, inhibit complement, and promote phagocyte lysis [11].

We need to better understand this complex bacterial pathogen, as a delicate interplay between bacterial virulence factor expression and interference with the host immune system, if we are to improve therapies or deliver effective vaccines.

Understanding the molecular basis of GAS-mediated immune resistance would reveal novel therapeutic targets for treatment or prevention of invasive infections. As safe and effective vaccines against GAS infections are not available, vaccine candidates or antimicrobial therapies are urgently needed.

It is also essential to explore new therapeutic strategies of coping with GAS infections. To date, my work represents one of the most comprehensive GAS analyses including both transcriptome and secretome analysis of an invasive *emm* type 32.2. Here we use an invasive and clinically relevant streptococcal *emm*32.2 and *emm*1.0 invasive and *emm*89.0 non-invasive to characterise the bacterial transcriptome and proteome in-depth. In order to investigate novel drug targets, it is fundamental to gain a better understanding of the proteins produced by the bacteria and, their cellular location, and their relative amounts. Therefore, one part of our work focused on the secretome because many extracellular proteins play an important role in streptococcal pathogenesis [199]. Thus, these proteins are interesting targets for drug development. We need to explore more about virulence factor genes and secreted proteins which are commonly targeted in vaccine research to develop new strategies to cope with GAS infections. We have shown both phenotypically and through comparative transcriptomic and proteomic analysis that isolates from *emm* type 32.2 are diverse

and potentially more capable of an invasive phenotype due to upregulation of key virulence genes.

The presented data provide new insights into the relative abundance of GAS secreted proteins, including non-M-protein antigens, which may contribute to the development of potential GAS vaccine strategies. Overall, the results obtained in this study may be used as a basis to further study the gene expression and proteins of various *emm*32.2 isolates in order to understand the mechanisms involved in invasive disease, which give an insight to the development of more efficient therapeutic and preventive strategies.

References

1. Jawetz, E., J. Melnich, and E. Adelbery, *Medical Microbiology LANGE.* New York, Mc Grow Hill, 2001: p. 88-108.
2. Todar, K., *Streptococcus pyogenes.* Todar's online textbook of bacteriology, University of Wisconsin-Madison Department of Bacteriology, 2002.
3. Lancefield, R.C., *A serological differentiation of human and other groups of hemolytic streptococci.* Journal of experimental medicine, 1933. **57**(4): p. 571-595.
4. Facklam, R., *What happened to the streptococci: overview of taxonomic and nomenclature changes.* Clinical microbiology reviews, 2002. **15**(4): p. 613-630.
5. Brandt, C.M., et al., *Characterization of Blood Culture Isolates ofStreptococcus dysgalactiae subsp. equisimilisPossessing Lancefield's Group A Antigen.* Journal of clinical microbiology, 1999. **37**(12): p. 4194-4197.
6. Malhotra-Kumar, S., et al., *Bacitracin-resistant clone of Streptococcus pyogenes isolated from pharyngitis patients in Belgium.* Journal of clinical microbiology, 2003. **41**(11): p. 5282-5284.
7. Mihaila-Amrouche, L., A. Bouvet, and J. Loubinoux, *Clonal spread of emm type 28 isolates of Streptococcus pyogenes that are multiresistant to antibiotics.* Journal of clinical microbiology, 2004. **42**(8): p. 3844-3846.
8. James, L. and R.B. McFarland, *An epidemic of pharyngitis due to a nonhemolytic group A streptococcus at Lowry Air Force Base.* New England journal of medicine, 1971. **284**(14): p. 750-752.
9. Beall, B., R. Facklam, and T. Thompson, *Sequencing emm-specific PCR products for routine and accurate typing of group A streptococci.* Journal of Clinical Microbiology, 1996. **34**(4): p. 953-958.
10. Steer, A.C., et al., *Global emm type distribution of group A streptococci: systematic review and implications for vaccine development.* The Lancet infectious diseases, 2009. **9**(10): p. 611-616.
11. Walker, M.J., et al., *Disease manifestations and pathogenic mechanisms of group A Streptococcus.* Clinical microbiology reviews, 2014. **27**(2): p. 264-301.
12. O'loughlin, R.E., et al., *The epidemiology of invasive group A streptococcal infection and potential vaccine implications: United States, 2000–2004.* Clinical Infectious Diseases, 2007. **45**(7): p. 853-862.

13. Dooling, K.L., et al., *Investigation of a prolonged group A streptococcal outbreak among residents of a skilled nursing facility, Georgia, 2009–2012.* Clinical infectious diseases, 2013. **57**(11): p. 1562-1567.
14. Deutscher, M., et al., *Investigation of a group A streptococcal outbreak among residents of a long-term acute care hospital.* Clinical infectious diseases, 2011. **52**(8): p. 988-994.
15. Harris, A.M., et al., *Community-acquired invasive GAS disease among native Americans, Arizona, USA, winter 2013.* Emerging infectious diseases, 2015. **21**(1): p. 177.
16. Smoot, J.C., et al., *Genome sequence and comparative microarray analysis of serotype M18 group A Streptococcus strains associated with acute rheumatic fever outbreaks.* Proceedings of the National Academy of Sciences, 2002. **99**(7): p. 4668-4673.
17. Stevens, D.L., *Streptococcal toxic-shock syndrome: spectrum of disease, pathogenesis, and new concepts in treatment.* Emerging infectious diseases, 1995. **1**(3): p. 69.
18. Ralph, A.P. and J.R. Carapetis, *Group a streptococcal diseases and their global burden*, in *Host-Pathogen Interactions in Streptococcal Diseases*. 2012, Springer. p. 1-27.
19. Wasserzug, O., et al., *A cluster of ecthyma outbreaks caused by a single clone of invasive and highly infective Streptococcus pyogenes.* Clinical Infectious Diseases, 2009. **48**(9): p. 1213-1219.
20. Efstratiou, A. and T. Lamagni, *Epidemiology of Streptococcus pyogenes.* 2016.
21. Martin, J.M., et al., *Group A streptococci among school-aged children: clinical characteristics and the carrier state.* Pediatrics, 2004. **114**(5): p. 1212-1219.
22. Steer, A.C., et al., *Invasive group A streptococcal disease.* Drugs, 2012. **72**(9): p. 1213-1227.
23. Carapetis, J.R., et al., *The global burden of group A streptococcal diseases.* The Lancet infectious diseases, 2005. **5**(11): p. 685-694.
24. Gerber, M.A., et al., *Prevention of rheumatic fever and diagnosis and treatment of acute streptococcal pharyngitis.* Circulation, 2009. **119**(11): p. 1541-1551.
25. Choby, B.A., *Diagnosis and treatment of streptococcal pharyngitis.* Am Fam Physician, 2009. **79**(5): p. 383-390.
26. Steer, A.C., et al., *Group A streptococcal vaccines: facts versus fantasy.* Current opinion in infectious diseases, 2009. **22**(6): p. 544-552.
27. Musser, J.M. and S.A. Shelburne, *A decade of molecular pathogenomic analysis of group A Streptococcus.* The Journal of clinical investigation, 2009. **119**(9): p. 2455-2463.
28. Nasser, W., et al., *Evolutionary pathway to increased virulence and epidemic group A Streptococcus disease derived from 3,615 genome sequences.* Proceedings of the National Academy of Sciences, 2014. **111**(17): p. E1768-E1776.
29. Guy, R., et al., *Increase in scarlet fever notifications in the United Kingdom, 2013/2014.* Eurosurveillance, 2014. **19**(12): p. 20749.

30. You, Y., et al., *Scarlet fever epidemic in China caused by Streptococcus pyogenes serotype M12: epidemiologic and molecular analysis.* EBioMedicine, 2018. **28**: p. 128-135.
31. Davies, M.R., et al., *Emergence of scarlet fever Streptococcus pyogenes emm12 clones in Hong Kong is associated with toxin acquisition and multidrug resistance.* Nature genetics, 2015. **47**(1): p. 84.
32. Zakour, N.L.B., et al., *Transfer of scarlet fever-associated elements into the group A Streptococcus M1T1 clone.* Scientific reports, 2015. **5**.
33. Athey, T.B., et al., *High incidence of invasive group A Streptococcus disease caused by strains of uncommon emm types in Thunder Bay, Ontario, Canada.* Journal of clinical microbiology, 2016. **54**(1): p. 83-92.
34. Turner, C.E., et al., *Emergence of a new highly successful acapsular group A Streptococcus clade of genotype emm89 in the United Kingdom.* MBio, 2015. **6**(4): p. e00622-15.
35. Tse, H., et al., *Molecular characterization of the 2011 Hong Kong scarlet fever outbreak.* The Journal of infectious diseases, 2012. **206**(3): p. 341-351.
36. Sumby, P., et al., *Evolutionary origin and emergence of a highly successful clone of serotype M1 group A Streptococcus involved multiple horizontal gene transfer events.* The Journal of infectious diseases, 2005. **192**(5): p. 771-782.
37. Wessels, M.R., *Pharyngitis and Scarlet Fever.* 2016.
38. Ebell, M.H., et al., *Does this patient have strep throat?* Jama, 2000. **284**(22): p. 2912-2918.
39. Shaikh, N., E. Leonard, and J.M. Martin, *Prevalence of streptococcal pharyngitis and streptococcal carriage in children: a meta-analysis.* Pediatrics, 2010. **126**(3): p. e557-e564.
40. Shulman, S.T., et al., *Clinical Practice Guideline for the Diagnosis and Management of Group A Streptococcal Pharyngitis: 2012 Update by the Infectious Diseases Society of America.* Clinical Infectious Diseases, 2012. **55**(10): p. e86-e102.
41. Wessels, M.R., *Streptococcal pharyngitis.* New England Journal of Medicine, 2011. **364**(7): p. 648-655.
42. Barash, J., *Rheumatic Fever and Post-Group A Streptococcal Arthritis in Children.* Current Infectious Disease Reports, 2013. **15**(3): p. 263-268.
43. Cunningham, M.W., *Streptococcus and rheumatic fever.* Current opinion in rheumatology, 2012. **24**(4): p. 408.
44. Block, S.L., *Streptococcal pharyngitis: Guidelines, treatment issues, and sequelae*. 2014, SLACK INC 6900 GROVE RD, THOROFARE, NJ 08086 USA.
45. Carapetis, J.R., M. McDonald, and N.J. Wilson, *Acute rheumatic fever.* The Lancet, 2005. **366**(9480): p. 155-168.
46. Veasy, L.G., et al., *Resurgence of acute rheumatic fever in the intermountain area of the United States.* New England journal of medicine, 1987. **316**(8): p. 421-427.
47. Stevens, D.L. and A.E. Bryant, *Severe group A streptococcal infections.* 2016.

48. Kihiczak, G., R. Schwartz, and R. Kapila, *Necrotizing fasciitis: a deadly infection.* Journal of the European Academy of Dermatology and Venereology, 2006. **20**(4): p. 365-369.
49. Herold, A.H., *Group A beta-hemolytic streptococcal toxic shock from a mild pharyngitis.* Journal of Family Practice, 1990. **31**(5): p. 549-552.
50. Bradley, J.S., P.M. Schlievert, and B.M. Peterson, *Toxic shock-like syndrome, a complication of strep throat.* The Pediatric infectious disease journal, 1991. **10**(10): p. 790.
51. Chapnick, E.K., et al., *Streptococcal toxic shock syndrome due to noninvasive pharyngitis.* Clinical infectious diseases, 1992. **14**(5): p. 1074-1077.
52. Holm, S.E., et al., *Aspects of pathogenesis of serious group A streptococcal infections in Sweden, 1988–1989.* Journal of Infectious Diseases, 1992. **166**(1): p. 31-37.
53. Sellers, B.J., et al., *Necrotizing group A streptococcal infections associated with streptococcal toxic shock syndrome.* The American journal of surgery, 1996. **172**(5): p. 523-528.
54. Kiska, D.L., et al., *Invasive group A streptococcal infections in North Carolina: epidemiology, clinical features, and genetic and serotype analysis of causative organisms.* Journal of Infectious Diseases, 1997. **176**(4): p. 992-1000.
55. Horn, D.L., et al., *Why have group A streptococci remained susceptible to penicillin? Report on a symposium.* Clinical infectious diseases, 1998. **26**(6): p. 1341-1345.
56. Bisno, A.L., et al., *Practice guidelines for the diagnosis and management of group A streptococcal pharyngitis.* Clinical Infectious Diseases, 2002: p. 113-125.
57. McNeil, S.A., et al., *Safety and immunogenicity of 26-valent group a streptococcus vaccine in healthy adult volunteers.* Clinical Infectious Diseases, 2005. **41**(8): p. 1114-1122.
58. Hu, M.C., et al., *Immunogenicity of a 26-valent group A streptococcal vaccine.* Infection and immunity, 2002. **70**(4): p. 2171-2177.
59. Dale, J.B., *Current status of group A streptococcal vaccine development*, in *Hot topics in infection and immunity in children IV*. 2008, Springer. p. 53-63.
60. Sauer, K., et al., *Pseudomonas aeruginosa displays multiple phenotypes during development as a biofilm.* Journal of bacteriology, 2002. **184**(4): p. 1140-1154.
61. Stoodley, P., et al., *Biofilms as complex differentiated communities.* Annual Reviews in Microbiology, 2002. **56**(1): p. 187-209.
62. O'toole, G.A. and R. Kolter, *Flagellar and twitching motility are necessary for Pseudomonas aeruginosa biofilm development.* Molecular microbiology, 1998. **30**(2): p. 295-304.
63. Doern, C.D., et al., *Biofilm formation by group A Streptococcus: a role for the streptococcal regulator of virulence (Srv) and streptococcal cysteine protease (SpeB).* Microbiology, 2009. **155**(1): p. 46-52.

64. Cho, K.H. and M.G. Caparon, *Patterns of virulence gene expression differ between biofilm and tissue communities of Streptococcus pyogenes.* Molecular microbiology, 2005. **57**(6): p. 1545-1556.
65. Akiyama, H., et al., *Assessment of Streptococcus pyogenes microcolony formation in infected skin by confocal laser scanning microscopy.* Journal of dermatological science, 2003. **32**(3): p. 193-199.
66. Neely, M.N., J.D. Pfeifer, and M. Caparon, *Streptococcus-zebrafish model of bacterial pathogenesis.* Infection and immunity, 2002. **70**(7): p. 3904-3914.
67. Lembke, C., et al., *Characterization of biofilm formation by clinically relevant serotypes of group A streptococci.* Applied and environmental microbiology, 2006. **72**(4): p. 2864-2875.
68. Sanchez, C.J., et al., *Streptococcus pneumoniae in biofilms are unable to cause invasive disease due to altered virulence determinant production.* PLoS One, 2011. **6**(12): p. e28738.
69. Mayanskiy, A., et al., *Biofilm formation by Streptococcus pneumoniae.* Molecular Genetics, Microbiology and Virology, 2015. **30**(3): p. 124-131.
70. Baldassarri, L., et al., *Therapeutic failures of antibiotics used to treat macrolide-susceptible Streptococcus pyogenes infections may be due to biofilm formation.* Journal of Clinical Microbiology, 2006. **44**(8): p. 2721-2727.
71. Marks, L.R., et al., *Streptococcus pyogenes biofilm growth in vitro and in vivo and its role in colonization, virulence, and genetic exchange.* The Journal of infectious diseases, 2014. **210**(1): p. 25-34.
72. Fiedler, T., T. Köller, and B. Kreikemeyer, *Streptococcus pyogenes biofilms—formation, biology, and clinical relevance.* Frontiers in cellular and infection microbiology, 2015. **5**: p. 15.
73. Dale, J.B., et al., *Hyaluronate capsule and surface M protein in resistance to opsonization of group A streptococci.* Infection and immunity, 1996. **64**(5): p. 1495-1501.
74. Levin, J.C. and M.R. Wessels, *Identification of csrR/csrS, a genetic locus that regulates hyaluronic acid capsule synthesis in group A Streptococcus.* Molecular microbiology, 1998. **30**(1): p. 209-219.
75. Bernish, B. and I. van de Rijn, *Characterization of a two-component system in Streptococcus pyogenes which is involved in regulation of hyaluronic acid production.* Journal of Biological Chemistry, 1999. **274**(8): p. 4786-4793.
76. Albertí, S., C.D. Ashbaugh, and M.R. Wessels, *Structure of the has operon promoter and regulation of hyaluronic acid capsule expression in group A Streptococcus.* Molecular microbiology, 1998. **28**(2): p. 343-353.
77. Biswas, I. and J.R. Scott, *Identification of rocA, a positive regulator of covR expression in the group A streptococcus.* Journal of bacteriology, 2003. **185**(10): p. 3081-3090.
78. Jain, I., et al., *RocA is an accessory protein to the virulence-regulating CovR/S two-component system in the group A Streptococcus.* Infection and immunity, 2017: p. IAI. 00274-17.

79. Wessels, M.R., et al., *Hyaluronic acid capsule is a virulence factor for mucoid group A streptococci.* Proceedings of the National Academy of Sciences, 1991. **88**(19): p. 8317-8321.
80. Stollerman, G.H. and J.B. Dale, *The importance of the group a streptococcus capsule in the pathogenesis of human infections: a historical perspective.* Clinical Infectious Diseases, 2008. **46**(7): p. 1038-1045.
81. Dinkla, K., et al., *Upregulation of capsule enables Streptococcus pyogenes to evade immune recognition by antigen-specific antibodies directed to the G-related α 2-macroglobulin-binding protein GRAB located on the bacterial surface.* Microbes and infection, 2007. **9**(8): p. 922-931.
82. Foley, M.J., M.R. Smith, and W.B. Wood Jr, *Studies on the Pathogenicity of Group A Streptococci: I. Its Relation to Surface Phagocytosis.* The Journal of experimental medicine, 1959. **110**(4): p. 603.
83. Cole, J.N., et al., *M protein and hyaluronic acid capsule are essential for in vivo selection of covRS mutations characteristic of invasive serotype M1T1 group A Streptococcus.* MBio, 2010. **1**(4): p. e00191-10.
84. Moses, A.E., et al., *Relative contributions of hyaluronic acid capsule and M protein to virulence in a mucoid strain of the group A Streptococcus.* Infection and immunity, 1997. **65**(1): p. 64-71.
85. Ashbaugh, C.D., et al., *Bacterial determinants of persistent throat colonization and the associated immune response in a primate model of human group A streptococcal pharyngeal infection.* Cellular microbiology, 2000. **2**(4): p. 283-292.
86. Panchaud, A., et al., *M-protein and other intrinsic virulence factors of Streptococcus pyogenes are encoded on an ancient pathogenicity island.* BMC genomics, 2009. **10**(1): p. 198.
87. McNamara, C., et al., *Coiled-coil irregularities and instabilities in group A Streptococcus M1 are required for virulence.* Science, 2008. **319**(5868): p. 1405-1408.
88. Courtney, H., et al., *Binding of Streptococcus pyogenes to soluble and insoluble fibronectin.* Infection and immunity, 1986. **53**(3): p. 454-459.
89. Kantor, F.S., *Fibrinogen precipitation by streptococcal M protein.* Journal of Experimental Medicine, 1965. **121**(5): p. 861-872.
90. Fischetti, V.A., *Streptococcal M protein: molecular design and biological behavior.* Clinical microbiology reviews, 1989. **2**(3): p. 285-314.
91. Johnson, D.R., et al., *Characterization of group A streptococci (Streptococcus pyogenes): correlation of M-protein and emm-gene type with T-protein agglutination pattern and serum opacity factor.* Journal of medical microbiology, 2006. **55**(2): p. 157-164.
92. Tandon, R., et al., *Revisiting the pathogenesis of rheumatic fever and carditis.* Nature Reviews Cardiology, 2013. **10**(3): p. 171-177.
93. Aziz, R.K. and M. Kotb, *Rise and persistence of global M1T1 clone of Streptococcus pyogenes.* Emerging infectious diseases, 2008. **14**(10): p. 1511.
94. Oehmcke, S., et al., *Streptococcal M proteins and their role as virulence determinants.* Clinica Chimica Acta, 2010. **411**(17): p. 1172-1180.

95. Nelson, D.C., J. Garbe, and M. Collin, *Cysteine proteinase SpeB from Streptococcus pyogenes–a potent modifier of immunologically important host and bacterial proteins.* Biological chemistry, 2011. **392**(12): p. 1077-1088.
96. Sumitomo, T., et al., *Group A streptococcal cysteine protease cleaves epithelial junctions and contributes to bacterial translocation.* Journal of Biological Chemistry, 2013. **288**(19): p. 13317-13324.
97. Aziz, R.K., et al., *Invasive M1T1 group A Streptococcus undergoes a phase-shift in vivo to prevent proteolytic degradation of multiple virulence factors by SpeB.* Molecular microbiology, 2004. **51**(1): p. 123-134.
98. Honda-Ogawa, M., et al., *Cysteine proteinase from Streptococcus pyogenes enables evasion of innate immunity via degradation of complement factors.* Journal of Biological Chemistry, 2013. **288**(22): p. 15854-15864.
99. Anderson, B. and J. Duncan, *Activation of human neutrophil metabolism by strepolysin 0.* J. Infect. Dis, 1980. **141**: p. 680-685.
100. Fontaine, M.C., J.J. Lee, and M.A. Kehoe, *Combined contributions of streptolysin O and streptolysin S to virulence of serotype M5 Streptococcus pyogenes strain Manfredo.* Infection and immunity, 2003. **71**(7): p. 3857-3865.
101. Limbago, B., et al., *Role of streptolysin O in a mouse model of invasive group A streptococcal disease.* Infection and immunity, 2000. **68**(11): p. 6384-6390.
102. Timmer, A.M., et al., *Streptolysin O promotes group A Streptococcus immune evasion by accelerated macrophage apoptosis.* Journal of Biological Chemistry, 2009. **284**(2): p. 862-871.
103. Madden, J.C., N. Ruiz, and M. Caparon, *Cytolysin-mediated translocation (CMT): a functional equivalent of type III secretion in gram-positive bacteria.* Cell, 2001. **104**(1): p. 143-152.
104. Ato, M., et al., *Incompetence of neutrophils to invasive group A streptococcus is attributed to induction of plural virulence factors by dysfunction of a regulator.* PloS one, 2008. **3**(10): p. e3455.
105. Brosnahan, A.J. and P.M. Schlievert, *Gram-positive bacterial superantigen outside-in signaling causes toxic shock syndrome.* FEBS Journal, 2011. **278**(23): p. 4649-4667.
106. Reglinski, M. and S. Sriskandan, *The contribution of group A streptococcal virulence determinants to the pathogenesis of sepsis.* Virulence, 2014. **5**(1): p. 127-136.
107. Sheeler, R.D., et al., *Accuracy of rapid strep testing in patients who have had recent streptococcal pharyngitis.* The Journal of the American Board of Family Practice, 2002. **15**(4): p. 261-265.
108. Feil, S.C., et al., *Structural studies of Streptococcus pyogenes streptolysin O provide insights into the early steps of membrane penetration.* Journal of molecular biology, 2014. **426**(4): p. 785-792.
109. Ghosh, J., et al., *Characterization of Streptococcus pyogenes β-NAD+ Glycohydrolase RE-EVALUATION OF ENZYMATIC PROPERTIES ASSOCIATED WITH PATHOGENESIS.* Journal of Biological Chemistry, 2010. **285**(8): p. 5683-5694.

110. Michos, A., et al., *Enhancement of streptolysin O activity and intrinsic cytotoxic effects of the group A streptococcal toxin, NAD-glycohydrolase.* Journal of Biological Chemistry, 2006. **281**(12): p. 8216-8223.
111. O'Seaghdha, M. and M.R. Wessels, *Streptolysin O and its co-toxin NAD-glycohydrolase protect group A Streptococcus from Xenophagic killing.* PLoS pathogens, 2013. **9**(6): p. e1003394.
112. Bricker, A.L., et al., *NAD+-glycohydrolase acts as an intracellular toxin to enhance the extracellular survival of group A streptococci.* Molecular microbiology, 2002. **44**(1): p. 257-269.
113. Ajdic, D., et al., *The NAD-glycohydrolase (nga) gene of Streptococcus pyogenes.* FEMS microbiology letters, 2000. **191**(2): p. 235-241.
114. Riddle, D.J., D.E. Bessen, and M.G. Caparon, *Variation in Streptococcus pyogenes NAD+ glycohydrolase is associated with tissue tropism.* Journal of bacteriology, 2010. **192**(14): p. 3735-3746.
115. Molloy, E.M., et al., *Streptolysin S-like virulence factors: the continuing sagA.* Nature Reviews Microbiology, 2011. **9**(9): p. 670-681.
116. Miyoshi-Akiyama, T., et al., *Cytocidal effect of Streptococcus pyogenes on mouse neutrophils in vivo and the critical role of streptolysin S.* The Journal of infectious diseases, 2005. **192**(1): p. 107-116.
117. Ofek, I., S. Bergner-Rabinowitz, and I. Ginsburg, *Oxygen-stable Hemolysins of Group A Streptococci VII. The Relation of the Leukotoxic Factor to Streptolysin S.* The Journal of infectious diseases, 1970: p. 517-522.
118. Carr, A., et al., *Similarities between complement-mediated and streptolysin S-mediated hemolysis.* Journal of Biological Chemistry, 2001. **276**(45): p. 41790-41796.
119. Ginsburg, I., *Is streptolysin S of group A streptococci a virulence factor?* Apmis, 1999. **107**(7-12): p. 1051-1059.
120. Yoshino, M., et al., *Nonhemolytic Streptococcus pyogenes isolates that lack large regions of the sag operon mediating streptolysin S production.* Journal of clinical microbiology, 2010. **48**(2): p. 635-638.
121. Wannamaker, L.W., *The differentiation of three distinct desoxyribonucleases of group A streptococci.* Journal of Experimental Medicine, 1958. **107**(6): p. 797-812.
122. Tiesler, E. and U. Beck, *Distribution of the isoenzymes of nucleases in group A Streptococci (author's transl).* Zentralblatt fur Bakteriologie, Parasitenkunde, Infektionskrankheiten und Hygiene. Erste Abteilung Originale. Reihe A: Medizinische Mikrobiologie und Parasitologie, 1976. **234**(4): p. 462-472.
123. Sumby, P., et al., *Extracellular deoxyribonuclease made by group A Streptococcus assists pathogenesis by enhancing evasion of the innate immune response.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(5): p. 1679-1684.
124. Walker, M.J., et al., *DNase Sda1 provides selection pressure for a switch to invasive group A streptococcal infection.* Nature medicine, 2007. **13**(8): p. 981-985.

125. Buchanan, J.T., et al., *DNase expression allows the pathogen group A Streptococcus to escape killing in neutrophil extracellular traps.* Current Biology, 2006. **16**(4): p. 396-400.
126. Frick, I.-M., et al., *SIC, a Secreted Protein of Streptococcus pyogenesThat Inactivates Antibacterial Peptides.* Journal of Biological Chemistry, 2003. **278**(19): p. 16561-16566.
127. Åkesson, P., A.G. Sjöholm, and L. Björck, *Protein SIC, a novel extracellular protein of Streptococcus pyogenes interfering with complement function.* Journal of Biological Chemistry, 1996. **271**(2): p. 1081-1088.
128. Fernie-King, B.A., et al., *Streptococcal inhibitor of complement inhibits two additional components of the mucosal innate immune system: secretory leukocyte proteinase inhibitor and lysozyme.* Infection and immunity, 2002. **70**(9): p. 4908-4916.
129. Fernie-King, B.A., et al., *Streptococcal inhibitor of complement (SIC) inhibits the membrane attack complex by preventing uptake of C567 onto cell membranes.* Immunology, 2001. **103**(3): p. 390-398.
130. Fernie-King, B.A., et al., *Streptococcal DRS (distantly related to SIC) and SIC inhibit antimicrobial peptides, components of mucosal innate immunity: a comparison of their activities.* Microbes and infection, 2007. **9**(3): p. 300-307.
131. Frick, I.M., et al., *The contact system—a novel branch of innate immunity generating antibacterial peptides.* The EMBO journal, 2006. **25**(23): p. 5569-5578.
132. Frick, I.-M., et al., *Antibacterial activity of the contact and complement systems is blocked by SIC, a protein secreted by Streptococcus pyogenes.* Journal of Biological Chemistry, 2011. **286**(2): p. 1331-1340.
133. Hoe, N.P., et al., *Insight into the molecular basis of pathogen abundance: group A Streptococcus inhibitor of complement inhibits bacterial adherence and internalization into human cells.* Proceedings of the National Academy of Sciences, 2002. **99**(11): p. 7646-7651.
134. Lukomski, S., et al., *Nonpolar inactivation of the hypervariable streptococcal inhibitor of complement gene (sic) in serotype M1 Streptococcus pyogenes significantly decreases mouse mucosal colonization.* Infection and immunity, 2000. **68**(2): p. 535-542.
135. Hoe, N.P., et al., *Rapid selection of complement-inhibiting protein variants in group A Streptococcus epidemic waves.* Nature medicine, 1999. **5**(8): p. 924-929.
136. Grifantini, R., et al., *Peroxide stimulon and role of PerR in group A Streptococcus.* Journal of bacteriology, 2011. **193**(23): p. 6539-6551.
137. McMillan, D.J., et al., *Immune response to superoxide dismutase in group A streptococcal infection.* FEMS Immunology & Medical Microbiology, 2004. **40**(3): p. 249-256.
138. Kwinn, L.A. and V. Nizet, *How group A Streptococcus circumvents host phagocyte defenses.* 2007.
139. Brenot, A., et al., *Contribution of glutathione peroxidase to the virulence of Streptococcus pyogenes.* Infection and immunity, 2004. **72**(1): p. 408-413.

140. Stenberg, L., P. O'Toole, and G. Lindahl, *Many group A streptococcal strains express two different immunoglobulin-binding proteins, encoded by closely linked genes: characterization of the proteins expressed by four strains of different M-type.* Molecular microbiology, 1992. **6**(9): p. 1185-1194.
141. Carlsson, F., et al., *Evasion of phagocytosis through cooperation between two ligand-binding regions in Streptococcus pyogenes M protein.* Journal of Experimental Medicine, 2003. **198**(7): p. 1057-1068.
142. Fagan, P.K., et al., *Identification and characterization of a novel secreted immunoglobulin binding protein from group A streptococcus.* Infection and immunity, 2001. **69**(8): p. 4851-4857.
143. Trastoy, B., et al., *Crystal structure of Streptococcus pyogenes EndoS, an immunomodulatory endoglycosidase specific for human IgG antibodies.* Proceedings of the National Academy of Sciences, 2014. **111**(18): p. 6714-6719.
144. Collin, M., et al., *EndoS and SpeB from Streptococcus pyogenes inhibit immunoglobulin-mediated opsonophagocytosis.* Infection and immunity, 2002. **70**(12): p. 6646-6651.
145. Sjögren, J., et al., *Study of the IgG endoglycosidase EndoS in group A streptococcal phagocyte resistance and virulence.* BMC microbiology, 2011. **11**(1): p. 120.
146. Cornick, J.E., et al., *Epidemiological and molecular characterization of an invasive group A streptococcus emm32. 2 outbreak.* Journal of clinical microbiology, 2017. **55**(6): p. 1837-1846.
147. McIver, K.S. and R.L. Myles, *Two DNA-binding domains of Mga are required for virulence gene activation in the group A streptococcus.* Molecular microbiology, 2002. **43**(6): p. 1591-1601.
148. Okada, N., R.T. Geist, and M.G. Caparon, *Positive transcriptional control of mry regulates virulence in the group A streptococcus.* Molecular microbiology, 1993. **7**(6): p. 893-903.
149. Kimura, K.R., et al., *Involvement of T6 pili in biofilm formation by serotype M6 Streptococcus pyogenes.* Journal of bacteriology, 2012. **194**(4): p. 804-812.
150. Hynes, W. and J. Ferretti, *Sequence analysis and expression in Escherichia coli of the hyaluronidase gene of Streptococcus pyogenes bacteriophage H4489A.* Infection and immunity, 1989. **57**(2): p. 533-539.
151. Hedges, A., *Estimating the precision of serial dilutions and viable bacterial counts.* International journal of food microbiology, 2002. **76**(3): p. 207-214.
152. Hathaway, L.J., et al., *Capsule type of Streptococcus pneumoniae determines growth phenotype.* PLoS pathogens, 2012. **8**(3): p. e1002574.
153. Henningham, A., et al., *Mutual exclusivity of hyaluronan and hyaluronidase in invasive group A Streptococcus.* Journal of Biological Chemistry, 2014. **289**(46): p. 32303-32315.
154. Hyams, C., et al., *C3b/iC3b deposition on Streptococcus pneumoniae is not affected by HIV infection.* PloS one, 2010. **5**(1): p. e8902.

155. Hynes, W.L. and J.R. Tagg, *A simple plate assay for detection of group A streptococcus proteinase.* Journal of microbiological methods, 1985. **4**(1): p. 25-31.
156. Courtney, H.S., et al., *Relationship between expression of the family of M proteins and lipoteichoic acid to hydrophobicity and biofilm formation in Streptococcus pyogenes.* PLoS One, 2009. **4**(1): p. e4166.
157. Eraso, J.M., et al., *Genomic landscape of intrahost variation in group A Streptococcus: repeated and abundant mutational inactivation of the fabT gene encoding a regulator of fatty acid synthesis.* Infection and immunity, 2016. **84**(12): p. 3268-3281.
158. Langmead, B. and S.L. Salzberg, *Fast gapped-read alignment with Bowtie 2.* Nature methods, 2012. **9**(4): p. 357.
159. Magoc, T., D. Wood, and S.L. Salzberg', *EDGE-pro: estimated degree of gene expression in prokaryotic genomes.* Evolutionary Bioinformatics, 2013. **9**: p. EBO. S11250.
160. Team, R.C., *R: A language and environment for statistical computing.* 2013.
161. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome biology, 2014. **15**(12): p. 550.
162. McClure, R., et al., *Computational analysis of bacterial RNA-Seq data.* Nucleic acids research, 2013. **41**(14): p. e140-e140.
163. Le Breton, Y., et al., *Essential genes in the core genome of the human pathogen Streptococcus pyogenes.* Scientific reports, 2015. **5**: p. 9838.
164. Kreikemeyer, B., et al., *The Streptococcus pyogenes serotype M49 Nra-Ralp3 transcriptional regulatory network and its control of virulence factor expression from the novel eno ralp3 epf sagA pathogenicity region.* Infection and immunity, 2007. **75**(12): p. 5698-5710.
165. Ferretti, J. and W. Köhler, *History of streptococcal research.* 2016.
166. Rice, R.H., G.E. Means, and W.D. Brown, *Stabilization of bovine trypsin by reductive methylation.* Biochimica et Biophysica Acta (BBA)-Protein Structure, 1977. **492**(2): p. 316-321.
167. Crater, D.L. and I. Van de Rijn, *Hyaluronic acid synthesis operon (has) expression in group A streptococci.* Journal of Biological Chemistry, 1995. **270**(31): p. 18452-18458.
168. Unnikrishnan, M., J. Cohen, and S. Sriskandan, *Growth-phase-dependent expression of virulence factors in an M1T1 clinical isolate of Streptococcus pyogenes.* Infection and immunity, 1999. **67**(10): p. 5495-5499.
169. Flores, A., et al., *Human disease isolates of serotype M4 and M22 group A*. 2012, Streptococcus.
170. Nelson, A.L., et al., *Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance.* Infection and immunity, 2007. **75**(1): p. 83-90.
171. Melin, M., et al., *Serotype-related variation in susceptibility to complement deposition and opsonophagocytosis among clinical isolates of Streptococcus pneumoniae.* Infection and immunity, 2010. **78**(12): p. 5252-5261.

172. Hynes, W., *Virulence factors of the group A streptococci and genes that regulate their expression.* Front Biosci, 2004. **9**: p. 3399-3433.
173. Sherwood, N.P., et al., *Studies on Streptococci: V. A Study of Streptococcal Proteinases.* The Journal of infectious diseases, 1954. **95**(1): p. 1-12.
174. Deibel, R., *Hydrolysis of proteins and nucleic acids by Lancefield group A and other streptococci.* Journal of bacteriology, 1963. **86**(6): p. 1270-1274.
175. Rothbard, S. and E.W. Todd, *Production of proteinase by hemolytic streptococci in various clinical conditions.* Journal of Experimental Medicine, 1948. **87**(4): p. 283-294.
176. Kansal, R.G., et al., *Inverse relation between disease severity and expression of the streptococcal cysteine protease, SpeB, among clonal M1T1 isolates recovered from invasive group A streptococcal infection cases.* Infection and immunity, 2000. **68**(11): p. 6362-6369.
177. Kappeler, K.V., et al., *A naturally occurring Rgg variant in serotype M3 Streptococcus pyogenes does not activate speB expression due to altered specificity of DNA binding.* Infection and immunity, 2009. **77**(12): p. 5411-5417.
178. Ma, Y., et al., *Identification and characterization of bicistronic speB and prsA gene expression in the group A Streptococcus.* Journal of bacteriology, 2006. **188**(21): p. 7626-7634.
179. Carroll, R.K. and J.M. Musser, *From transcription to activation: how group A Streptococcus, the flesh-eating pathogen, regulates SpeB cysteine protease production.* Molecular microbiology, 2011. **81**(3): p. 588-601.
180. Ribardo, D.A. and K.S. McIver, *Defining the Mga regulon: comparative transcriptome analysis reveals both direct and indirect regulation by Mga in the group A streptococcus.* Molecular microbiology, 2006. **62**(2): p. 491-508.
181. Zhu, L., et al., *Contribution of secreted NADase and streptolysin O to the pathogenesis of epidemic serotype M1 Streptococcus pyogenes infections.* The American journal of pathology, 2017. **187**(3): p. 605-613.
182. Manetti, A.G., et al., *Environmental acidification drives S. pyogenes pilus expression and microcolony formation on epithelial cells in a FCT-dependent manner.* PLoS One, 2010. **5**(11): p. e13864.
183. Sugareva, V., et al., *Serotype-and strain-dependent contribution of the sensor kinase CovS of the CovRS two-component system to Streptococcus pyogenes pathogenesis.* BMC microbiology, 2010. **10**(1): p. 34.
184. D'Urzo, N., et al., *Acidic pH strongly enhances in vitro biofilm formation by a subset of hypervirulent ST-17 Streptococcus agalactiae strains.* Applied and environmental microbiology, 2014. **80**(7): p. 2176-2185.
185. Lowe, R., et al., *Transcriptomics technologies.* PLoS computational biology, 2017. **13**(5): p. e1005457.
186. Freiberg, J.A., et al., *Global Analysis and Comparison of the Transcriptomes and Proteomes of Group A Streptococcus Biofilms.* MSystems, 2016. **1**(6): p. e00149-16.

187. Vähäkuopus, S., et al., *Distribution of emm types in invasive and non-invasive group A and G streptococci.* European journal of clinical microbiology & infectious diseases, 2012. **31**(6): p. 1251-1256.
188. Shea, P.R., et al., *Group A Streptococcus emm gene types in pharyngeal isolates, Ontario, Canada, 2002–2010.* Emerging infectious diseases, 2011. **17**(11): p. 2010.
189. Bättig, P., et al., *Serotype-specific invasiveness and colonization prevalence in Streptococcus pneumoniae correlate with the lag phase during in vitro growth.* Microbes and infection, 2006. **8**(11): p. 2612-2617.
190. Kreikemeyer, B., K.S. McIver, and A. Podbielski, *Virulence factor regulation and regulatory networks in Streptococcus pyogenes and their impact on pathogen–host interactions.* Trends in microbiology, 2003. **11**(5): p. 224-232.
191. McIver, K.S., A.S. Heath, and J.R. Scott, *Regulation of virulence by environmental signals in group A streptococci: influence of osmolarity, temperature, gas exchange, and iron limitation on emm transcription.* Infection and immunity, 1995. **63**(11): p. 4540-4542.
192. Callegari, E.A., *Rgg regulates growth phase-dependent expression of proteins associated with secondary metabolism and stress in Streptococcus pyogenes.* Journal of bacteriology, 2004. **186**(21): p. 7091-7099.
193. Fiedler, T., et al., *Insights into Streptococcus pyogenes pathogenesis from transcriptome studies.* Future microbiology, 2010. **5**(11): p. 1675-1694.
194. Bastiat-Sempe, B., et al., *Streptolysin O and NAD-glycohydrolase prevent phagolysosome acidification and promote group A streptococcus survival in macrophages.* MBio, 2014. **5**(5): p. e01690-14.
195. Sharma, O., et al., *NAD+-glycohydrolase promotes intracellular survival of group A streptococcus.* PLoS pathogens, 2016. **12**(3): p. e1005468.
196. Van Epps, D.E. and B.R. Andersen, *Streptolysin O inhibition of neutrophil chemotaxis and mobility: nonimmune phenomenon with species specificity.* Infection and immunity, 1974. **9**(1): p. 27-33.
197. Maxted, W., *The indirect bactericidal test as a means of identifying antibody to the M antigen of Streptococcus pyogenes.* British journal of experimental pathology, 1956. **37**(4): p. 415.
198. Cunningham, M.W., *Pathogenesis of group A streptococcal infections.* Clinical microbiology reviews, 2000. **13**(3): p. 470-511.
199. Hynes, W. and M. Sloan, *Secreted extracellular virulence factors.* 2016.
200. Ji, Y., et al., *Intranasal immunization with C5a peptidase prevents nasopharyngeal colonization of mice by the group A Streptococcus.* Infection and immunity, 1997. **65**(6): p. 2080-2087.
201. Rasmussen, M. and L. Björck, *Unique regulation of SclB–a novel collagen-like surface protein of Streptococcus pyogenes.* Molecular microbiology, 2001. **40**(6): p. 1427-1438.
202. Lukomski, S., et al., *Identification and Characterization of a Second Extracellular Collagen-Like Protein Made by Group AStreptococcus: Control of Production at the Level of Translation.* Infection and immunity, 2001. **69**(3): p. 1729-1738.

203. Squeglia, F., et al., *The crystal structure of the streptococcal collagen-like protein 2 globular domain from invasive M3-type group A Streptococcus shows significant similarity to immunomodulatory HIV protein gp41.* Journal of Biological Chemistry, 2014. **289**(8): p. 5122-5133.
204. Anbalagan, S., *Transcriptional regulation of a bacteriophage encoded extracellular DNase (Spd-3) by Rgg in Streptococcus pyogenes.* PloS one, 2013. **8**(4): p. e61312.
205. Nyberg, P., M. Rasmussen, and L. Björck, *α2-Macroglobulin-proteinase complexes protect Streptococcus pyogenes from killing by the antimicrobial peptide LL-37.* Journal of Biological Chemistry, 2004. **279**(51): p. 52820-52823.
206. Kihlberg, B.-M., et al., *Biological properties of a Streptococcus pyogenes mutant generated by Tn916 insertion in mga.* Microbial pathogenesis, 1996. **19**(5): p. 299-315.
207. Stockbauer, K.E., et al., *A natural variant of the cysteine protease virulence factor of group A Streptococcus with an arginine-glycine-aspartic acid (RGD) motif preferentially binds human integrins αvβ3 and αIIbβ3.* Proceedings of the National Academy of Sciences, 1999. **96**(1): p. 242-247.
208. Shelburne III, S.A., et al., *Contribution of AmyA, an extracellular α-glucan degrading enzyme, to group A streptococcal host–pathogen interaction.* Molecular microbiology, 2009. **74**(1): p. 159-174.
209. Commons, R.J., et al., *Streptococcal superantigens: categorization and clinical associations.* Trends in molecular medicine, 2014. **20**(1): p. 48-62.
210. Talkington, D.F., et al., *Association of phenotypic and genotypic characteristics of invasive Streptococcus pyogenes isolates with clinical components of streptococcal toxic shock syndrome.* Infection and immunity, 1993. **61**(8): p. 3369-3374.
211. Musser, J.M., et al., *Streptococcus pyogenes causing toxic-shock-like syndrome and other invasive diseases: clonal diversity and pyrogenic exotoxin expression.* Proceedings of the National Academy of Sciences, 1991. **88**(7): p. 2668-2672.
212. Eriksson, B.K., et al., *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.* The Journal of infectious diseases, 1999. **180**(2): p. 410-418.
213. Friães, A., et al., *Superantigen gene complement of Streptococcus pyogenes—relationship with other typing methods and short-term stability.* European journal of clinical microbiology & infectious diseases, 2013. **32**(1): p. 115-125.
214. Meisal, R., et al., *Streptococcus pyogenes isolates causing severe infections in Norway in 2006 to 2007: emm types, multilocus sequence types, and superantigen profiles.* Journal of clinical microbiology, 2010. **48**(3): p. 842-851.
215. Vega, L.A., H. Malke, and K.S. McIver, *Virulence-related transcriptional regulators of Streptococcus pyogenes.* 2016.

216. Graham, M.R., et al., *Virulence control in group A Streptococcus by a two-component gene regulatory system: global expression profiling and in vivo infection modeling.* Proceedings of the National Academy of Sciences, 2002. **99**(21): p. 13855-13860.
217. Savic, D.J., W.M. McShan, and J.J. Ferretti, *Autonomous expression of the slo gene of the bicistronic nga-slo operon of Streptococcus pyogenes.* Infection and immunity, 2002. **70**(5): p. 2730-2733.
218. Lyon, W.R., C.M. Gibson, and M.G. Caparon, *A role for trigger factor and an rgg-like regulator in the transcription, secretion and processing of the cysteine proteinase of Streptococcus pyogenes.* The EMBO journal, 1998. **17**(21): p. 6263-6275.
219. Neely, M.N., et al., *Role of RopB in growth phase expression of the SpeB cysteine protease of Streptococcus pyogenes.* Journal of bacteriology, 2003. **185**(17): p. 5166-5174.
220. Dmitriev, A., E. McDowell, and K. Kappeler, *Chaussee MA, Rieck LD, Chaussee MS. The Rgg regulator of Streptococcus pyogenes influences utilization of nonglucose carbohydrates, prophage induction, and expression of the NAD-glycohydrolase virulence operon.* J Bacteriol, 2006. **188**(20): p. 7230-7241.
221. Kietzman, C.C. and M.G. Caparon, *CcpA and LacD. 1 affect temporal regulation of Streptococcus pyogenes virulence genes.* Infection and immunity, 2010. **78**(1): p. 241-252.
222. Shelburne, S.A., et al., *A combination of independent transcriptional regulators shapes bacterial virulence gene expression during infection.* PLoS pathogens, 2010. **6**(3): p. e1000817.
223. Anderson, N.G. and N.L. Anderson, *Twenty years of two-dimensional electrophoresis: past, present and future.* Electrophoresis, 1996. **17**(3): p. 443-453.
224. Callegari, E.A. and M.S. Chaussee, *The Streptococcal Proteome.* 2016.
225. Omenn, G.S., et al., *Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database.* Proteomics, 2005. **5**(13): p. 3226-3245.
226. Carr, S., et al., *The need for guidelines in publication of peptide and protein identification data working group on publication guidelines for peptide and protein identification data*. 2004, ASBMB.
227. Bradshaw, R.A., et al., *Reporting protein identification data the next generation of guidelines*. 2006, ASBMB.
228. Wilk, L., et al., *Comprehensive Mass Spectrometric Survey of Streptococcus pyogenes Subcellular Proteomes.* Journal of proteome research, 2017. **17**(1): p. 600-617.
229. Sharma, A., et al., *Identification of potential universal vaccine candidates against group A Streptococcus by using high throughput in silico and proteomics approach.* Journal of proteome research, 2012. **12**(1): p. 336-346.

230. Bensi, G., et al., *Multi high-throughput approach for highly selective identification of vaccine candidates: the Group A Streptococcus case.* Molecular & Cellular Proteomics, 2012: p. mcp. M111. 015693.
231. Lawrenson, R.A. and S. Sriskandan, *Cell Envelope Proteinase A (Streptococcus)*, in *Handbook of Proteolytic Enzymes (Third Edition)*. 2013, Elsevier. p. 3195-3202.
232. Zinkernagel, A.S., et al., *The IL-8 protease SpyCEP/ScpC of group A Streptococcus promotes resistance to neutrophil killing.* Cell host & microbe, 2008. **4**(2): p. 170-178.
233. Simpson, W., et al., *Coregulation of type 12 M protein and streptococcal C5a peptidase genes in group A streptococci: evidence for a virulence regulon controlled by the virR locus.* Journal of bacteriology, 1990. **172**(2): p. 696-700.
234. DeMaster, E., et al., *M+ group A streptococci are phagocytized and killed in whole blood by C5a-activated polymorphonuclear leukocytes.* Infection and immunity, 2002. **70**(1): p. 350-359.
235. Bisno, A.L., M. Brito, and C. Collins, *Molecular basis of group A streptococcal virulence.* The Lancet infectious diseases, 2003. **3**(4): p. 191-200.
236. Herwald, H., et al., *Streptococcal cysteine proteinase releases kinins: a virulence mechanism.* Journal of Experimental Medicine, 1996. **184**(2): p. 665-673.
237. Dale, J.B., et al., *Current approaches to group A streptococcal vaccine development.* 2016.
238. Liu, M., et al., *Active and passive immunizations with the streptococcal esterase Sse protect mice against subcutaneous infection with group A streptococci.* Infection and immunity, 2007. **75**(7): p. 3651-3657.
239. Zhu, H., et al., *The secreted esterase of group a streptococcus is important for invasive skin infection and dissemination in mice.* Infection and immunity, 2009. **77**(12): p. 5225-5232.
240. Sutcliffe, I.C. and D.J. Harrington, *Pattern searches for the identification of putative lipoprotein genes in Gram-positive bacterial genomes.* Microbiology, 2002. **148**(7): p. 2065-2077.
241. Lei, B., et al., *Identification of new candidate vaccine antigens made by Streptococcus pyogenes: purification and characterization of 16 putative extracellular lipoproteins.* The Journal of infectious diseases, 2004. **189**(1): p. 79-89.
242. Mortensen, R., et al., *Identifying protective Streptococcus pyogenes vaccine antigens recognized by both B and T cells in human adults and children.* Scientific reports, 2016. **6**: p. 22030.
243. Okamoto, A., et al., *Application of both high-performance liquid chromatography combined with tandem mass spectrometry shotgun and 2-D polyacrylamide gel electrophoresis for streptococcal exoproteins gave reliable proteomic data.* Microbiology and immunology, 2011. **55**(2): p. 84-94.
244. Hegde, P.S., I.R. White, and C. Debouck, *Interplay of transcriptomics and proteomics.* Current opinion in biotechnology, 2003. **14**(6): p. 647-651.

245. Maier, T., et al., *Quantification of mRNA and protein and integration with protein turnover in a bacterium.* Molecular systems biology, 2011. **7**(1): p. 511.
246. Chaussee, M.S., R.L. Cole, and J.P.v. Putten, *MOLECULAR AND CELLULAR PATHOGENESIS-Streptococcal Erythrogenic Toxin B Abrogates Fibronectin-Dependent Internalization of Streptococcus pyogenes by Cultured Mammalian Cells.* Infection and Immunity, 2000. **68**(6): p. 3226-3232.
247. Kapur, V., et al., *A conserved Streptococcus pyogenes extracellular cysteine protease cleaves human fibronectin and degrades vitronectin.* Microbial pathogenesis, 1993. **15**(5): p. 327-346.
248. Kreikemeyer, B., et al., *Genomic organization, structure, regulation and pathogenic role of pilus constituents in major pathogenic Streptococci and Enterococci.* International Journal of Medical Microbiology, 2011. **301**(3): p. 240-251.
249. Sylva, G.L., et al., *Rgg influences the expression of multiple regulatory loci to coregulate virulence factor expression in Streptococcus pyogenes.* Infection and immunity, 2002. **70**(2): p. 762-770.
250. Dmitriev, A.V., et al., *The Rgg regulator of Streptococcus pyogenes influences utilization of nonglucose carbohydrates, prophage induction, and expression of the NAD-glycohydrolase virulence operon.* Journal of bacteriology, 2006. **188**(20): p. 7230-7241.
251. Zhu, L., et al., *A molecular trigger for intercontinental epidemics of group A Streptococcus.* The Journal of clinical investigation, 2015. **125**(9): p. 3545-3559.
252. Asanuma, H., et al., *Isolation and characterization of mouse nasal-associated lymphoid tissue.* Journal of immunological methods, 1997. **202**(2): p. 123-131.
253. Koornstra, P., et al., *The Waldeyer ring equivalent in the rat: A model for analysis of oronasopharyngeal immune responses.* Acta oto-laryngologica, 1991. **111**(3): p. 591-599.
254. Park, H.-S., et al., *Membranous cells in nasal-associated lymphoid tissue: a portal of entry for the respiratory mucosal pathogen group A streptococcus.* The Journal of Immunology, 2003. **171**(5): p. 2532-2537.
255. Fan, X., et al., *Sortase A induces Th17-mediated and antibody-independent immunity to heterologous serotypes of group A streptococci.* PLoS One, 2014. **9**(9): p. e107638.
256. Hurdle, J.G., et al., *Targeting bacterial membrane function: an underexploited mechanism for treating persistent infections.* Nature Reviews Microbiology, 2011. **9**(1): p. 62.
257. Sanyahumbi, A.S., et al., *Global disease burden of group A Streptococcus.* 2016.
258. Control, C.f.D. and Prevention, *Antibiotic resistance threats in the United States, 2013. Atlanta: CDC; 2013*. 2014.