GLOBAL EMERGENCE AND PERSISTENCE OF HYPERVIRULENT CAPSULAR GROUP W SEQUENCE TYPE (ST)-11 *NEISSERIA MENINGITIDIS*

by

Mustapha Munir Mustapha

MBBS, Bayero University, Nigeria 2007

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2015

UNIVERSITY OF PITTSBURGH

Graduate School of Public Health

This dissertation was presented

by

Mustapha Munir Mustapha

It was defended on

September 28, 2015

and approved by

Dissertation Advisor:
Lee H. Harrison, MD
Professor
Departments of Epidemiology and Medicine
Graduate School of Public Health and School of Medicine
University of Pittsburgh

Committee Members:
Maria Mori Brooks, PhD
Associate Professor and Vice Chair of Education
Department of Epidemiology
Graduate School of Public Health
University of Pittsburgh

Jeffrey Lawrence, PhD
Professor
Department of Biological Sciences
Dietrich School of Arts and Sciences
University of Pittsburgh

N. Luisa Hiller, PhD
Assistant Professor
Department of Biological Sciences
Carnegie Mellon University
Pittsburgh, Pennsylvania

Lee H. Harrison, MD

GLOBAL EMERGENCE AND PERSISTENCE OF HYPERVIRULENT CAPSULAR GROUP W SEQUENCE TYPE (ST) - 11 *NEISSERIA MENINGITIDIS*

Mustapha M Mustapha, PhD

University of Pittsburgh, 2015

ABSTRACT

Neisseria meningitidis is a leading bacterial cause of sepsis and meningitis globally. Beginning with an epidemic among Hajj pilgrims in 2000, capsular group W (W) sequence type (ST) 11 emerged as a leading cause of epidemic meningitis in the African 'meningitis belt' and endemic cases in South America, Europe, Middle East and China. Previous genotyping studies were unable to reliably discriminate less virulent W ST-11 strains in circulation since 1970 from the Hajj epidemic strain (Hajj clone). It is also unclear what proportion of more recent W ST-11 disease clusters were caused by direct descendants of the Hajj clone. This work analyzes whole genome sequences of a global collection of over 250 meningococcal strains isolated from patients with invasive meningococcal disease globally from 1970 to 2014 using phylogenetic analyses, detailed examination of the capsule gene cluster (cps) and genes encoding major surface antigens.

We found that W ST-11 strains were descendants of an ancestral strain that had undergone unique capsular switching events. We identified two distinct, conserved, recombination events within W ST-11 cps genes with W ST-22 and Y ST-23 as most likely donor lineages. In addition, the Hajj clone and its descendants were distinct from other W ST-11 strains in that they

shared a common antigen gene profile and had undergone further recombination involving virulence genes encoding factor H binding protein (*fHbp*), nitric oxide reductase (*nor*), and nitrite reductase (*aniA*). Emergence of the Hajj clone may be related to recent acquisition of a distinct antigen-encoding gene profile and variations in meningococcal virulence genes.

This study resolves questions about the Hajj epidemic strain that were unanswered for 15 years. Furthermore, the findings of this study help illuminate genomic factors associated with emergence and evolution of virulent meningococcal strains.

Public Health significance: This dissertation provides genomic markers that reliably distinguish epidemic from sporadic W ST-11 strains that are applicable to molecular surveillance of *N. meningitidis*. Data presented in this work also demonstrate the need for a group W vaccine disease in the meningitis belt that can be potentially used beyond the meningitis belt in South Africa, parts of Latin America, and Europe that are facing the emergence of W ST-11.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

1.1 BACKGROUND

Neisseria meningitidis (meningococcus, Nm) is a leading cause of meningitis and sepsis[1] that is estimated to cause over 500,000 disease cases and 50,000 deaths worldwide. There is wide variation in incidence and rates and strain distribution between global regions and over time. *N. meningitidis,* a gram negative intracellular bacterium, is an exclusive human pathogen that is part of the normal pharyngeal flora of about 10% of the adult population. Occasionally, pathogenic meningococcal strains cross the mucosal barrier of the nasopharynx to cause invasive meningococcal disease (IMD). IMD most commonly manifests as bacteremia and/or meningitis.[2] Less common manifestations of IMD include pneumonia, septic arthritis or pericardial infection.[3] IMD is often a fatal disease; mortality rates range between 5-10% with survivors suffering from long term sequelae such as hearing loss and other neurologic deficits, limb amputation, and renal damage.[4] Transmission is through close contact with airway droplets of asymptomatically colonized individuals. Risk factors for invasive meningococcal disease are close contact with an IMD patient, cigarette smoking, kissing, immune complement deficiency and crowding in bars, dormitories or large human gatherings.[5]

Jafri et al[6] classified global meningococcal disease incidence pattern into countries with epidemic and/or high endemic disease with incidence rates >10 cases per 100,000 population that comprise most African countries, Uruguay, New Zealand and Mongolia. The highest incidence rates of 100-1000 per 100,000 are witnessed during occasional epidemics across 21 countries[7] in Africa collectively referred to as the 'meningitis belt' (Figure 1).[8] Moderate endemic rates are between 2-10 cases per 100,000 population and occur in South Africa, some European countries,

Brazil, Cuba and Australia. Remaining countries have low endemic rates (<2/100,000). IMD in some countries shows a bimodal incidence pattern with a large peak in infants and children less than five years old and a smaller peak among adolescents and young adults.

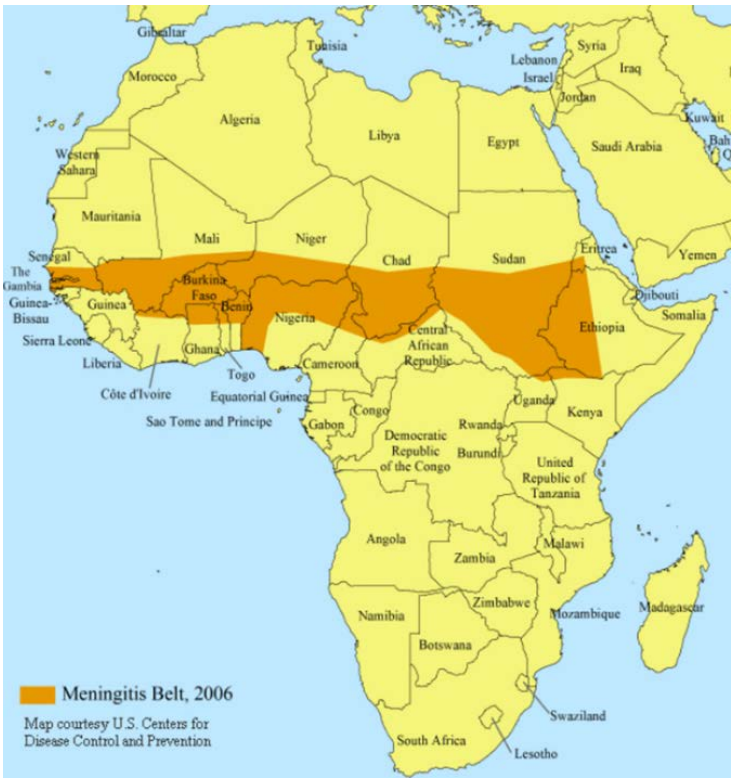

Figure 1: Meningitis belt, 2006

(Source: U.S. Centers for Disease Control and Prevention)

1.1.1 Molecular Characterization of *N. meningitidis*

Meningococci are classified into capsular groups (serogroups) based on serologic and genetic characteristics of their polysaccharide capsule, a major virulence determinant and the target of capsular group-specific vaccines. Of thirteen known capsular groups, six major groups A, B, C, W, X and Y account for almost all cases of invasive meningococcal disease.[9,10] The

epidemiologic pattern of meningococcal disease differs by capsular type over time. Group A is most common cause of epidemic meningitis; for example, an outbreak of group A meningococcal disease in 1996 in the meningitis belt resulted in over 250,000 cases and 25,000 deaths.[11] Capsular group B causes sporadic disease and small outbreaks in low to moderate endemic countries[12,13] but is very uncommon in the meningitis belt. Group C also causes sporadic disease and small case clusters globally,[14] although group C strains occasionally cause large epidemics in the meningitis belt.[15] In recent decades, groups W and Y have emerged to cause significantly higher number of meningococcal disease cases globally[16-22] while smaller group X[23,24] epidemics have occurred in the meningitis belt.

Multilocus sequence typing (MLST), which is based on sequencing DNA fragments of 7 housekeeping genes, is the method of choice for determining the genetic lineage of *N. meningitidis.*[25] Strains of the same sequence type (ST) share a common lineage and closely related STs are grouped under the same clonal complex (cc). A relatively small number of MLST lineages, termed either 'hypervirulent' or 'hyperinvasive' lineages cause a disproportionately greater percentage of IMD cases. There is a loose correlation between MLST lineage and capsular group that changes over time.

Major outer membrane proteins PorA and PorB are important virulence determinants and the basis of serologic classification of meningococci.[9] PorA genotype, which is based on peptide sequence variation within two variable regions, VR1 and VR2, combined with MLST is a very powerful tool for surveillance. *PorB* and *fetA* genotyping schemes are often used to further characterize invasive meningococcal strains.[5] Factor H binding protein (FHbp), along with other

surface proteins Nhba and NadA, is an antigen used in polysaccharide capsule-independent vaccines that have been developed for prevention of group B infection.[26]

In recent years whole genome sequencing (WGS) of meningococcal strains is expanding the scope of meningococcal strain typing by providing unprecedented level of detail on genetic diversity of meningococci. WGS studies have shed light on the genetic mechanisms underlying meningococcal strain emergence[27,28], virulence[29] and vaccine design.[30] WGS studies are facilitated by sites that make data publically available and provide comparative tools such as PubMLST (pubmlst.org/neisseria/).[31,32] PubMLST is especially well suited for comparison of newly sequenced meningococcal isolates to wide selection of historical strains.

1.1.2 Meningococcal Virulence

Meningococcal outer membrane contains several components that are important for virulence of which polysaccharide capsule is the most important. Of 13 known capsular groups, six (A, B, C, Y, W, X) account for nearly all cases of IMD. Also, meningococci lacking a capsule, though common among asymptomatic carriage strains, almost never cause invasive disease. The capsule is involved in resisting phagocytosis, opsonization and complement dependent lysis during human infection.[3] On the outer membrane, type IV pili are involved in meningococcal adhesion and together with opacity proteins Opa and Opc mediate translocation through human endothelium. Other adhesins such as Nhha, NadA, Nadr and App also mediate meningococcal attachment and translocation. Factor H binding protein, FHbp prevents human complement activation by binding to human factor H.[26] Other virulence factors are outer membrane proteins PorA, PorB, FetA, iron uptake proteins and endotoxin.[9]

N. meningitidis frequently undergoes genetic change through spontaneous mutation and horizontal transfer of genetic material from meningococcal or other bacterial species. Homologous recombination drives widespread virulence gene exchange among *Neisseria* species[29] causing substantial genetic variability even among meningococcal isolates of the same clonal complex.[33] Recombination involving major outer membrane antigen genes, also known as "antigenic shift", has been linked to increased incidence of meningococcal disease.[20] Capsular switching, which involves acquisition of novel capsule specific genes through recombination, also facilitates the emergence and persistence of virulent meningococcal lineages.[34-36] Through capsular switching, defined as presence of different capsular phenotypes within the same clonal complex, meningococcal strains belonging to virulent lineages may escape vaccine induced immunity. Surveillance studies have demonstrated that capsular switching is common among IMD isolates.[35,36]

1.1.3 Epidemiology of capsular group W *N. meningitidis*

N. meningitidis capsular group W (W) was first observed among military recruits in the United States in the late 1960s.[37] W was associated with sporadic IMD cases globally in the 1970s through the mid-1990s. Active surveillance in the United States identified W in 2-4% of invasive meningococcal isolates from 1992-1998.[38] Likewise in Canada, W caused 3-6% of annual reported IMD cases in 1997-1999.[39] In the meningitis belt, W was identified in 3.1% of 349 endemic meningococcal disease strains from Senegal and Niger in 1981-1982 and 7.3% of 41 strains from the Gambia.[40] In 1992-1995, an unexpectedly high proportion of W ST-11 (43%) was observed in 14 endemic case isolates in Gambia and Mali.[41] In Brazil, W represented <2% of all meningococcal disease isolates from 1990-2001.[42]

W emerged as an important cause of IMD following a large outbreak among Hajj pilgrims in Mecca, Saudi Arabia in 2000 which resulted in over 400 cases and 52 deaths were recorded among pilgrims and their close contacts across 16 different countries in Europe, Africa, Middle East, Southeast Asia and the United States.[16,43] In 2002, a quadrivalent polysaccharide vaccine that covers W capsular antigen was made a visa requirement for all pilgrims traveling to Saudi Arabia, where no subsequent W outbreaks have been reported since then.[44]

The Hajj outbreak strain was characterized as ST-11, PorA antigen gene type P1.5,2, with a distinct pulsed field gel electrophoresis (PFGE) pattern and 16S ribosomal RNA (16S) type 31.[45] Since 2000, several W epidemics have occurred in the meningitis belt and endemic clusters have been reported in South America, Europe, United States and China.[17,46-49] W ST-11 was equally as prevalent as group A IMD strains in Burkina Faso and Niger in 2001.[50] In 2002, the largest known W meningococcal outbreak was reported in Burkina Faso and caused an estimated 12,000 cases and more than 1400 deaths.[37] With the recent introduction of group A meningococcal conjugate vaccine into the meningitis belt, W ST-11 became the leading cause of IMD in that region.[7,17,51]

Outside the meningitis belt, South Africa witnessed a doubling of meningococcal disease incidence during 2000 to 2005.[52] High number of endemic W ST-11 cases were reported in Latin America,[19,42,48,53,54] UK[18] and Middle East.[55] Smaller clusters were reported in France,[56] United States[46] and China.[49]

Molecular Characteristics of the Hajj clone

Molecular epidemiologic studies conducted at the time of the Hajj epidemic revealed that the Hajj clone was found to be very similar to sporadic isolates obtained from Scotland and Canada

in 1970, USA (1996, 1998) Europe, and the meningitis belt. The Hajj clone was indistinguishable from sporadic W ST-11 strains by MLST, multi-locus DNA finger-print (MLDF), serologic typing and *porA* sequencing.[50,57,58] However, pre-Hajj isolates were distinguished by being predominantly 16S type 13 and 14.[45] The high genetic and antigenic similarity between the Hajj clone and pre-Hajj W ST-11 strains by traditional molecular techniques markedly contrasts with the abrupt and dramatic changes in the epidemiology of W ST-11 that began with the Hajj outbreak in 2000.[59] Emergence of W ST-11 epidemics and endemic cases beginning with the Hajj in 2000 raised a number of fundamental questions.[37,42,45,59,60]

1) W ST-11 strains were hypothesized to have arisen result of C to W capsular switching event(s). W ST-11 strains are genetically and phenotypically closely related to group C ST-11 strains which were historically the predominant capsular group within the ST-11 lineage.[59] However, the presence and extent recombination event(s) that caused C to W capsular switching were not established. It was also unknown whether the Hajj clone and other historical W ST-11 strains shared the same capsular switching event(s). On one hand, high genetic similarity between the Hajj clone and historical W ST-11 strains supported a common ancestral recombination but does not explain why the Hajj clone was much more virulent compared to historical W ST-11 strains.

2) A second question was whether there were genetic factors that could not be detected by traditional genotyping techniques either within or outside the capsule genes that could explain such dramatic increase in virulence.

3) It was also hypothesized that conditions associated with the Hajj such as large numbers of susceptible pilgrims from all over the world, overcrowding, and polysaccharide group A/C

vaccines that were a visa requirement for pilgrims at the time of the outbreak contributed to the emergence of the Hajj clone.[59]

1.2 HYPOTHESIS

The objective of this dissertation is to study the genetic factors associated with the change in epidemiologic behaviour of W ST-11 meningococcal strains, with particular reference to the W ST-11 strain that caused the global epidemic in 2000. Based on limited genotyping data, the prevailing model of evolution for the W ST-11 lineage was at the time based on a number of key assumptions (Figure 2). First was the assumption that the W ST-11 lineage arose from an ancestral group C ST-11 strain, although the exact recombination event(s) and direction of capsular switch were not established.[37,59] The second assumption was that a historical sporadic W ST-11 strain underwent genetic change that led to emergence of the Hajj clone.[5,45,59] Third was the assumption that all subsequent W ST-11 epidemics and endemic clusters were caused by global spread of the Hajj clone.[42,60,61] WGS provides excellent resolution of genetic content of bacterial strains that could illuminate the genetic factors responsible for the sudden emergence of W ST-11 in the year 2000 and beyond.

The primary research question of this dissertation is: *What are the dynamic genomic changes within W ST-11 from 1970-2014 that could be responsible for the abrupt change in W ST-11 epidemiology?* To address this question WGS from a large global collection of group W ST-11 strains were studied.

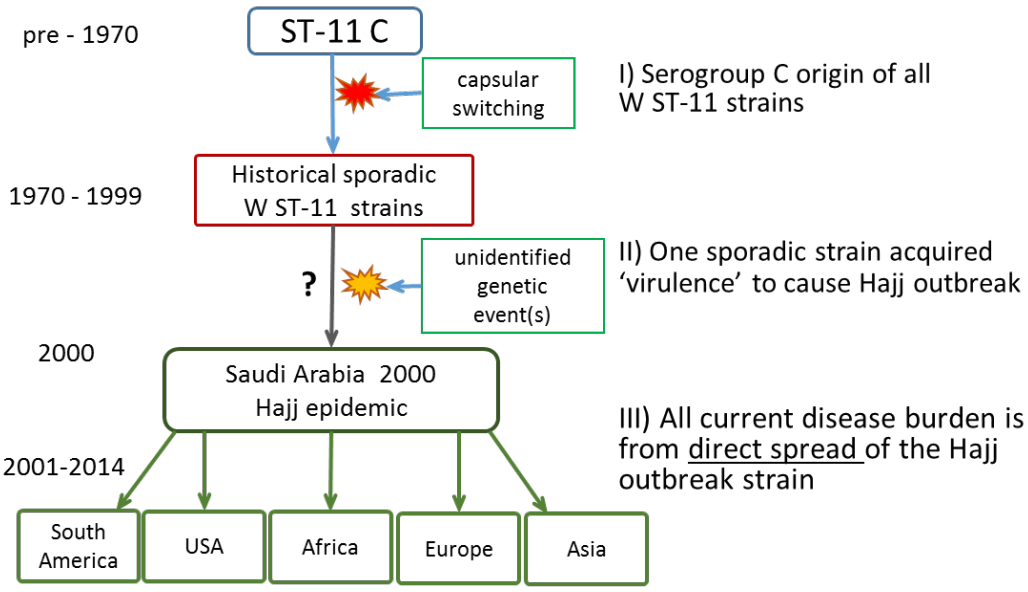

Figure 2: Traditional model of ST-11 W evolution

1.2.1 Specific aims

Specific aim 1: To assess phylogenetic relationships among the global collection of W and C ST-11 strains. We hypothesize that W ST-11 strains shared a close relationship representing clonal expansion within ST-11 lineage. We also hypothesize that more recent W ST-11 strains isolated following the Hajj 2000 epidemic (post Hajj strains) share even closer phylogenetic relationship to the Hajj clone and are phylogenetically distinct from historical W ST-11 isolates.

Specific aim 2: To characterize capsular genes and identify recombination event(s) within both the Hajj and pre-Hajj capsule genes. We hypothesize that the Hajj clone had a genetic differences within capsule genes as compared to pre-Hajj strains.

Specific aim 3: To determine whether the Hajj clone has a unique set of virulence genes not present in pre-Hajj strains. We hypothesize that the Hajj clone acquired a unique set of virulence

genes as compared to pre-Hajj strains and that these genes could account for increased virulence of the Hajj clone.

2.0 PAPER 1: GLOBAL EPIDEMIOLOGY OF CAPSULAR GROUP W MENINGOCOCCAL DISEASE, 1970-2014: EMERGENCE AND PERSISTENCE OF THE HYPERVIRULENT ST-11 LINEAGE (REVIEW)

Mustapha M Mustapha, MBBS[1] Jane W Marsh, PhD[1] and Lee H Harrison, MD[1*]

[1]Infectious Diseases Epidemiology Research Unit, University of Pittsburgh, A525 Crabtree Hall,130 Desoto Street, Pittsburgh, 15261PA, USA.

*Corresponding Author: lharriso@edc.pitt.edu.

Running title: Capsular Group W Meningococcal Disease

Key words: *Neisseria meningitidis***,** W135, meningitis surveillance, molecular epidemiology, Factor H binding protein FHbp.

2.1 ABSTRACT

Following an outbreak in Mecca Saudi Arabia in 2000, meningococcal strains expressing capsular group W (W) emerged as an important cause of invasive meningococcal disease (IMD) worldwide. The Saudi Arabian outbreak strain (Hajj clone) belonged to the ST-11 clonal complex similar to W ST-11 causing occasional sporadic disease before 2000. Since 2000, W ST-11 has caused large meningococcal disease epidemics in the African meningitis belt and endemic disease in parts of Latin America, Europe and China. Traditional molecular epidemiologic methods have been unable to discriminate epidemic W ST-11 from rare sporadic W ST-11 strains. Genetic diversity within the gene encoding factor H binding protein *fHbp*, helps to discriminate among W ST-11 strains. Further genetic characterization of global group W strains is required to elucidate the role of genetic factors in the emergence of the Hajj clone and to monitor continued evolution of other group W lineages.

2.2 BACKGROUND

Neisseria meningitidis (meningococcus), is an important cause of bacterial meningitis and sepsis worldwide that shows dynamic changes in strain distribution over time.[5] The highest incidence of endemic disease and periodic epidemics occur in 13 African countries of the 'meningitis belt.[5] The most important meningococcal virulence factor is the polysaccharide capsule, which is the target of capsular group specific vaccines.[10] Capsular groups A, B, C, W (formerly W-135), Y and X cause almost all invasive meningococcal disease (IMD). Other virulence factors include major outer membrane proteins (OMP) PorA, PorB, FetA, FHbp and lipooligosaccharide.[5] *N. meningitidis* undergoes genetic change through phase variation, homologous recombination and acquisition of genetic material from commensal *Neisseria* and/or other bacterial species.[33] These phenomena result in substantial genetic variability even among meningococcal isolates of the same lineage and facilitate emergence and persistence of virulent lineages.[62]

N. meningitidis capsular group W (W) was first observed among military recruits in the United States in the late 1960s. W emerged as an important cause of IMD following a large outbreak among Hajj pilgrims in Mecca, Saudi Arabia in 2000 which resulted in over 400 cases and 52 deaths; previously, W had not been known to cause outbreaks.[16] Since 2000 several W epidemics have occurred in the meningitis belt and endemic clusters have been reported in South America, Europe, United States and China.[17,46-49] Hajj outbreak strain was characterized as ST-11, PorA antigen gene type P1.5,2, with a distinct pulsed field gel electrophoresis (PFGE) pattern and 16S ribosomal RNA (16S) type 31.[45]

In this review W:P1.5,2:ST-11 (capsular group:PorAVR1,VR2:sequence type) strains with direct epidemiologic link to Hajj 2000-2001 outbreaks are referred to as "Hajj clone." The global

distribution and molecular characteristics of W from 1970-2014 is described with emphasis on the emergence of endemic and epidemic W ST-11 disease since 2000.

2.3 MOLECULAR CHARACTERISTICS OF GROUP W

Multilocus sequence typing (MLST), which is based on sequencing DNA fragments of 7 housekeeping genes, is the method of choice for determining the genetic lineage of *N. meningitidis*.[25] Strains of the same sequence type (ST) share a common lineage and closely related STs are grouped under the same clonal complex (cc). Historical sporadic W strains mostly belonged to ST-22, ST-174, and less often ST-11 clonal complexes. Following the Hajj 2000 epidemic, ST-11 became the predominant W lineage globally and the only lineage associated with epidemic disease while sporadic W cases caused by other clonal complexes have persisted at low levels globally.[30,42]

Outer membrane protein (OMP)genes (porA, porB, fetA): PorA and PorB are important virulence determinants and the basis of serologic classification of meningococci. The *porA* genotype based on sequence variation within two variable regions, VR1 and VR2, combined with MLST is a very powerful tool for surveillance. The *porB* and *fetA* genotyping schemes are often used to further characterize invasive meningococcal strains.[5] W ST-22 and ST-174 isolates have marked variability in their OMP allelic profiles consistent with non-clonal sporadic disease incidence. Conversely, the W ST-11 lineage has a remarkably stable outer membrane allelic profile P1.5,2:F1-1:2-2 (PorA VR1,VR2:FetA:PorB) despite global spread and different epidemiologic patterns over four decades.

Factor H binding protein (FHbp) is a surface protein that helps meningococci evade innate host immunity by binding human complement factor H and is also one of the antigens of

polysaccharide capsule-independent vaccines that have been developed for prevention of group B infection.[26] There is marked variability in the *fHbp* gene among W ST-11 strains globally. Endemic W ST-11 strains from Chile in 2012 and UK from 2010-2014 shared FHbp protein ID 22 (variant group 2); Hajj related W ST-11 isolates from the UK 2000-2002 had FHbp 9 (variant group 1), while FHbp 23 and 9 were most common in W from the meningitis belt.

Pulse field gel electrophoresis (PFGE) is often used in epidemiologic investigations to discriminate outbreak strains from genetically similar meningococcal strains unrelated to the outbreak. In a molecular epidemiologic study,[45] six of 19 (32%) sporadic pre-Hajj W ST-11 isolates were indistinguishable by PFGE from the Hajj clone with remaining 13 isolates having >85% PFGE similarity to the Hajj clone. Sporadic non-Hajj related W ST-11 from Younde, Cameroon in 1999-2000 also had only 1-2 band differences from Hajj clone by PFGE.[15] Thus, PFGE discriminates different W lineages but is less discriminatory for epidemic and sporadic W ST-11 strains. In spite of this limitation, PFGE has been utilized to describe W strains as likely descendants of the Hajj clone.

16S ribosomal RNA gene (16S) sequencing showed promise in discriminating the Hajj clone from sporadic W ST-11. The Hajj clone contained 16S type 31 while sporadic W ST-11 had 16S types 13 and 14.[42,45]

As a result of increased availability and reduced costs, *whole genome sequencing* (WGS) data are now publically available for several hundreds of genomes. In a genomic comparison of W ST-11 isolates collected from UK in 2010-2013 and earlier Hajj-related UK isolates from 2000-2002, over 200 allelic gene differences were identified.[3] This study illustrates how in-depth

WGS analyses combined with epidemiological data could potentially illuminate the genetic basis of W ST-11 lineage emergence and virulence.

"*Capsular switching*," refers to the expression of a different polysaccharide capsule through allelic exchange of capsular genes. This phenomenon has been linked with an increased incidence of meningococcal disease.[34,36] Surveillance studies have demonstrated that capsular switching is present in a substantial minority of invasive meningococcal disease isolates. [35,36] Before 2000, the ST-11 genetic lineage was most commonly associated with invasive group C strains.[59] This observation suggested that W ST-11 strains arose through C→W capsular switch, although the true direction of the genetic exchange is unknown.

2.4 GLOBAL INCIDENCE OF GROUP W 1970 – 1999

Capsular group W was associated with sporadic IMD cases globally in the 1970s through the mid-1990s. Active surveillance in the United States identified W in 2-4% of invasive meningococcal isolates from 1992-1998.[38] Likewise in Canada W caused 3-6% of annual reported IMD cases in 1997-1999. [39] In the meningitis belt, W ST-11 was identified in 3.1% of 349 endemic meningococcal disease strains from Senegal and Niger in 1981-1982 and 7.3% of 41 strains from the Gambia. In 1992-1995, an unexpectedly high proportion of W ST-11 (43%) was observed in 14 endemic case isolates in Gambia and Mali.[41] In Brazil, W represented <2% of all meningococcal disease isolates from 1990-2001.[42] Lemos et al[42] reported an increase in W ST-11 from 35% of 63 isolates in 1990-1995 to 80% of 56 isolates in 1996-1999 in Brazil. In the same period, a decline in W ST-174 occurred from 52% to 16% (Figure 3). All 67 W ST-11 strains from Brazil in 1990-1999 were indistinguishable from the Hajj clone by *porA* and *fetA* genotyping but differed from the Hajj clone by both16S and PFGE.[42] Strains genetically similar

to the Hajj clone were also identified in the 1990s in Taiwan,[63] UK, France[61] and Sweden.[58] In England and Wales, average annual W cases increased to 42 cases/year in 1996/97-1998/99 compared to 18 cases/year in 1993/94-1995/96.[64]

In summary, W primarily caused sporadic disease globally. However, some surveillance data from the African meningitis belt, UK, Brazil and Taiwan suggest an increase in W ST-11 in the mid- to late 1990s, predating the Hajj 2000 epidemic.

2.5 GROUP W INCIDENCE: 2000 AND BEYOND

2.5.1 Saudi Arabia (Hajj) 2000-2001

During the Hajj 2000 outbreak, over 400 cases were recorded among pilgrims and their close contacts across 16 different countries in Europe, Africa, Middle East, Southeast Asia and the United States (Figure 4).[43] In 2001, a similar outbreak among pilgrims from Saudi Arabia occurred with 106 meningococcal cases reported that were genotypically indistinguishable from the 2000 Hajj clone. In 2001, a quadrivalent polysaccharide vaccine that covers W capsular antigen was made a visa requirement for all pilgrims traveling to Saudi Arabia, where no subsequent W outbreaks have been reported.[44]

2.5.2 Middle East: endemic W following Hajj epidemic

Following the epidemics in Saudi Arabia, W has emerged as a leading cause of endemic meningococcal disease across the Middle East. While Saudi Arabia witnessed a decline in meningococcal disease from 338 cases in 2000 to 6 cases in 2009, reports from Qatar, Oman and United Arab Emirates indicated that W was the most commonly identified meningococcal group

in 2001-2009.[65] Epidemiologic data suggested that these endemic W cases represented spread and persistence of the Hajj clone.

Laboratory surveillance of IMD strains from Egypt identified W ST-11 in 4.5% of 67 IMD isolates from 1998-2003.[8] During 2005-2012 W ST-11 was identified as the leading cause of childhood acute bacterial meningitis in Turkey accounting for 38.1% of 333 confirmed cases.[32] These epidemiologic surveys illustrate the endemic nature of W ST-11 in the Middle East after the Hajj outbreaks.

2.5.3 Africa: emergence of W ST-11 epidemics

Cases of W ST-11 among Hajj pilgrims and their contacts were reported across Africa in Morocco, Sudan, Chad, Central African Republic, and Burkina Faso in 2000 and 2001.[7] In 2001, disease caused by W ST-11 was equally as common as group A in a non-representative sample of cases from outbreaks in Burkina Faso and Niger.[50] The following year (2002), the largest known W epidemic occurred in Burkina Faso, with over 13,000 cases and 1,400 deaths. More than 80% of cases were by strains indistinguishable from the Hajj clone. [66] In 2003-2008, a significant proportion of endemic W disease in the meningitis belt was caused by ST-174 clonal complex.[15,30] These data demonstrate the presence of other clonal W lineages in the meningitis belt of Africa post Hajj. Further W ST-11 epidemics were recorded across the meningitis belt in 2009-2014[4,7,17] making W the most commonly isolated capsular group across the meningitis belt in 2010-2014 (Figure 5).

The introduction of a highly effective group A polysaccharide-protein conjugate vaccine (MenAfriVac) in 2010 – 2014 led to a dramatic reduction in group A meningococci both in

disease cases and asymptomatic carriage[51] that could lead to elimination of group A epidemics within the region, which underscores the importance of the recent emergence of W disease. Epidemic W ST-11 strains from the meningitis belt were indistinguishable from the Hajj clone by PorA and FetA genotyping despite having 1-4 PFGE band differences. [60] However, these strains had diverse FHbp protein profiles [67]beyond the beyond the commonly tested OMP antigens.

South Africa witnessed a doubling of meningococcal disease due primarily to an increase in W which caused 5% of meningococcal cases in 2000 and 62% in 2005. The majority of these cases (93%) belonged to W ST-11 similar to the Hajj clone.[52] Similarly, meningococcal disease incidence attributed to W ST-11 increased in neighboring Mozambique from 2005-2008.[68] Again, it is not known whether endemic W ST-11 clusters in southern Africa represent global spread of the Hajj clone, regional spread of endemic strains from the meningitis belt or persistence of locally endemic strains.

2.5.4 Europe: emergence of endemic W ST-11 clusters and persistence of sporadic W ST-22

Cases of W ST-11 among returning pilgrims and their close contacts were reported across several European countries. [43] An overall increase in W disease occurred in France. Of 101 W cases in 2001– 2002, 45% (45/101) belonged to ST-11 and were indistinguishable from the Hajj clone by serotyping and PFGE.[61] Yet again, increase in W ST-11 in France could represent diversification of imported Hajj clone or expansion of locally endemic strains before the Hajj epidemic. In 2006-2008, W ST-11 cases declined and were replaced by ST-22.[3] A cluster of imported W ST-11 cases from the meningitis belt was reported in France during 2012[47] further providing evidence of the continued potential for international spread of W ST-11.

A similar increase in W cases related to the Hajj outbreak was seen in the England and Wales 2000-2002;[64] W cases returned to pre-2000 levels in 2003 – 2008 and then increased again in in 2009-2014, with 15% of all meningococcal disease cases in 2013-2014 being group W. Over 67% of all W cases in the UK 2010-2013 belonged to ST-11 and were indistinguishable from the Hajj clone by PorA and FetA while ST-22 persisted as a cause of sporadic disease.[3] Limited whole genome sequencing analysis revealed that recent endemic W ST-11 from the UK 2010-2013 had phylogenetic differences from earlier W ST-11 strains related to the Hajj outbreak. Strains from the UK and Ireland (2009-2014) shared gene alleles corresponding to FHbp 22 as compared to the FHbp 9 found in the Hajj clone.[3]

2.5.5 Latin America: Emergence of endemic W ST-11 clusters

Southern Brazil, Argentina and Chile witnessed the emergence of W ST-11 as a major cause of endemic meningococcal disease after 2003. Case clusters first appeared in southern Brazil where W represented 17.5% of all invasive strains in 2003-2005 compared to 6.2% in 1995-2002. [69] In Argentina, W cases increased from 7% in 2006 to 50% of all isolates in 2008-2011.[19] Similarly, the proportion of W ST-11 increased in Chile from 6.6% in 2010 to 58.3% in 2012 (Figure 6).[48] Similar to UK, W ST-11 from Chile also shared *fHbp* genotypes corresponding to protein ID 22.

2.5.6 USA and Canada

No increase in endemic W in the United States was observed despite presence Hajj related W ST-11 cases in 2000.[45] W made up 2.3% (48/1979) of all strains isolated in the United States during 2000-2010[13,35] and belonged predominantly to ST-22 clonal complex with only 2 cases of W ST-11. However, a cluster of 14 W ST-11 cases was reported in 2008-2009 in Florida.[46] Low

incidence rates of W disease (0.01-0.05 per 100,000) have persisted in Canada from 1990-2011 accounting for 5% of all IMD cases.[70]

2.5.7 Asia

In Singapore, Hajj related and endemic W ST-11 cases were reported in 2000-2002. W ST-11 cases have not been reported since 2003 suggesting disappearance of endemic W ST-11 in Singapore.[71] In Taiwan 2000-2001, W ST-11 increased to 34 cases/year compared to 4 cases/year in 1996-1999.[63] Endemic W ST-11 cases emerged in China (2011-2012) where 24.4% (11/45) confirmed IMD in 2011 -2012) cases were W ST-11.[49] Yet again, it is unclear whether W ST-11 in Taiwan and China represent expansion of local endemic strains or continued global spread of the Hajj clone.

2.5.8 Australia and New Zealand

In Australia, W caused 1.5-5% of annual IMD cases in 1997-2012. Australian W isolates were of diverse PorA genotypes including a few with P1.5,2 genotype similar to that found in the Hajj.[72] Disease caused by W continued to be rare in New Zealand, accounting for 3.5% of 700 reported IMD cases in 2008-2013.[12] One case of W:P1.5,2:ST-11 was reported in New Zealand[36] while a cluster of W:P1.7-2,4:ST-11 cases genetically distinct from Hajj clone was reported in 2003. This strain was shown to have arisen through capsular switching between locally endemic group C and a carriage group W strain.

2.6 GEOGRAPHIC AND GENETIC ORIGINS OF THE HAJJ CLONE

The Hajj clone represents clonal expansion of a historical capsular switch strain in circulation at least since the 1970s.[45] Environmental factors unique to the Hajj pilgrimage may have facilitated emergence of the W ST11 Hajj clone. [59] The Hajj, an annual gathering of over 2 million people

from around the world is an ideal setting for the selection and rapid global dissemination of virulent meningococci. For example, in 1987, a Hajj-related meningococcal outbreak facilitated the spread of virulent group A ST-5 meningococci from south Asia to sub-Saharan Africa.[44] It is hypothesized that the polysaccharide group A/C vaccine requirement for Hajj pilgrims in the 1990s lead to selection of W capsular switch due to vaccine-induced immunologic pressure. The W ST-11 Hajj clone and its descendants may have subtle genetic changes that cannot be detected by traditional molecular methods. Even small genetic changes in key virulence genes can lead to significant increase in transmission efficiency and virulence potential.[62] Point mutations and/or recombination event(s) can differentiate the W ST-11 Hajj clone from endemic W ST-11. Detailed characterization of a global collection of meningococcal strains using whole genome sequencing is needed to discern genetic and epidemiologic relatedness between W ST-11 strains. Limited surveillance of meningococcal disease and carriage strains in many regions in the decade before the Hajj epidemic makes accurate assessment of geographic origins of the Hajj clone difficult.

Although epidemic W ST-11 IMD began with the 2000 Hajj outbreak, surveillance data showed an upward trend in W ST-11 sporadic cases in the mid to late 1990s in the meningitis belt,[41,57] Brazil,[42] UK, and Taiwan.[63] Also, the southern cone of Latin America, one of the regions facing recent emergence of W ST-11 witnessed no Hajj or other travel related W ST-11 cases making it more likely that W ST-11 emergence in the region represented regional expansion of rare sporadic W ST-11 strains. These studies suggest multifocal emergence of W ST-11 clusters some of which may be unrelated to the Hajj.[60]

2.7 CONCLUSION

W has emerged during the past 15 years as an important global cause of meningococcal disease. The recent success with the control of MenA disease in the meningitis belt has resulted in the fact that W is now the predominant cause of meningococcal disease in the region. This, in combination with the emergence of disease caused by capsular groups X and C, underscore the need for polyvalent vaccines for the meningitis belt. Meningococcal disease surveillance on a global scale and broader application of whole genome characterization will improve our understanding of the factors contributing to the continued spread of the Hajj clone and the evolution of other W lineages.

Funding: None

Conflicts of Interest: All authors have no conflicts of interest.

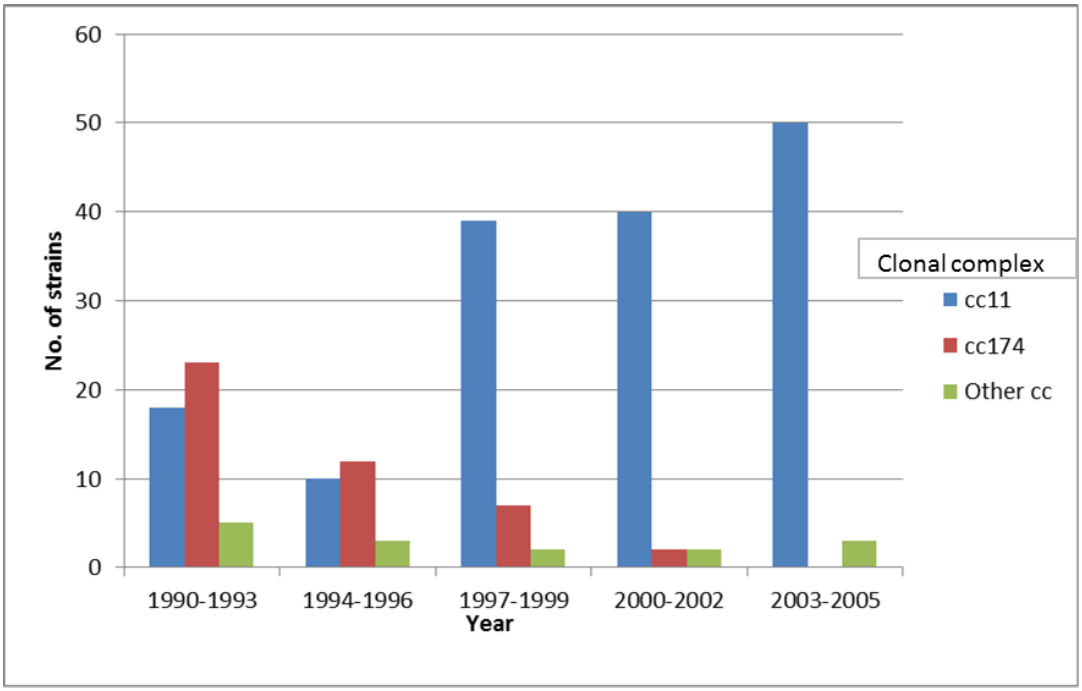

Figure 3: Molecular typing of invasive capsular group W *N. meningitidis* in Brazil, 1990-2005. Based on data from Lemos et al[42]

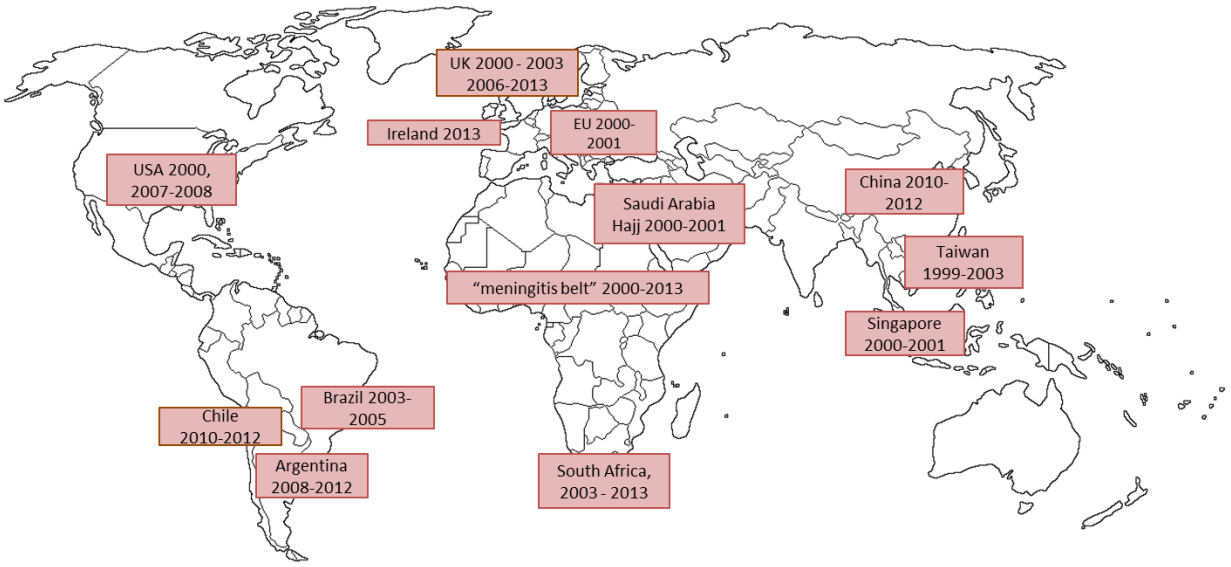

Figure 4: Global spread of capsular group W ST-11 meningococcal disease case clusters and epidemics, 2000-2014.

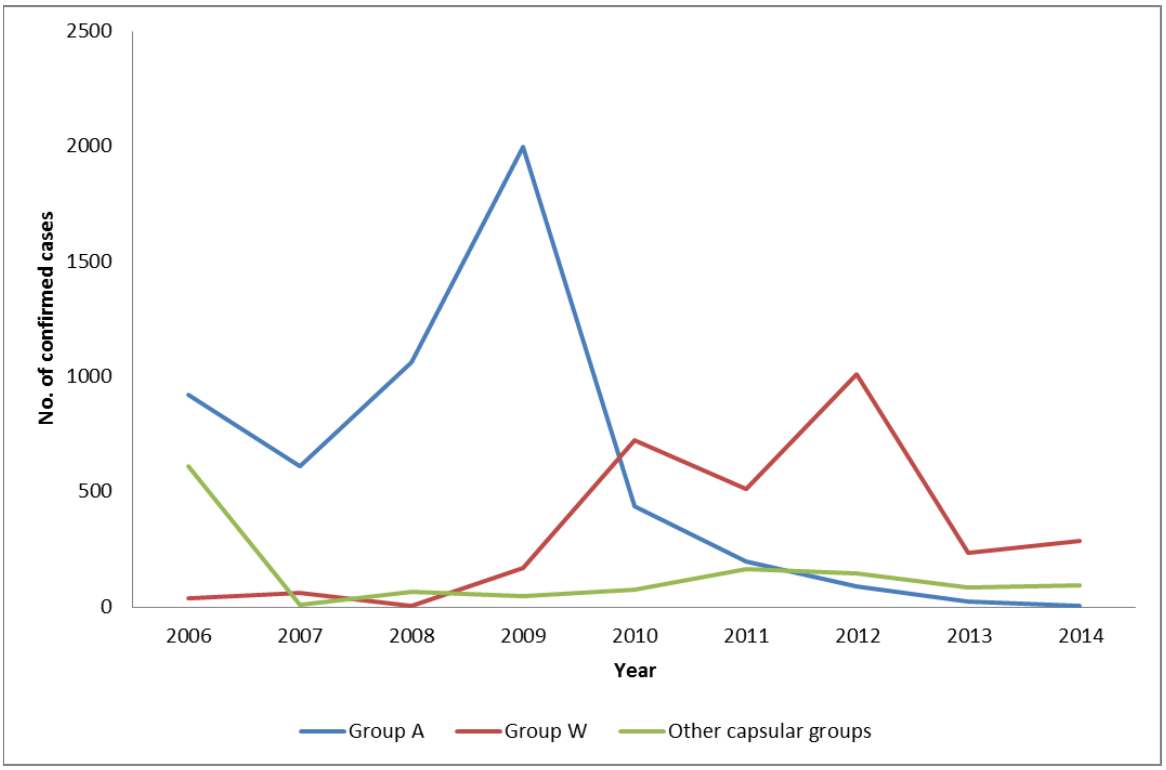

Figure 5: Distribution of capsular groups among confirmed invasive meningococcal disease cases, meningitis belt 2006-2014.

(source: World Health Organization, www.who.int/csr/disease/meningococcal/en/).

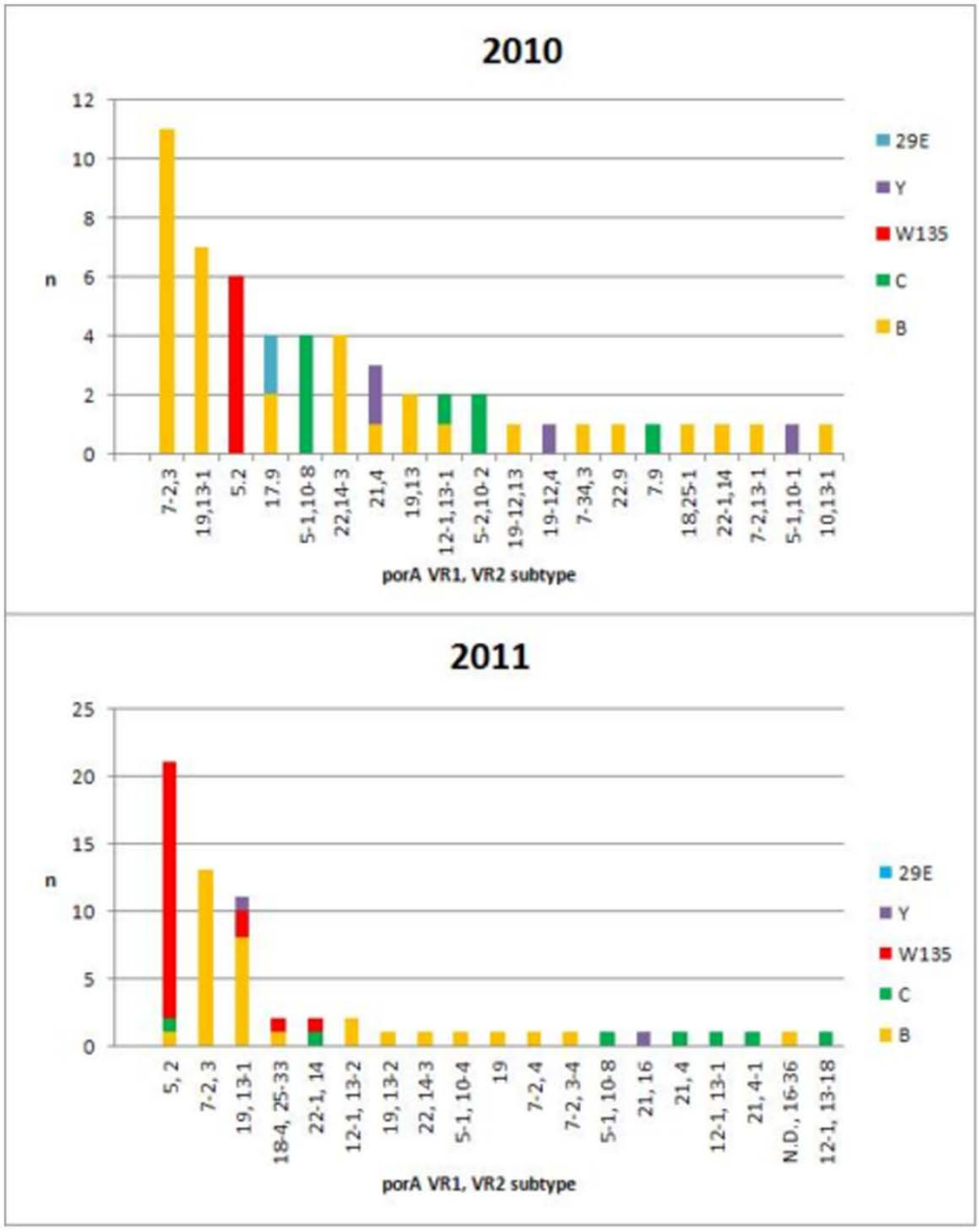

Figure 6: Molecular characterization of invasive meningococcal strains, Chile 2010-2011.[54]

3.0 PAPER 2: GENOMIC EPIDEMIOLOGY OF HYPERVIRULENT SEROGROUP W, ST-11 *NEISSERIA MENINGITIDIS*

Authors: Mustapha M. Mustapha[1*], Jane W. Marsh[1], Mary G. Krauland[8], Jorge O. Fernandez[2], Ana Paula S. de Lemos[3], Julie C. Dunning Hotopp[4], Xin Wang[5], Leonard W. Mayer[5], Jeffrey G. Lawrence[6], N. Luisa Hiller[7], Lee H. Harrison[1]

*Corresponding author (email: mmm147@pitt.edu)

[1]Infectious Diseases Epidemiology Research Unit, University of Pittsburgh, Pittsburgh, PA 15261, USA.

[2]Molecular Genetics Laboratory, Public Health Institute of Chile, Santiago, Chile.

[3]Department of Bacteriology, Instituto Adolfo Lutz (IAL), São Paulo, Brazil.

[4]The Institute for Genome Sciences, University of Maryland School of Medicine, University of Maryland, Baltimore, Maryland, USA.

[5]Meningitis and Vaccine Preventable Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA.[+]

[6]Department of Biological Sciences, University of Pittsburgh, USA.

[7]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, USA.

[8]Public Health Dynamics Laboratory, Graduate School of Public Health, University of Pittsburgh, USA.

[+]*The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.*

3.1 ABSTRACT

Neisseria meningitidis is a leading bacterial cause of sepsis and meningitis globally with dynamic strain distribution over time. Beginning with an epidemic among Hajj pilgrims in 2000, serogroup W (W) sequence type (ST) 11 emerged as a leading cause of epidemic meningitis in the African 'meningitis belt' and endemic cases in South America, Europe, Middle East and China. Previous genotyping studies were unable to reliably discriminate sporadic W ST-11 strains in circulation since 1970 from the Hajj outbreak strain (Hajj clone). It is also unclear what proportion of more recent W ST-11 disease clusters are caused by direct descendants of the Hajj clone. Whole genome sequences of 270 meningococcal strains isolated from patients with invasive meningococcal disease globally from 1970 to 2013 were compared using whole genome phylogenetic and major antigen-encoding gene sequence analyses. We found that all W ST-11 strains were descendants of an ancestral strain that had undergone unique capsular switching events. The Hajj clone and its descendants were distinct from other W ST-11 strains in that they shared a common antigen gene profile and had undergone recombination involving virulence genes encoding factor H binding protein, nitric oxide reductase, and nitrite reductase. These data demonstrate that recent acquisition of a distinct antigen-encoding gene profile and variations in meningococcal virulence genes were associated with the emergence of the Hajj clone. Importantly, W ST-11 strains unrelated to the Hajj outbreak contribute a significant proportion of W ST-11 cases globally. This study helps illuminate genomic factors associated with meningococcal strain emergence and evolution.

Keywords: Hajj clone, invasive meningococcal disease, W135, whole genome sequencing, virulence factors, FHbp

3.2 RESEARCH IN CONTEXT

Neisseria meningitidis, a bacterial cause of frequently fatal brain (meningitis) and blood stream (sepsis) infections, has variable strain distribution over time. Serogroup W sequence type 11 (W ST-11) lineage is associated on one hand with strains causing only rare (sporadic) disease cases, and the Hajj clone – a major global cause of epidemic and endemic meningococcal disease. In this study we analyzed complete genome sequences of a global collection of 270 W ST-11 isolates causing meningococcal disease from 1970-2013. The Hajj clone acquired novel gene sequences within genes involved in nitrogen metabolism (nitrogen oxide reductase, nitrite reductase) and evasion of human immune response (factor H binding protein).These genes may be the cause of increased virulence of the Hajj clone and can be used to trace continuing spread of the clone. These results shed light on mechanisms of meningococcal strain emergence.

3.3 INTRODUCTION

Neisseria meningitidis is a major global cause of meningitis and sepsis with large variations in disease incidence rates and strain distribution globally.[1] Incidence rates range from 0.5-15 cases per 100,000 population across most global regions. Very high incidence rates of 100-1000 per 100,000 are witnessed during occasional epidemics across 21 countries[7] in Africa collectively referred to as the 'meningitis belt.'[8] Meningococci are classified into serogroups based on biochemical properties of their polysaccharide capsule – the primary determinant of meningococcal virulence and major vaccine target. Serogroups A, B, C, W (formerly W-135) and Y cause almost all invasive disease cases. Other virulence determinants are lipooligosaccharide and several outer membrane proteins. Multilocus sequence typing (MLST),[25] based on DNA sequence of seven housekeeping genes, is used to classify meningococci into lineages (sequence types, ST). Closely related STs are termed 'clonal complex.'

Outer membrane proteins, porins A and B (PorA, PorB) and iron-regulated enterobactin (FetA) are used for "fine typing" of meningococcal isolates while factor H binding protein (FHbp), *Neisserial* adhesion A (NadA) and *Neisserial* heparin binding antigen (Nhba) are components of capsule-independent vaccines developed for prevention of serogroup B disease. Meningococci have a very plastic genome as a result of frequent acquisition of genetic material from other *Neisseria* or more distant bacterial species through recombination. Recombination involving major outer membrane antigen genes - "antigenic shift" has been linked to increased incidence of meningococcal disease. Capsular switching – acquisition of novel capsule specific genes through recombination has also facilitates the emergence and persistence of virulent meningococcal lineages. Through capsular switching, defined as presence of different capsular phenotypes within the same clonal complex, meningococcal strains belonging to virulent lineages may escape vaccine induced immunity.

In the 1970s to 1990s, serogroup W was a rare cause of meningococcal disease. In 2000, the first recorded serogroup W meningococcal disease outbreak occurred during the annual Hajj pilgrimage in Mecca, Saudi Arabia.[16,43] The Hajj outbreak strain, referred to as the Hajj clone, was characterized as belonging to the hypervirulent sequence type (ST)-11 genetic lineage and having the PorA antigen-encoding gene type P1.5,2.[45] Since 2000, W ST-11 strains that are genetically similar to the Hajj clone have caused large epidemics in the African meningitis belt[17,66,73] and have caused case clusters and smaller outbreaks in South Africa,[52] China,[49] Taiwan,[63] Brazil,[42] and most recently in Argentina,[53] Chile[54] and the United Kingdom.[3]

In the 1960s-1980s more than 90% of ST-11 strains expressed serogroup C capsule.[59] As a result, W ST-11 is thought to have emerged from serogroup C ST-11 lineage through capsular switching though the direction of capsular switching has not been established.[59] The Hajj clone

was highly similar by most molecular typing techniques including pulsed field gel electrophoresis (PFGE), PorA and FetA genotyping to historical rare sporadic group W ST-11 strains isolated globally from 1970-1999.[45,57,61] These "pre-Hajj" strains were distinguished by 16S ribosomal RNA (rRNA) genotyping as 16S type 13 and type 14 compared to type 31 associated with the Hajj clone.[45] The high genetic and antigenic similarity between the Hajj clone and pre-Hajj W ST-11 strains by traditional molecular techniques markedly contrasts with the abrupt and dramatic changes in the epidemiology of W ST-11 that began with the Hajj outbreak in 2000.[59] There is a need for molecular surveillance laboratories to reliably identify epidemic from locally endemic sporadic strains.

In a recent study, we compared capsular gene sequences from the Hajj clone and 24 other invasive W ST-11 strains isolated from 1970-2012.[74] Phylogenetic and BLAST analyses against a database of over 2300 genome sequences demonstrated that the Hajj clone and 24 other invasive W ST-11 strains shared identical capsular recombination events, with a sporadic group W strain and Y ST-23 as the most likely donor lineages into a serogroup C ST-11 strain.[74] Taken together, these studies suggest that historical sporadic W ST-11 strains and the Hajj clone have a common origin and likely emerged from a capsular switching event in a single ancestral C ST-11 strain. Furthermore, these studies suggested that genetic differences outside the capsule region may be responsible for increased virulence seen in the Hajj clone when compared to historical sporadic W ST-11 strains.

Whole genome sequencing (WGS) could potentially illuminate genetic differences not detectable by older genotyping techniques that could account for the observed epidemiologic differences between W ST-11 strains. We analyzed whole genome sequencing (WGS) data from invasive serogroup W ST-11 strains isolated globally from 1970-2013 to generate a model of W ST-11

evolution and global spread. We postulated that subtle genetic variations through recombination and/or mutation, outside of the capsule gene cluster may distinguish these W ST-11 strains and could explain the dramatic emergence and increased virulence of the Hajj clone. We identified genomic markers that are distinct from the Hajj clone, traced continued global spread of the Hajj clone, and interestingly, found that some of the current W ST-11 case clusters are caused by W ST-11 strains that are not direct descendants of the Hajj clone.

3.4 METHODS

Strain selection: We analyzed 270 *Neisseria meningitidis* serogroup W ST-11 genomes isolated from disease cases over four decades (1970-2013). Twenty-six genomes were newly sequenced while 244 genome sequences were identified from public databases (Table 2). All 26 newly sequenced genomes including 14 genomes that were part of the Genomic Sequencing Center for Infectious Disease (GSCID, gscid.igs.umaryland.edu/) project are made publically available in both PubMLST (pubmlst.org/neisseria/) and GenBank databases. In the PubMLST database, 273 genome sequences were designated serogroup W ST-11 by the end of June 2014, of which 31 genomes were excluded because they had missing information in at least one of six antigen gene alleles. Two genomes were obtained from *Neisseria* Base (nbase.biology.gatech.edu/).[31,75]

Strain classification: W ST-11 invasive meningococcal strains were classified as isolated before, during or after the Hajj 2000 outbreak (Figure 7). Pre-Hajj strains (n = 13) were isolated globally from 1970-1999, the Hajj clone (n = 1) (Strain ID: *M7124*) was isolated in Saudi Arabia during the Hajj outbreak in 2000[45], post-Hajj (n = 256) were isolated during 2000 – 2013. All 270 whole genome sequences were characterized using antigen-encoding gene profiles and presence of genomic regions associated with allelic exchange. Forty-eight isolates representative of the

antigen-encoding gene and geographic diversity of all 270 strains were selected for whole genome phylogenetic analyses.

Whole genome sequencing, assembly and genome annotation: A total of 26 W ST-11 isolates were newly sequenced using Pacific Biosciences RS II (PacBio, www.pacificbiosciences.com) Single Molecule Real Time (SMRT) sequencing (n = 7), paired-end Illumina HiSeq only (Illumina, www.illumina.com, n = 10), paired-end Illumina and Roche 454 (454.com/) pyrosequencing (Illumina+454, n = 3) and Life Technologies (www.lifetechnologies.com) Ion Torrent PGM sequencing (n = 9) platforms. Library preparation and sequencing were done according to manufacturer protocols. Four out of 7 isolates sequenced using PacBio had previously been sequenced using Illumina+454 (*M7124*) or Ion Torrent (*NM3682, NM3686, NM3687*). *M7124* was re-sequenced using PacBio to provide very high quality Hajj reference genome sequence.

De novo assembly was done using Hierarchical Genome Assembly Process (HGAP) v4.0 for PacBio, Celera v7.0 for Illumina/454 and Mira 3.0 for Ion Torrent with default settings and assembly qualities were checked as described in Supplementary Methods.[76,77] Contiguous genomic DNA sequences (contigs) from *de novo* assemblies were annotated on RAST and IGS annotation servers.[78] Assembled contig files were downloaded from public databases for the remaining 244 genome sequences.

Whole genome alignment and phylogenetic analyses: Forty-eight representative W ST-11 strains underwent whole genome phylogenetic analyses. Core genome alignment of 1,014,185 sites was generated using Mauve v2.3[79] and a Maximum-Likelihood phylogenetic tree was constructed using general time reversible model with invariant sites (GTR+Γ+I) with 100 bootstrap replicates

using PhyML v3.0.[80] The effect of recombination on phylogenetic relationships was assessed using SplitsTree v4[81] and ClonalFrame v1.2.[82] ClonalFrame is a phylogenetic reconstruction method that determines phylogenetic relationships between bacterial strains after detecting and accounting for recombinant sequences based on a coalescent model of evolution.

Antigen gene allele designations for full length *porA, porB, fetA, nadA, nhbA* and *fHbp* genes were obtained by comparing assembled contigs to reference alleles downloaded from the *Neisseria* PubMLST database (pubmlst.org/).[31] Phylogenetic trees of catenated, aligned antigen-encoding gene sequences were constructed using ClonalFrame and phylogenetic trees were visualized on CLC Genomics workbench v7 (www.clcbio.com) and MEGA v5.2.[83]

Gene gain/loss: Annotated gene sequences for the Hajj reference strain, *M7124*, were compared to W ST-11 strains using BLASTn with a cut-off of 70% alignment coverage and 70% identity. Universally present genes were aligned using MAFFT[84] v7 and Maximum-Likelihood tree was constructed as described above.

Single nucleotide polymorphism (SNP) analysis: Raw sequence reads of 23 W ST-11 isolates for which we have short read sequence data were aligned to the Hajj reference genome *M7124* using BWA v0.6[85]. Aligned reads were indexed, sorted and filtered using VCFtools v0.1 with default options. All insertions and deletions (indels), diploid SNPs and SNPs with phred-like quality score, Q < 30 were excluded. Hajj specific *SNPs* were defined as those SNPs found in all Hajj cluster strains, but not non Hajj-cluster strains. Hajj cluster strains were defined as all strains phylogenetically and antigenically very closely related to the Hajj clone. We determined which discriminatory SNPs were acquired by recombination as opposed to spontaneous mutation by phylogenetic and BLASTN analyses of genomic sequences adjacent each Hajj specific SNP.

Bioinformatics analyses were done on BioLinux 7 server[86] and Windows 7 computing environments.

Multilocus sequence typing (MLST) and outer membrane protein (OMP) gene sequencing: MLST and OMP sequence typing of the *porA* VR1, VR2, and *fetA* VR gene fragments were performed as described in Supplementary Methods.

16S rRNA gene sequencing: PCR amplification and sequence analysis of the 16S rRNA genes were performed using modifications of published methods[42,87] as described in Supplementary Methods.

Funding: This study was supported by a career development award to Dr. Harrison, NIAID (K24 AI52788) and by a grant from the University of Pittsburgh Department of Epidemiology Small Grants Program. This project has also been funded in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract number HHSN272200900009C.

Role of funding source: Funding sources played no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication

3.5 RESULTS

The Hajj clone (strain ID: *M7124*), and six other W ST-11 strains that were sequenced using PacBio provided very high quality reference sequences (Table 2), with excellent resolution of capsular genes and other highly repetitive genomic regions.

W ST-11 strains in this study were closely related to serogroup C ST-11 strains and clustered into two main groups based on genetic relatedness to the Hajj clone. Out of 270 total strains, 125 (46.3%) were most closely related to the Hajj clone based on antigen-encoding gene profiles (Figure 8), presence of recombinant alleles (Table 1, Figures 20-23) and whole genome phylogenetic analyses (Figures 9 and 24). We collectively refer to these strains as Cluster 1 (Hajj cluster). All strains in this cluster were isolated during or after the Hajj 2000 epidemic (post-Hajj strains). Remaining 145 strains were more heterogeneous having antigen-encoding gene, phylogenetic and SNP differences within key virulence genes compared to *M7124* and were collectively referred to as Cluster 2 strains. This cluster included both historical W ST-11 strains isolated before 2000 (pre-Hajj strains) and post-Hajj strains (2000-2013).

3.5.1 Phylogenetic analyses

Phylogenetic trees constructed using three different methods, Maximum Likelihood, SplitsTree and ClonalFrame consistently discriminated Cluster 1 from Cluster 2. Phylogenetic analyses of aligned whole genome sequences (Figures 9 and 24) and concatenated antigen-encoding gene sequences (Figure 8) demonstrated that several strains (Cluster 1/Hajj cluster) were nearly identical to *M7124* suggesting that these strains were direct descendants of the Hajj clone. The other strains (Cluster 2) were phylogenetically diverse and included a combination of recent and historical strains from the 1970s and 1980s that were most closely related to serogroup C ST-11 strains. SplitsTree phylogenetic network constructed from aligned core genome sequences support a clonal phylogenetic relationship between Cluster 1 strains (Figure 9). Also, a comparison of SNPs in 23 W ST-11 strains relative to *M7124* showed that Cluster 1 strains had fewer SNP differences than Cluster 2 strains, which collectively support indicate that Cluster 1 (Hajj cluster) strains represent clonal expansion within the W ST-11 lineage. Furthermore, these

results demonstrate that many recent W ST-11 case clusters are more closely related to pre-Hajj sporadic strains and are unlikely to be direct descendants of the Hajj clone.

Genomic differences between the Cluster 1 and Cluster 2 strains were further explored by mapping the genomic locations of Hajj specific SNPs among 23 selected strains. Of 48 Hajj specific SNPs, 46 were located in four genomic regions that ranged in size from 1·2-4·3 kb. Phylogenetic and BLAST analyses suggest these genomic regions were acquired through homologous recombination (allelic exchange) [Figures 20-23]. Two of these recombinant regions encode known meningococcal antigens and/or virulence proteins – factor H binding protein (*fHbp*), nitric oxide reductase (*nor*), and nitrite reductase (*aniA*).

To explore the presence of these identified recombinant regions in all 270 strains in our collection, we queried nucleotide sequences from these recombinant regions in *M7124* against assembled contigs from all 270 genomes using BLASTN with 99% sequence identity and coverage. All four recombinant regions present in the Hajj clone were also found in 95·1% (118/124) of Cluster 1 strains. Six other Cluster 1 strains from South Africa 2005-2013 (*21583, 29326, 29336, 29387, 29402* and *29393*) had allelic profiles within one of four recombinant regions that were different from Hajj clone alleles. No Cluster 2 strain had a sequence closely matching any of the four recombinant regions associated with the Hajj clone. These results demonstrate that these genomic regions represented areas with allelic replacement through recombination likely involving donor sequences from meningococci outside ST-11 lineage and commensal *Neisseria* species.

3.5.2 Antigen-encoding genes

Cluster 1 (Hajj cluster): The Hajj clone and 43% (116/270) of all W ST-11 strains in this study had an identical antigen-encoding gene profile: *porA* 1, *porB* 1, *fetA* 13, *nadA* 5, *nhba* 72 and *fHbp* 9. Strains with this profile were isolated from the meningitis belt in 2000 – 2005, Mauritius (2001), South Africa (2003 – 2013), United States (2000) and the UK (2000 – 2004, 2007, 2011). The Hajj clone antigen-encoding gene profile is characterized by a unique *fHbp* allele 9 (FHbp peptide 9, variant group 1/sub-family B). This *fHbp* allele was present only in Cluster 1 strains and not in sporadic pre-Hajj W ST-11 strains. *fHbp* allele 9 was most likely acquired through allelic exchange within one of the four recombinant regions identified earlier (Figures 20-23). Nine other strains all isolated from South Africa in 2010-2013, differed only at the *fetA* gene compared to the Hajj clone, these strains were classified as Cluster 1 strains based on whole genome phylogenetic analysis and shared recombinant alleles with the Hajj clone.

Cluster 2: Antigen-encoding gene profiles (Figure 8) were more heterogeneous within Cluster 2 with two different alleles each for *nadA* and *nhba* and 8-11 different alleles for *porA, porB, fetA,* and *fHbp* genes. There were 33 different allelic combinations within Cluster 2 with 67.6% (98/145) having one of three predominant *porA/porB/fetA/nadA/nhba/fHbp* allelic profiles: 1/244/13/5/17/22, 1/311/13/5/17/160, and 1/1/13/3/17/22. None of the pre-Hajj strains shared an identical antigen-encoding gene profile with the Hajj clone. The *fHbp* locus was the most divergent between Cluster 1 and Cluster 2. The most common *fHbp* allele in Cluster 2 strains was allele 22 belonging to variant 2/family A present in 80.2% (117/145). None of the Cluster 2 strains had the Hajj clone *fHbp* gene allele. The remaining five antigen-encoding genes – *porA, porB, fetA, nadA* and *nhba* genes–were more closely related to the Hajj clone alleles.

Strains in Cluster 2 were also geographically and temporally diverse, with W ST-11 strains isolated from the UK from 1970-1975 and Netherlands in 1985 being antigenically most closely related to the serogroup C ST-11 reference strain *FAM18* (Figure 2, 3A-C, Supplementary Figure 3A-B). Five of nine strains from the meningitis belt isolated in 2000-2005 and 72.4% of 98 post-Hajj strains from South Africa (2003 – 2013) had the Hajj related antigen-encoding gene profile. In contrast, only one of three strains from the USA (2000, 2008-2009) and none of the five post-Hajj strains from Brazil and Chile had the Hajj related antigen-encoding gene profile. In the UK, strains with identical antigen-encoding gene profile to the Hajj clone predominated in 2000 – 2004 (97.8% of 45 strains) but were uncommon during 2005 – 2013 (3.3% of 90 strains). These data are consistent with antigenic diversification of W ST-11 strains presumably following an ancestral C to W ST-11 capsular switch, and subsequent emergence of the Hajj cluster from within Cluster 2 strains.

Gene content: All annotated genes within the *M7124* genome were found in at least one pre-Hajj strain, suggesting that the Hajj clone did not acquire additional genes but gained novel alleles by recombination.

16S Ribosomal RNA gene (16S) typing: 16S rRNA type 31 was shared by the Hajj clone and six out of eight Cluster 1 strains while 14 of 16 Cluster 2 strains (87.5%) strains had 16S type 13.

3.6 DISCUSSION

In this study, we demonstrate that W ST-11 strains are closely related to serogroup C ST-11 *N. meningitidis* and likely arose from a single ancestral capsular switching event. We also demonstrate that the emergence of the Hajj clone in 2000 was caused by a meningococcal strain that was distinct from other circulating serogroup W ST-11 strains. This conclusion is supported

by the congruent results from detailed whole genome phylogeny, antigen-encoding gene characterization, and identification of recombinant virulence gene alleles that were unique to the Hajj clone. This study clarifies the recent emergence of serogroup W ST-11 disease globally, which, based on previous limited genetic analyses, appeared to be caused by highly-related strains. To our knowledge, this is the first detailed genomic analysis of serogroup W ST-11 strains that have emerged globally as a major virulent meningococcal lineage.

Based on these results and those from previous studies,[45,74] we propose a model whereby W ST-11 diverged from a C ST-11 ancestral strain through capsular switching before 1970 (Figure 10). From the 1970s onward, W ST-11 strains disseminated to cause sporadic disease and case clusters globally (Cluster 2). Cluster 1 strains (Hajj cluster) evolved from sporadic Cluster 2 strains through allelic exchange within four recombinant regions two of which encode FHbp, nitric oxide reductase and nitrite reductase. This model supports global co-circulation of both Cluster 1 (Hajj cluster) and Cluster 2.

Sequence variation within the *fHbp* gene can potentially be used as a marker to identify the Hajj clone and closely related Cluster 1 strains. For example, a study of *fHbp* gene profiles from 47 endemic W ST-11 strains from 16 African countries isolated from 1980 – 2006 demonstrated that 34% of the strains shared the Hajj clone *fHbp* allele 9.[67] Also, all W ST-11 strains with the Hajj clone *fHbp* allele 9 (variant family 1) were isolated after the Hajj 2000 epidemic while *fHbp* alleles belonging to variant family 2 and 3 were identified both before and after Hajj 2000. These findings are consistent with our results and support our evolutionary model.

Since 2001, Hajj related and endemic non Hajj W ST-11 strains have co-circulated across the meningitis belt. In 2002, the largest recorded epidemic of W ST-11 occurred in Burkina Faso

with 12,000 cases and 1,400 deaths.[88] It was generally believed that the Burkina Faso and other African W ST-11 epidemics were caused by direct spread of the Hajj clone. However, three Burkina Faso strains from 2001 – 2002 analyzed in this study all had antigen-encoding gene and other genomic markers consistent with non Hajj-cluster endemic W ST-11 strains. Additionally, Pajon et al[67] reported that 76.5% (17/ 22) W ST-11 strains from Burkina Faso 2001 – 2003 had *fHbp* genotypes associated with endemic non Hajj strains. Epidemics of W ST-11 subsided in the meningitis belt from 2003-2009 despite persistence of small case clusters but resurfaced in 2010 – 2013.[73,89] Detailed antigenic and genomic characterization of more recent W ST-11 strains from the meningitis belt will be needed to monitor the continued evolution of the Hajj clone and endemic W ST-11 strains.

Outside the meningitis belt, serogroup W strains accounted for 62% of all invasive meningococcal disease strains in South Africa in 2005 compared to 5% in 2000, with 93% of W strains belonging to ST-11 lineage.[90] Our study reveals that 71·4% of 98 South African W ST-11 strains from 2003 – 2013 belonged to the Hajj cluster. These results are consistent with the finding that 85% of invasive W ST-11 strains isolated in South Africa in 2005 had the Hajj - related *fHbp* allele 9.[90] Taken together, these data show that the Hajj cluster strains were predominant in South Africa. In the UK, Hajj related W ST-11 strains predominated in 2000 – 2004 but were replaced by endemic non Hajj strains thereafter.[91] Likewise, our results suggest that the small case clusters of W ST-11 in the United States 2008-2009,[46] and larger clusters in south Brazil 2003-2005,[42] and Chile 2010-2012[54] represent the local spread of endemic strains with no evidence of direct spread of the Hajj clone.

Gene sequencing for 16S ribosomal RNA was previously the most discriminatory test for differentiating the Hajj clone, which exhibited 16S type 31 compared to type 13 and 14 in

sporadic W ST-11 strains.[45] However, our data demonstrate evolution of the 16S rRNA allele within strains linked to the Hajj outbreak and the presence of 16S type 31 in strains that are genetically unrelated to the Hajj clone.

There are several hypotheses that could explain the emergence of the Hajj clone and subsequent W ST-11 outbreaks worldwide.[59] FHbp is a major meningococcal antigen and a virulence determinant that is a component of vaccines developed for protection against serogroup B strains. The introduction of a novel FHbp antigenic type into an immunologically naïve population may have played a part in the emergence of the Hajj clone. In support of this hypothesis, the *fHbp* allele 9 unique to the Hajj related strains, belongs to variant group 1/sub-family B and has limited immunologic cross reactivity with variant groups 2 and 3/sub-family A, which were prevalent among Cluster 2 strains.[92,93] Similarly, antigenic shift was associated with increases in serogroup C and serogroup Y meningococcal disease in the U.S. in the 1990s.[94]

Alternatively, the genomic changes we observed in the Hajj clone may be associated with increased virulence. For example, the nitrite reductase (*aniA*) gene – encoding a major outer membrane copper-containing protein, and the nitric oxide reductase (*nor*, sometimes referred to as *norB*) gene are both essential for overcoming oxidative stress and resistance to phagocytic lysis by macrophages.[95] Also, *N. meningitidis* lacking the *nor* gene have been shown to survive poorly in human nasopharyngeal tissue.[96] Together, these data suggest that allelic variation in key virulence determinants may have a potential contribution to W ST-11 emergence. Then again, the genomic events observed could simply be markers of other events that lead to changing epidemiologic behavior of W ST-11.

This study provides increasing evidence on the role of recombination in the emergence and persistence of meningococcal lineages and demonstrates the role of recombinant gene alleles in molecular epidemiologic typing of meningococcal isolates. We also add to the body of evidence showing the suitability of *fHbp* gene sequencing for routine meningococcal surveillance. A limitation of this study is incomplete data as a result of variations in meningococcal disease surveillance by country and over the study period particularly in the 1970s-1990s when serogroup W strains were thought to be less virulent than other meningococcal lineages.

In summary, this study describes evidence of an ancestral capsular switching event and a model for the emergence, persistence and global spread of W ST-11 strains that are highly related to the Hajj 2000 outbreak strain. These data also demonstrate the co-circulation of W ST-11 strains that are phylogenetically and antigenically distinct from the Hajj clone and still cause disease in the African meningitis belt and globally. The emergence of the Hajj clone may have occurred because of the recent acquisition of a distinct antigen-encoding gene profile and genetic variations in meningococcal virulence genes.

3.7 LEGENDS

Table 1: Homologous recombination regions associated with Hajj specific SNPs. Start and end indicate up- and downstream recombination breakpoints. Numbers and annotations are relative to the Hajj clone reference strain, *M7124*. Known virulence genes are shown in bold. No. of SNPs - Number of highly discriminatory SNPs within genomic region. Closest match - Closest match based on BLASTn query on PubMLST database (pubmlst.org/neisseria/).

Table 2: Characteristics of 26 meningococcal genomes sequenced in this study.

Figure 7: Flow chart showing how 270 invasive W ST-11 meningococcal strains were grouped based on date of isolation (epidemiologic classification) and the combined results of whole genome sequence analysis (genomic classification). All 270 strains were classified into Cluster 1 and 2 based on antigen-encoding gene analyses and presence of recombinant genomic regions. Several strains from each cluster were further analyzed by whole genome phylogenetic and Hajj specific SNP analyses as described in the Methods.

Figure 8: Antigen-encoding gene profiles of 270 invasive serogroup W and historical serogroup C ST-11 strains. On the left is ClonalFrame consensus tree constructed using concatenated full length antigen-encoding gene sequences from *porA, porB, fetA, nadA, nhba* and *fHbp* genes. Color chart in the center depicts antigen gene allelic differences among W ST-11 strains compared to *M7124* allele. Year(s) reflect earliest and most recent isolation dates for strains with the listed antigen-encoding gene profile; N: Total number of strains with listed antigen-encoding gene profile. Numbers in parentheses indicate total number of strains with identical antigen-encoding gene profile to the adjacent strain for profiles shared by more than one strain. On the right, green open triangles mark Cluster 1 strains, red open squares mark Cluster 2 and blue open circles mark serogroup C ST-11 strains. Antigen-encoding gene allele numbers were obtained from www.pubmlst.org/neisseria. Scale bar represents time (coalescent units).

Figure 9: Core genome phylogenetic tree of serogroup W ST-11 and historical serogroup C ST-11 *N. meningitidis* strains (A). A maximum likelihood phylogenetic tree was constructed from aligned universally present genome sequences with 100 bootstrap iterations using the General Time Reversible model, gamma distribution of rate variation with invariant sites (GTR+Γ+I). Scale bar represents phylogenetic distance. Tree is rooted using the serogroup B reference strain *MC58* as outgroup (not shown). Bootstrap support values <60% are not shown. SplitsTree

phylogenetic network of generated from the alignment described above (B). In A-B above, several serogroup W ST-11 strains (Cluster 1) are phylogenetically very closely related to the Hajj clone reference strain *M7124*; remaining W ST-11 strains (Cluster 2) are phylogenetically diverse.

Figure 10: Genomic model of serogroup W ST-11 emergence and global spread. W ST-11 lineage likely emerged from an ancestral serogroup C ST-11 strain through 'capsular switching' – recombination within the capsular gene cluster (*cps*) involving donor sequences from a sporadic W strain and Y ST-23[74] with subsequent antigenic diversification giving rise to sporadic W ST-11 strains (Cluster 2). The Hajj clone (Cluster 1) emerged through recombination within antigen-encoding and virulence genes *fHbp, nor* and *aniA.*

Figures 20-23: Phylogenetic tree of Hajj cluster recombinant regions 1-4. Gene sequences involved in recombination are highlighted in blue; Numbers at the top of annotations represents genomic locations of recombinant regions within *M7124* genome. Maximum Likelihood trees with 500 bootstrap replicates were constructed using the Hasegawa, Kishino and Yano (HKY) model of evolution under uniform rate of substitution.

Figure 24: Phylogenetic tree of serogroup W, C and B ST-11 strains in relation to other hyperinvasive meningococcal lineages. A maximum likelihood phylogenetic tree was constructed from concatenated 140,903 SNP positions among 153 meningococcal strains using the General Time Reversible model, gamma distribution of rate variation with invariant sites (GTR+Γ+I). Green triangles represent Hajj cluster (Cluster 1) W ST-11 strains, blue diamond represents Cluster 2, white diamonds and red squares represent serogroup C and B ST-11 respectively. Scale bar represents phylogenetic distance.

3.8 AUTHOR CONTRIBUTIONS

Conception of study hypothesis, aims and analytic plans: MMM, JWM, JGL, NLH, LHH. Acquisition and molecular characterization of meningococcal isolates: JWM, JOF, APSL, XW, LWM, LHH. Meningococcal genome sequencing, data analysis and interpretation: MMM, JWM, MGK, APSL, JOF, JCDH, XW, LWM, JGL, NLH, LHH. Initial draft and revision of study manuscript: MMM, JWM, MGK, JCDH, XW, LWM, JGL, NLH, LHH. Read and approved final manuscript: MMM, JWM, MGK, JOF, APSL, JCDH, XW, LWM, JGL, NLH, LHH.

3.9 DECLARATION OF INTERESTS

Dr. Harrison reports grants and personal fees from Sanofi Pasteur, personal fees from GSK, personal fees from Merck, personal fees from Novartis, personal fees from Pfizer, outside the submitted work; and All relationships with industry were terminated before I became a voting member of the Advisory Committee on Immunization Practices on July 1, 2012. Dr Lemos reports travel grants and personal fees from Novartis, personal fees from Sanofi Pasteur , travel grants from GSK, outside the submitted work. Other co-authors have no interests to declare.

3.10 ACKNOWLEDGEMENTS

We thank Maria Brooks for thoughtful discussions about the study. We also thank Jessica Schlackman for DNA extraction and genotyping of bacterial strains and Jeremy Nelsen for assistance with bioinformatics analyses.

This publication made use of the *Neisseria* Multi Locus Sequence Typing website (pubmlst.org/ neisseria/) developed by Keith Jolley and sited at the University of Oxford.[31] The development of this site has been funded by the Wellcome Trust and European Union.

This publication also made use of the Meningitis Research Foundation Meningococcus Genome Library (www.meningitis.org/research/genome) developed by the Health Protection Agency, the Wellcome Trust Sanger Institute and the University of Oxford as a collaboration.

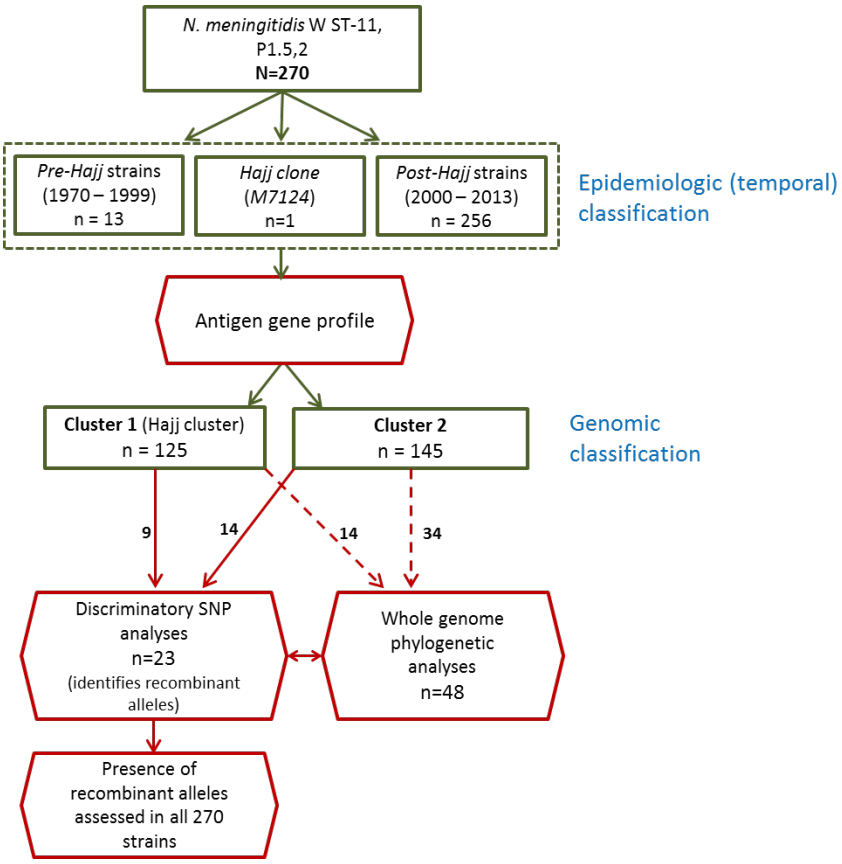

Figure 7: Flow chart showing classification of invasive W ST-11 meningococcal strains.

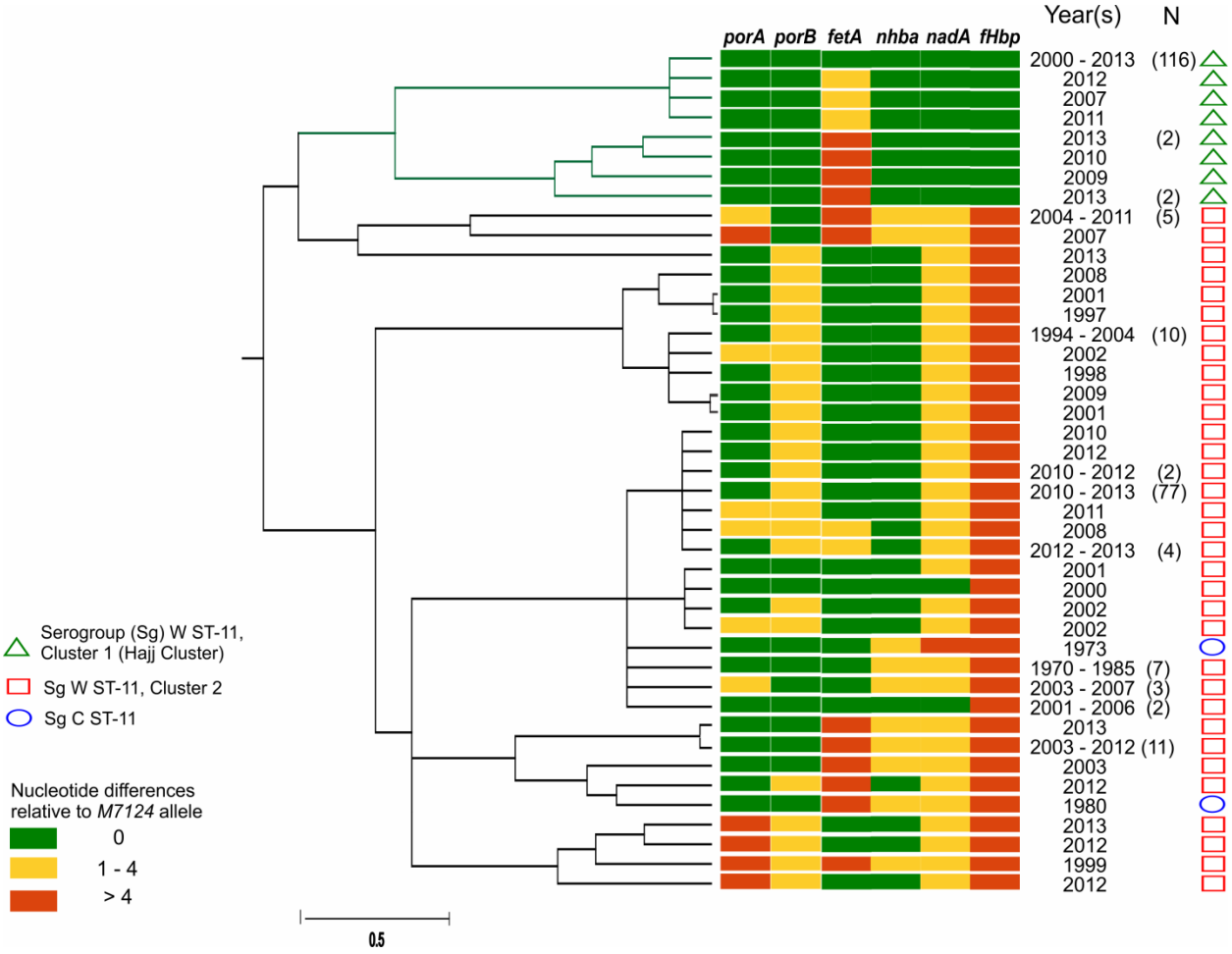

Figure 8: Antigen-encoding gene profiles of 270 invasive serogroup W and historical serogroup C ST-11 strains.

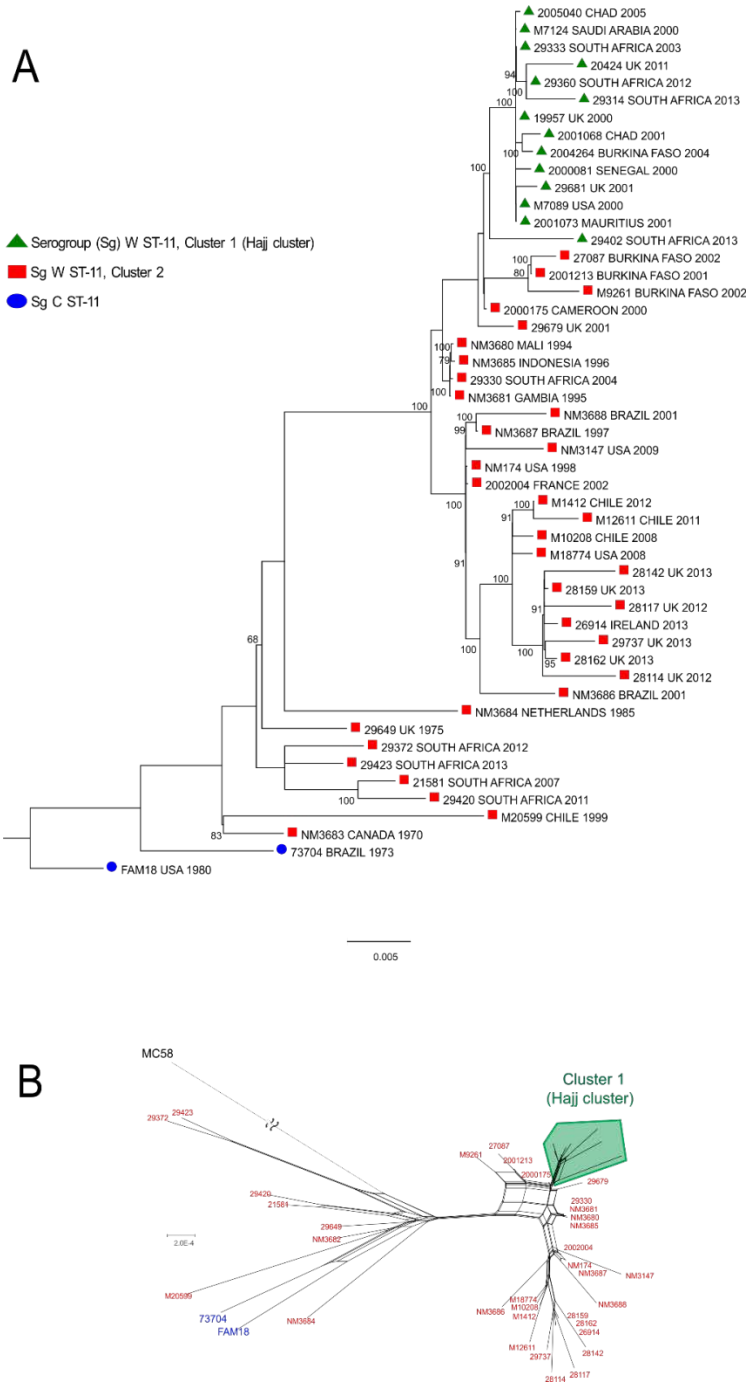

Figure 9: Phylogenetic trees of serogroup W ST-11 and historical serogroup C ST-11 *N. meningitidis* strains.

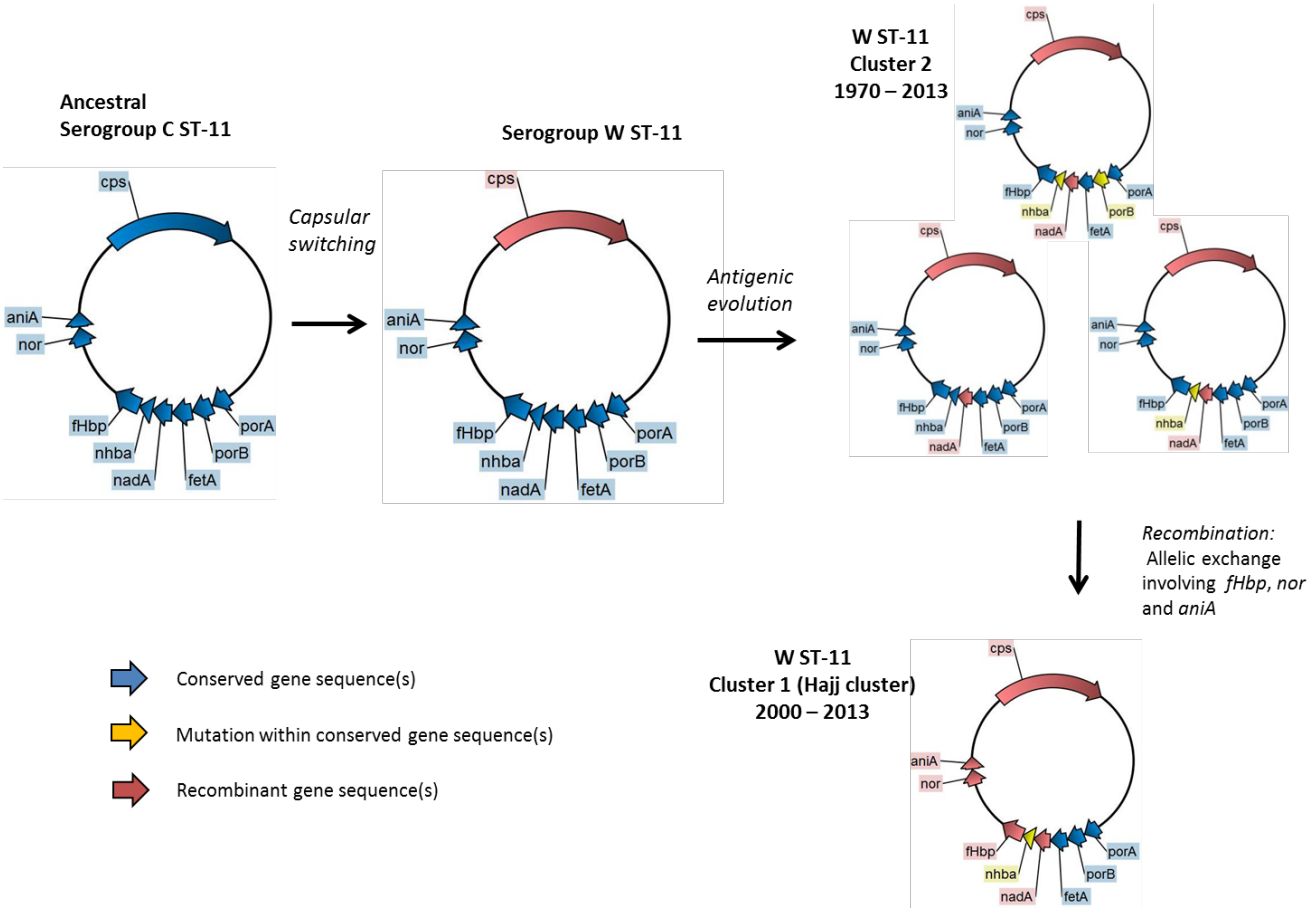

Figure 10: Genomic model of serogroup W ST-11 emergence and global spread.

Table 1. Homologous recombination regions associated with Hajj specific SNPs.

	Genomic location of Hajj specific SNPs					
Recombinant region	**start**	**end**	**size (kb)**	**No. of SNPs**	**Annotations**	**Closest match**
1	329041	330204	1.2	5	Phosphopantetheine adenylyltransferase, *coaD*; Ribosomal large subunit pseudouridine synthase D	*Neisseria cinerea*
2	628712	632970	4.3	13	fructose-bisphosphate aldolase, *fba*; **Factor H binding protein, *fHbp***; glycoprotease family protein; ribosomal-protein-alanine acetyltransferase, *rimI*; uracil DNA glycosylase; Orotate phosphoribosyltransferase, *pyrE*; MJ0042 family finger-like domain protein; amino-acid N-acetyltransferase, *argA*	*Neisseria meningitidis* serogroup B ST-639 (cc32)
3	896826	900840	4.0	13	Argininosuccinate lyase, *argH* ; **UTP-Glucose-1-phosphate uridyl transferase, *galU*;** purine NTP pyrophosphatase, *rdgB*; dATP pyrophosphohydrolase, *ntpA*; inorganic pyrophosphatase, *ppa*	*Neisseria spp* ST-6263 (cc -)
4	1844312	1846343	2.0	15	**Nitric oxide reductase, *nor*; Nitrite reductase, *aniA***	*Neisseria meningitidis* ST-461 (cc461)

4.0 PAPER 3: GENOMIC INVESTIGATION REVEALS HIGHLY CONSERVED, MOSAIC, RECOMBINATION EVENTS ASSOCIATED WITH CAPSULAR SWITCHING AMONG *NEISSERIA MENINGITIDIS* GROUP W ST-11 STRAINS.

Authors: Mustapha M. Mustapha[1*], Jane W. Marsh[1], Mary G. Krauland[8], Jorge O. Fernandez[2], Ana Paula S. de Lemos[3], Julie C. Dunning Hotopp[4], Xin Wang[5], Leonard W. Mayer[5], Jeffrey G. Lawrence[6], N. Luisa Hiller[7], Lee H. Harrison[1]

Affiliations: [1]Infectious Diseases Epidemiology Research Unit, University of Pittsburgh, Pittsburgh, PA 15261, USA. [2]Molecular Genetics Laboratory, Public Health Institute of Chile, Santiago, Chile. [3]Department of Bacteriology, Instituto Adolfo Lutz (IAL), São Paulo, Brazil. [4]The Institute for Genome Sciences, University of Maryland School of Medicine, University of Maryland, Baltimore, Maryland, USA. [5]Meningitis and Vaccine Preventable Diseases Branch, Division of Bacterial Diseases, National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA.[+] [6]Department of Biological Sciences, University of Pittsburgh, USA. [7]Department of Biological Sciences, Carnegie Mellon University, Pittsburgh, PA, USA. [8]Public Health Dynamics Laboratory, Graduate School of Public Health, University of Pittsburgh, USA.

*Corresponding author (email: mmm147@pitt.edu)

Running title: W ST-11 meningococcal capsular switch.

Key words: *capsule switch, meningococcal virulence, serogroup W135.*

4.1 ABSTRACT

Neisseria meningitidis group W ST-11 is an important cause of meningococcal disease globally. ST-11 is a hypervirulent meningococcal lineage historically associated with group C capsule type and is believed to have acquired the W capsule through a C to W capsular switching event. We studied the sequence of capsule gene cluster (cps) and adjoining genomic regions of 525 global W ST-11strains.We identified recombination breakpoints corresponding to two distinct recombination events within W ST-11. 1) An 8.4 kb recombinant region likely acquired from W ST-22 that includes the sialic acid transferase gene, *csw* resulted in a C→W change in capsular phenotype. 2) A 13.7 kb recombinant segment likely acquired from Y ST-23 lineage that includes 4.5kb of cps genes and 8.2kb outside the cps cluster and led to allelic changes in capsule translocation genes. A vast majority of W ST-11 strains (497/524, 94.8%) have identical or very closely related capsular allelic profiles. These data suggest that the W ST-11 capsular switch involved two separate recombination events and that current global W ST-11 meningococcal disease is caused by strains bearing this capsular switch.

4.2 INTRODUCTION

Polysaccharide capsule is the most important virulence determinant of *Neisseria meningitidis*, a normal commensal of the human nasopharynx[2] that occasionally causes invasive meningococcal disease (IMD). Even though un-encapsulated meningococci are common in asymptomatic carriage, they are very rarely associated with IMD.[97] This polysaccharide capsule is the target of capsule-specific vaccines and its biochemical and genetic properties form the basis of classifying meningococci into capsular groups (serogroups). Out of 13 capsular groups, groups A, B, C, W, X and Y cause almost all cases of IMD [2].

Capsular group W sequence type (ST) 11 evolved from a rare cause of meningococcal disease in the 1970s-1990s [41] to a leading cause of invasive meningococcal disease globally in 2000-2015 [3,73]. Epidemic W ST-11, often referred to as the Hajj clone emerged in Mecca, Saudi Arabia among Hajj pilgrims and their close contacts in 2000 [16]. W ST-11 is a major global hypervirulent lineage responsible for large epidemics in the 'meningitis belt'[17,66] and high burden of endemic cases in Europe [3], Latin America [48] and the Middle East[65]. Whole genome phylogenetic studies suggest that recent global W ST-11 disease burden represents both continued global spread of the Hajj clone and persistence of endemic W ST-11 strains unrelated to the Hajj outbreak. [98,99]

N. meningitidis has high frequency of genetic change through homologous recombination that drives meningococcal genetic diversity and plays a role in the emergence and persistence of virulent strains [100]. Capsular switching, acquisition of novel capsular genes, is one mechanism by which *N. meningitidis* may escape population immunity [34]. Capsular switching is a relatively common phenomenon among invasive meningococcal isolates [35,36,101].

The capsule gene cluster (cps), a 24 kb pathogenicity island, contains 5 gene regions involved in capsule synthesis, transport and assembly (cps regions A-E) [10]. Region A contains genes that synthesize capsular polysaccharide and are capsular-group specific while conserved regions B and C are involved in capsule translocation. Groups B, C, W and Y have capsules containing sialic acid residues[10,34]. The sialic acid transferase gene, located in region A, dictates sialic acid linkage and thus determines group phenotype. Regions D and D' are involved in lipopolysaccharide synthesis while the function of genes within region E as not been clearly defined. The genetic basis of capsular switching is homologous recombination involving allelic exchange of cps genes. For a recombination event to result in change of capsular phenotype, allelic changes must include region A genes, particularly the sialic acid transferase [34].

Historically, ST-11 was predominantly associated with virulent group C strains and W ST-11 is thought to have arisen from C→W capsular switching[37,59] However, the genetic events associated with this C→W capsular switching have not been established. Furthermore, it is not established whether the Hajj clone underwent recombination event(s) that were distinct from those in W ST-11 strains causing rare sporadic disease cases that occurred before 2000.

The purpose of this study was to analyze capsular gene sequences among meningococcal isolates from a global collection to define the capsular switching events that occurred in the W ST-11 lineage.

4.3 METHODS

Study isolates: A total of 529 serogroup W ST-11 whole genome sequences were identified from PubMLST (www.pubmlst.org/neisseria), a database that captures *Neisseria* genetic diversity by assigning allele numbers for all *Neisseria* genes. Four isolates (*19369, 19377, 19379*, and *19381*) were excluded because they had missing data for a majority of cps alleles. Capsular contigs and allele designations for 525 isolates were included in the study. One of these isolates, *M7124*, identified in Saudi Arabia during the Hajj 2000 epidemic is a well characterized reference strain for the W ST-11 lineage and is used in this study as W ST11 cps reference genome.[99]

Identification of recombination donors and breakpoints for M7124: To determine the organization of cps genes, contiguous nucleotide sequences (contigs) containing cps sequences were extracted for the Hajj clone reference strain, *M7124*.

M7124 cps sequences were compared using SplitsTree[81] phylogenetic networks and visual inspection of sequence alignments with previously described cps sequences for reference meningococcal strains [10]. Also, potential recombination donor lineages were assessed by querying *Neisseria* genome sequences on PubMLST and GenBank databases to determine closest matches outside of the W ST-11 lineage. A strain was considered a potential donor for *M7124* capsular genes if it had ≤1 nucleotide difference per 1000bp over a span of at least 2kb compared to *M7124*. Recombination breakpoints were identified from sequence alignments as points of abrupt change in sequence similarity between *M7124* and a potential donor strain.

Comparison of cps alleles within W ST-11 lineage: M7124 cps gene sequences were compared to 524 W ST-11 strains with whole genome sequence data isolated from 21 different countries. Majority (93.3%, 489/524) were isolated in 2000 – 2014 with the remaining 35 isolated in 1970-

1999. Most isolates were from in the United Kingdom (331/524, 63.2%), South Africa (116/524, 22.1%) or the meningitis belt (34/524, 6.5%). Gene alleles were downloaded for sixteen cps genes (*ctrA-G, cssA-C, csw, tex, galE, galE2, rfbC* and *rfbC2*) that had complete sequences for ≥80% of the study isolates. Allele information for *rfbA* and *rfbB* genes was excluded due to missing data among 91% (479/525) and 79% (416/525) of total isolates respectively. Also, gene elements not curated by PubMLST including insertion sequences, putative fragments in close proximity to insertion sequences and two pseudo-genes within region E were all excluded from further analyses. Over 70% of isolates (368/524) had complete sequence data for all 16 cps genes, 22.3% (117/524) had missing data for one or two genes while 39 (7.3%) had missing data for three to five genes. *GalE2* (99/524, 19.9%), and *galE* (94/524, 17.9%) genes had the highest frequency of missing data.

Isolates that had an identical allelic profile to M7124 or had 1-3 nucleotide differences across all non-missing cps genes were considered to share the same capsular gene structure with M7124. Isolates with >3 nucleotide differences across all cps genes were further examined to identify whether nucleotide differences clustered into one or more genes. Any gene that differed from M7124 by >3 nucleotide was defined as an 'allelic shift' suggestive of a unique recombination in that particular gene in a given isolate compared to M7124. Presence and total number of allelic shifts were assessed for each isolate.

Contiguous genome sequences (contigs) containing cps genes for 65 representative isolates selected to capture geo-temporal and allelic diversity of the entire collection, were concatenated and aligned. MEGA v5.2 [83] and CLC genomics workbench v8.0 (www.clcbio.com) were used to generate phylogenetic trees and sequence alignments. Maximum likelihood phylogenetic trees

with 500 bootstrap replicates were constructed under the Hasegawa, Kishino and Yano model of evolution with Γ-distribution of substitution rates and invariant sites (HKY+ Γ+I).

4.4 RESULTS

The cps locus of M7124 is 27kb with genes organized into regions D-A-C-E-D'-B (Figure 11) similar to that described for other meningococcal lineages[10]. When *M7124* cps sequences were compared to other cps reference sequences, a number of points marking abrupt changes in sequence similarity consistent with distinct recombination events were identified:

1) *M7124* shares very high sequence similarity with group W ST-22 reference strain α-275 over an 8.4 kb segment that includes all cps region A genes. Genes involved in this recombination event (listed from 5' to 3') included *galE, galU, ctrG, cssF, csW, cscC, cssB and cssA*. (Figure 11). Only 3 nucleotide differences are evident between *M7124* and *α-275* over the entire 8.4 kb segment (Figure 11). Recombination breakpoints at nucleotide positions 858 and 804 on *galE* and *cssA* genes respectively were demonstrated by alignment of sequences from representative meningococcal isolates (Supplementary Figure 11 and 12). The sialic acid transferase gene, *csw* acquired within this recombination is responsible for capsular group W phenotype (Figure 11). Phylogenetic networks (Figure 12) shows that *cssA-C* and *csw* gene sequences on M7124 were most closely related W ST-22 reference capsule α-275.

To assess what proportion of W ST-22 strains shared the same recombinant segment with M7124, we compared *cssB-C, csw* and *ctrG* gene alleles between M7124 and 55 W ST-22 strains in the PubMLST database. Eighty percent (44/55) of W ST-22 strains shared a common allelic profile that differed from M7124 by two nucleotide substitutions within *csw*. These data demonstrate that the recombinant segment is predominant among W ST-22 strains.

2) A 13.7kb region of very high sequence similarity between M7124 and the group Y ST-23 includes 4.5kb of cps and 9.2kb downstream of the cps cluster (Figure 11). The entire region B (*ctrE* and *ctrF*) and parts of region D' (*rfbA2* and *rfbC2*) were part of this recombinant segment (Figure 11-12). This recombinant segment extends beyond the cps genes and includes the genes *gltS*, encoding a sodium/glucose transport protein, a pyruvate kinase gene (*pykA*), *sel1* and several genes encoding hypothetical proteins. A corresponding recombination breakpoint was identified at position 611 of *rfbA2* gene (Figure 11-12, Supplementary Figure 18) while a downstream breakpoint is located outside the cps cluster at the 5' end of *pykA* gene.

We further compared *ctrE* and *ctrF* gene alleles between M7124 and 201 Y ST-23 strains in the PubMLST database. Over 88% (177/201) of all Y ST-23 isolates shared identical *ctrE* and *ctrF* alleles with M7124 demonstrating that this allelic combination is predominant among Y ST-23 strains.

M7124 gene sequences were then compared to whole genome sequence data from 524 W ST-11 strains isolated from 21 different countries. When compared to M7124, the majority of W ST-11 strains had cps gene sequences that were either identical to M7124 or differed only by ≤ 3 point mutations; 63.7% (334/524) of isolates had cps allelic profiles that were identical to M7124, another 162 isolates (30.9%) differed from M7124 by a total of 1-3 point mutations across all cps genes. Thus 94.7% (496/524) of W ST-11 strains shared the same cps gene structure as M7124, which predominated across the entire study period from 1970-2014 (Figure 13 and 14). Phylogenetic tree constructed using maximum likelihood shows strains with same cps structure as M7124 (Figure 15, red boxes) belonged to a single phylogenetic cluster while strains with >3 SNPs relative to M7124 (Figure 15, blue dots) were diverse; some belonged to the same cluster as M7124 while majority were distinct from M7124.

Twenty nine strains (29/524, 5.5%) with non M7124 cps gene structure were further examined. All had one or more genes that exhibited allelic shift, defined as >3 nucleotide differences within a single (Figure 13, blue boxes). Genes with allelic shift likely differed from M7124 reference allele through recombination. Among these 29 isolates, 44.2% (13/29) had allelic shift in a single cps gene while the remaining 16 isolates shared 2-7 allelic shifts (Figure 13, blue rectangles). Seven strains isolated in South Africa in 2006-2013 each had three allelic shifts and shared a common cps allelic combination in keeping with geographically limited clonal spread of strains with this particular cps allelic profile (Figure 13). Twenty two remaining strains with one or more allelic shifts were diverse without any predominant allelic profile and were isolated throughout the study period (Figure 14).

4.5 DISCUSSION

In this study, we present detailed characterization of the capsule gene locus, the primary meningococcal virulence determinant within the context of globally emergent hypervirulent group W ST-11 lineage. To our knowledge, this is the first in-depth study of capsule gene sequence variation within a single meningococcal lineage.

Taken together, these data suggest that W ST-11 likely acquired capsule polymerase (*csw*) gene allele along with other region A genes from a W ST-22 strain and acquired capsule translocation gene alleles from Y ST-23 lineage. Also, an overwhelming majority of W ST-11 strains had cps gene alleles very closely related to M7124 in keeping with a shared capsule recombination structure and a common ancestor. Strains with one or more divergent alleles were a very small minority but occurred throughout the study period. These data are consistent with recent whole genome sequencing studies showing that all W ST-11 strains belonged to a single lineage that is

phylogenetically distinct from group C and B ST-11 lineages.[98] Very high degree of sequence conservation among W ST-11 capsule genes markedly contrasts the substantial global and temporal variability in W ST-11 disease incidence patterns, a lineage associated with both rare sporadic disease[49,58] and epidemics[17] In fact, our previous study identified that the Hajj clone and its descendants acquired virulence gene alleles, outside of the cps cluster, through a set of unique recombination events affecting factor H binding protein (*fHbp*), nitric oxide reductase (*nor*) and nitrite reductase (*aniA*) genes.[99]

These data suggest that the Hajj clone and a vast majority of W ST-11 strains acquired an 8.4kb recombinant fragment from group W ST-22 lineage. This recombination event includes sialic acid transferase gene that mediates formation of (α2→6) sialic acid and glucose heteropolymers characteristic of group W capsule.[10] These findings are consistent with those of Swartley et al[34] that identified allelic exchange spanning over 9kb of cps genes including the entire region A and resulting in a group B to C ST-32 capsular switch among isolates obtained during a group B epidemic in Oregon, USA. Wang et al[102] also reported a switch from group A ST-7 to C in China consisting of a 12kb recombination involving the entire cps region A and flanking genes within region D and C. These studies highlight that even though allelic exchange within a single gene (sialic acid transferase) may be sufficient to cause capsular switching, large recombination segments affecting several genes are not uncommon.

We also identified additional recombination involving genes associated with capsule translocation (*ctrE, ctrF*) from Y ST-23 donor lineage. This recombination event did not involve capsule synthesis genes in region A and therefore had no corresponding change in capsular group. However, even without an obvious change in capsular phenotype, such allelic exchanges

outside region A could affect meningococcal virulence by altering capsule transport and modification. [103] *CtrE,* for example, is up-regulated during meningococcal invasion of human cells [104] and recombination within this gene could potentially enhance virulence through enhanced intracellular survival of W ST-11. Persistence of one particular cps allelic combination despite continuous generation of closely related allelic variants is consistent with the 'genocloud' concept in which a main allelic profile persists despite presence of a few, mostly transient, escape variants [64]. The dominant allelic type likely represents a set of mutually co-adapted genes that may confer a fitness advantage to W ST-11 strains.[105]

In summary, we have demonstrated that the W ST-11 lineage arose through recombination from ancestral group C ST-11 and group Y ST23 strains. Remarkably, this recombination pattern has persisted despite decades of global spread and emergence of the epidemic Hajj clone.

4.6 DECLARATION OF INTERESTS

Dr Lemos reports travel grants and personal fees from Novartis, personal fees from Sanofi Pasteur, travel grants from GSK, outside the submitted work. Other co-authors have no interests to declare.

4.7 ACKNOWLEDGEMENTS

This publication made use of the *Neisseria* Multi Locus Sequence Typing website (pubmlst.org/neisseria/) developed by Keith Jolley and sited at the University of Oxford.[31] The development of this site has been funded by the Wellcome Trust and European Union.

This publication also made use of the Meningitis Research Foundation Meningococcus Genome Library (www.meningitis.org/research/genome) developed by the Health Protection Agency, the Wellcome Trust Sanger Institute and the University of Oxford as a collaboration.

4.8 FIGURE LEGENDS

Figure 11: Capsule gene cluster (cps) for W ST-11 reference strain M7124. Cps is comprised of five regions A-E arranged as shown. Two recombination events are depicted; Blue bars:: an 8.4kb W ST-22 donor fragment; a 13.7kb Y ST-23 donor fragment; Blue arrows indicate associated recombination breakpoints within *galE, cssA* and *rfbA2*.

Figure 12: Phylogenetic relationships of M7124, W ST-11 to cps from meningococcal reference genomes. MC58, B ST-32; H4476, B ST-41/44 alpha275, W ST-22; WUE171 W ST-11; FAM18, C ST-11; 8013 C ST-18; 53442 C ST-4821; alpha162 Y ST-23. Dotted red boxes depict margins of sequence alignments used to generate corresponding phylogenetic network.

Figure 13: Graphic depiction of 524 W ST-11 cps allelic profiles. Genes with ≤3 nucleotide differences to M7124 allele are represented by red arrows, blue arrows represent genes with >3 nucleotide differences to M7124 within a single gene; white boxes represent genes with incomplete sequence data. Strain ID indicates strain identifier of a representative strain within each cps allelic profile. N- number of isolates.

Figure 14: Distribution of W ST-11 cps sequences over time. Red bars depict strains with ≤3 nucleotide differences across all cps alleles relative to M7124; blue bars depict strains with >3 SNPs compared to M7124.

Figure 15: Maximum likelihood tree depicting phylogenetic relationship between selected W ST-11 cps sequences. Tree shows strains with ≤3 nucleotide differences to M7124 (red star and red squares) and those with ≥4 nucleotide differences to M7124 (blue dots).

Figure 16: Sequence alignments *galE* gene showing recombination breakpoints within M7124. Sequence alignment of *galE* genes showing recombination breakpoint at position 858 relative to

W ST-22 strain, alpha275. M7124, W ST-11; WUE171, W ST-11; alpha275, W ST-22; FAM18, C ST-11; alpha162, Y ST-23; MC58, B ST-32; H4476, B ST-44; 053442, C ST-4821; 8013, C ST-18.

Figure 17: Sequence alignments *cssA* gene showing recombination at position 804.

Figure 18: Sequence alignment of *rfbA2* and *rfbC2* genes showing recombination breakpoint within *rfbA2* position 634.

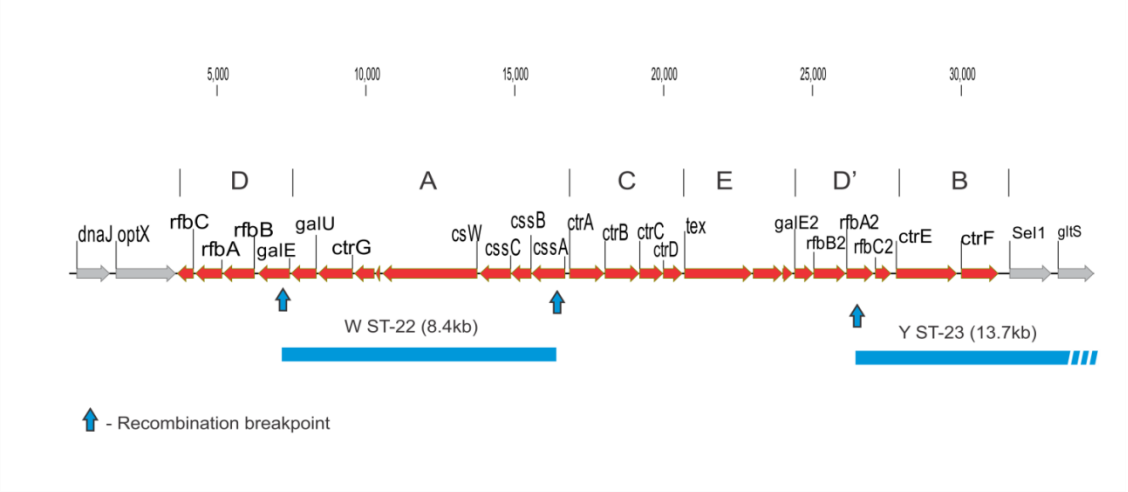

Figure 11: Capsule gene cluster (cps) for W ST-11 reference strain M7124.

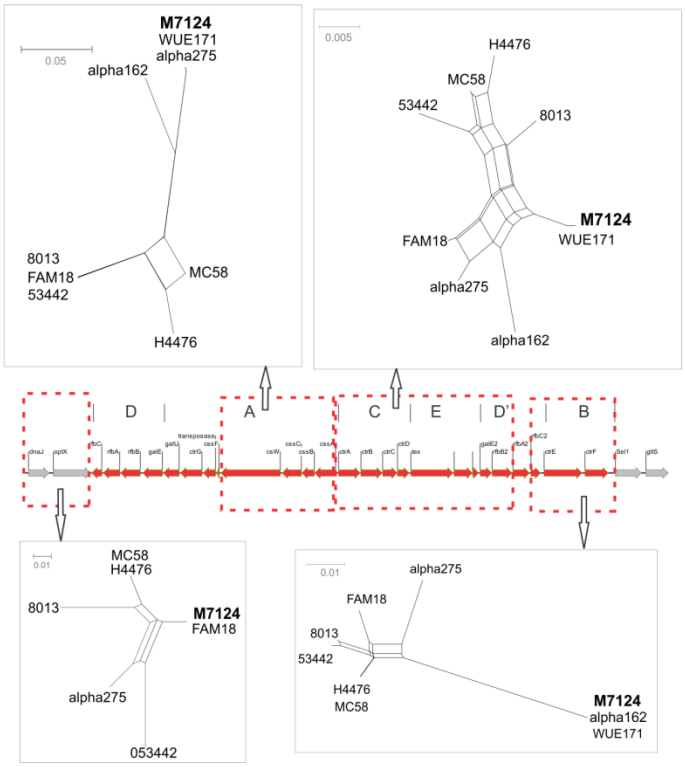

Figure 12: Phylogenetic relationships of W ST-11 and cps sequences from meningococcal reference genomes.

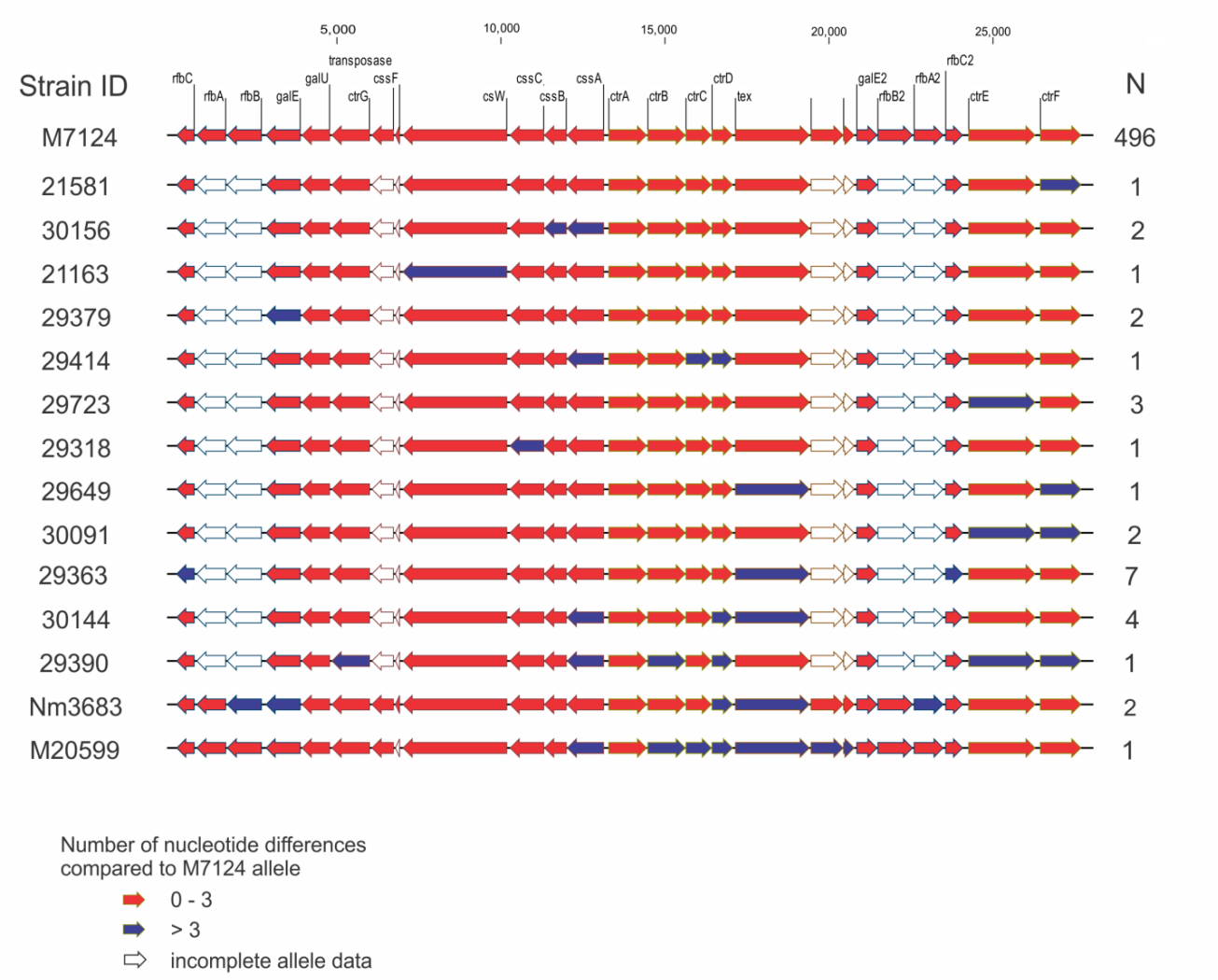

Figure 13: Graphic depiction of 524 W ST-11 cps allelic profiles.

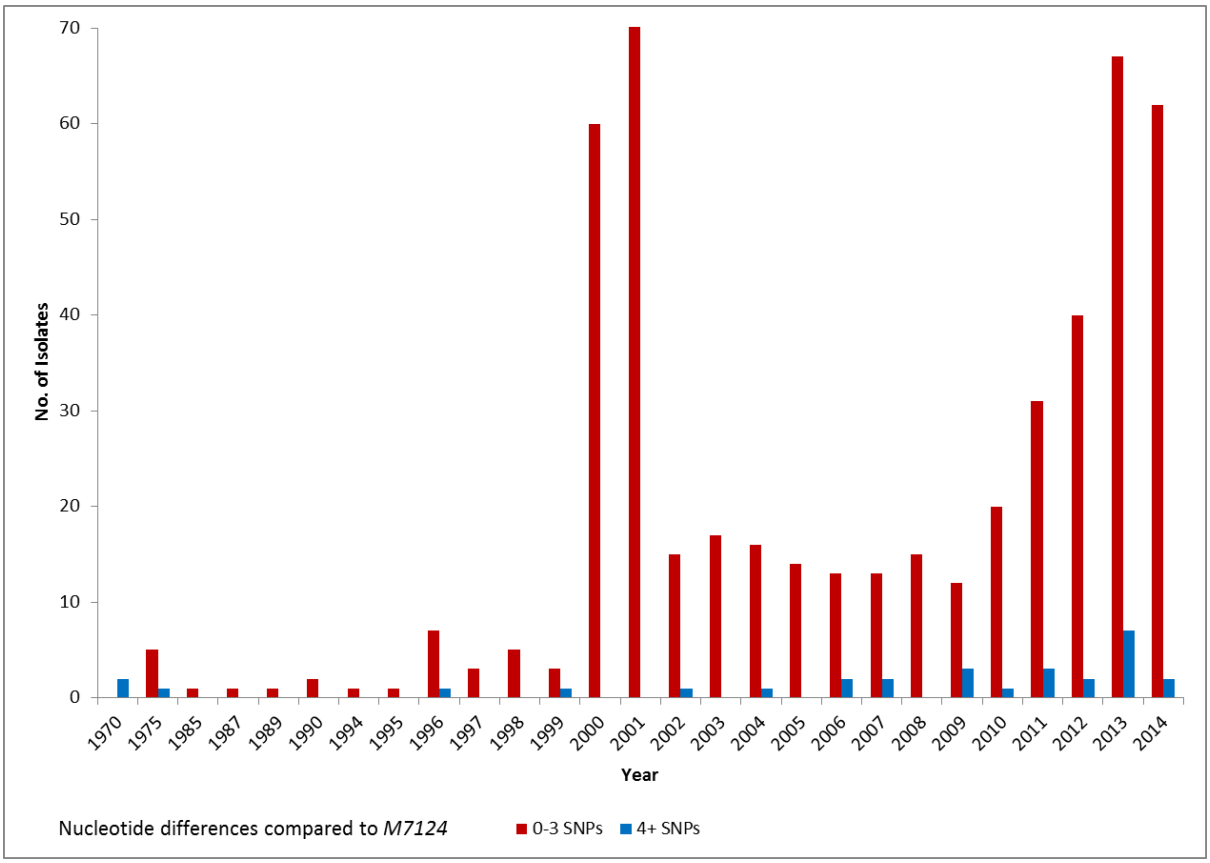

Figure 14: Distribution of W ST-11 cps sequences over time.

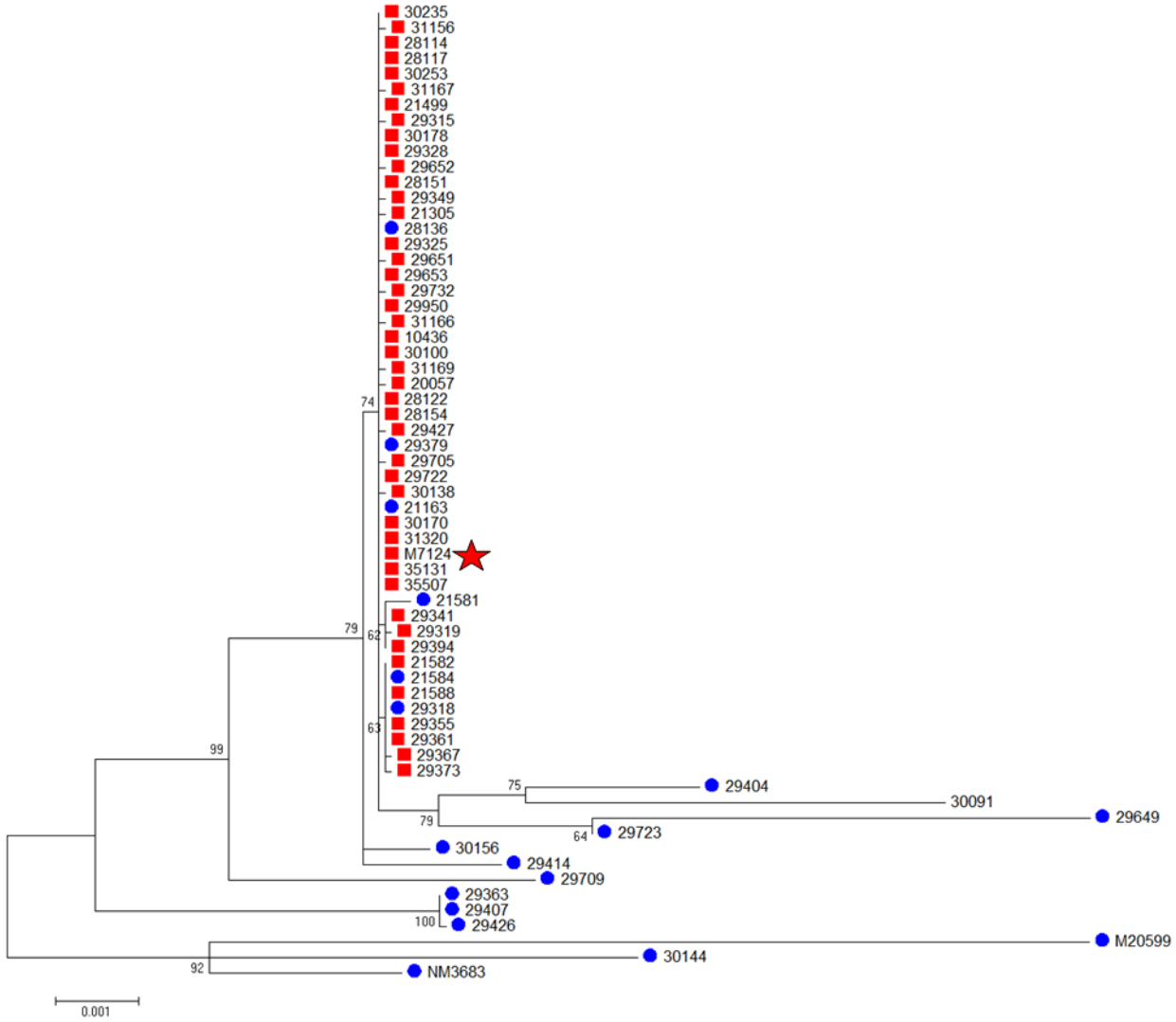

Figure 15: Phylogenetic relationship between W ST-11 cps sequences.

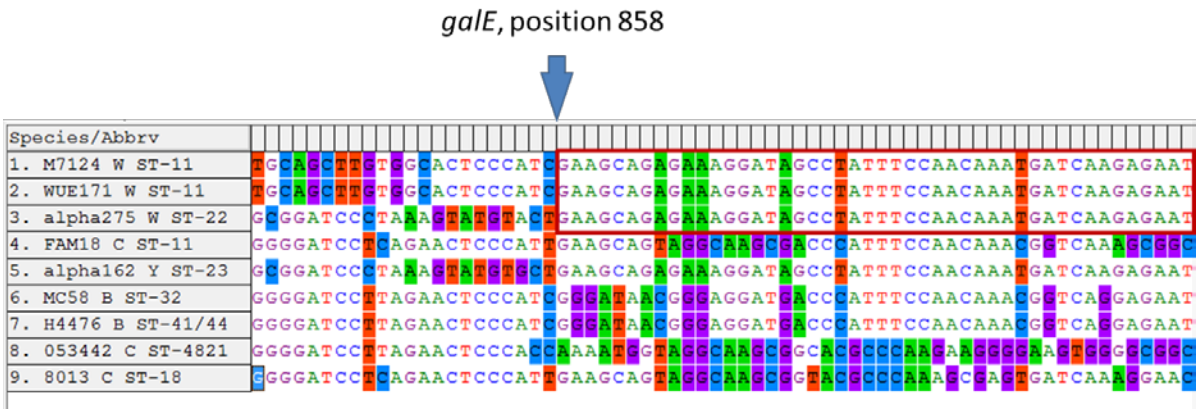

Figure 16: Sequence alignments of *galE* gene showing recombination breakpoint.

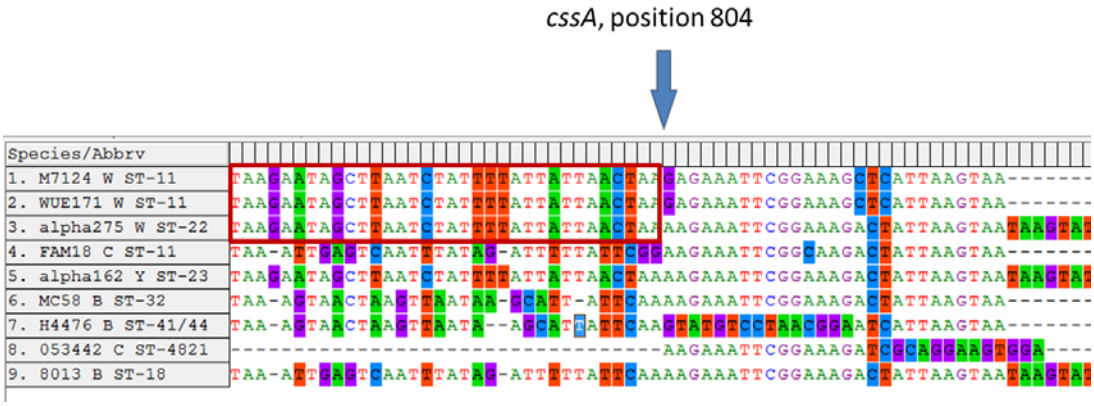

Figure 17: Sequence alignments of *cssA* gene showing recombination breakpoint.

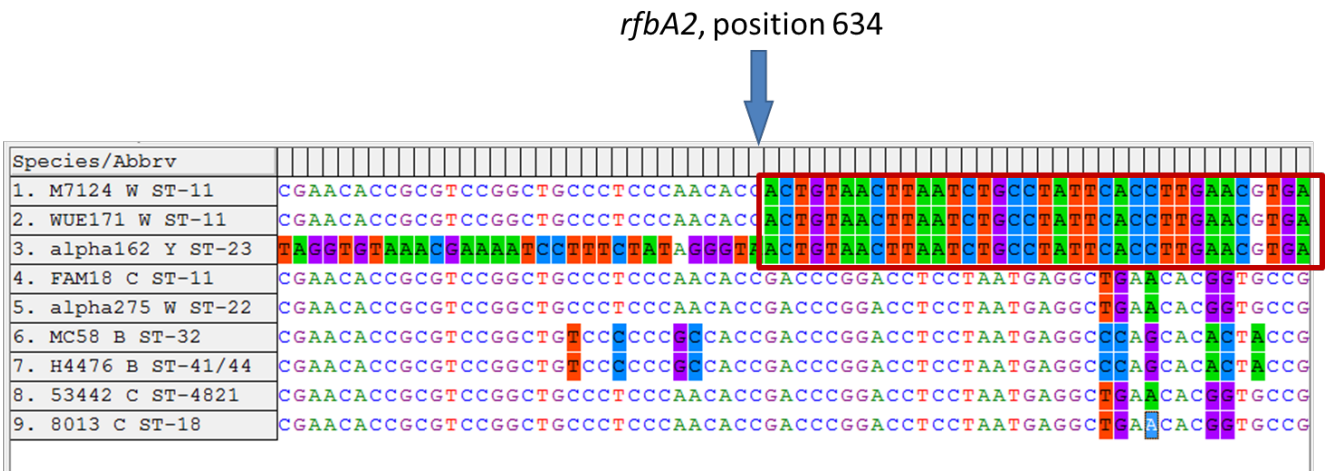

Figure 18: Sequence alignment of *rfbA2* genes showing recombination breakpoint.

5.0 DISCUSSION

5.1 MAJOR FINDINGS

This dissertation resulted in key findings that illuminate several factors that lead to the emergence of W ST-11 as a hypervirulent meningococcal lineage.

1. *All W ST-11 strains shared extensive ancestral mosaic recombination within their capsule genes*: This finding together with phylogenetic evidence of W ST-11 representing clonal expansion within the larger group C ST-11 lineage provides strong evidence that the majority of global W ST-11 strains represent clonal expansion of a single strain that had undergone capsular switching. On one hand, this capsular switch strain has spread and persisted globally for at least half a century, suggesting that the specific combination of capsular genes confers fitness to W ST-11 strains. On the other hand, a varying degree of virulence among W ST-11 strains suggests that genetic variation outside of the capsule genes were associated with differences in W ST-11 epidemiology.
2. *Current W ST-11 strains represent multifocal emergence, as opposed to simple global spread of a single strain that was suggested by the traditional model of ST-11 W evolution:* Whole genome sequence analysis confirmed that several recent endemic W ST-11 strains were phylogenetically distinct from the Hajj clone.
3. *The Hajj clone underwent genetic change that differentiates it from endemic W ST-11 strains:* We identified distinct recombination in antigen encoding and metabolic genes in the Hajj clone and its descendants that may have accounted for increased virulence of the Hajj clone (Figure 19). These data support the key role of allelic exchange within core

genes in the emergence and persistence and evolution of hypervirulent meningococcal lineages.

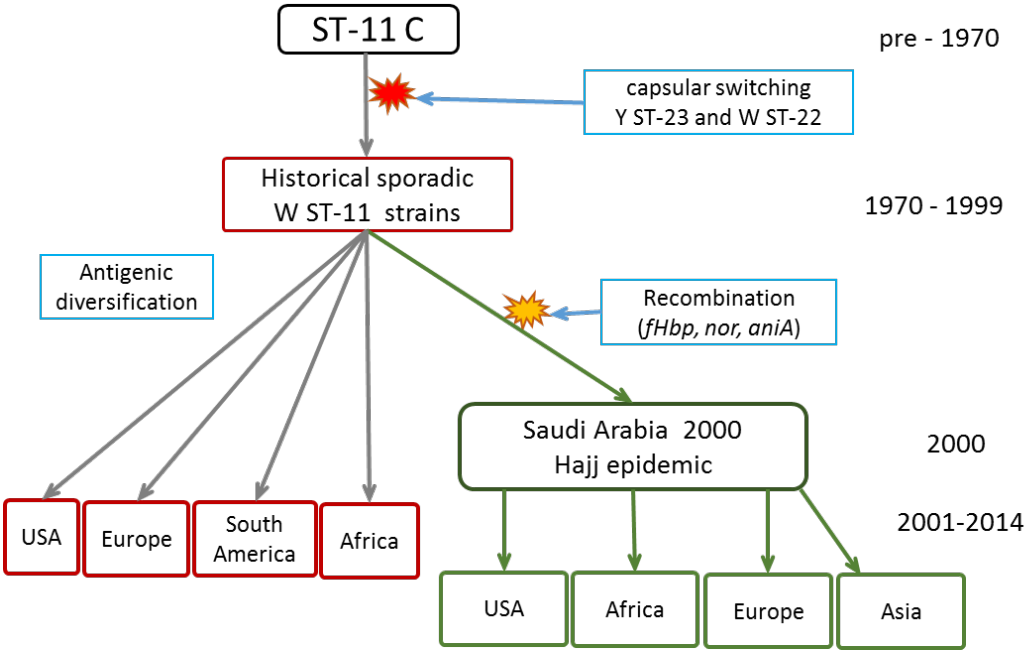

Figure 19: Revised genomic model of W ST-11 evolution based on this dissertation research.

5.2 SIGNIFICANCE

5.2.1 Public Health significance and implications for vaccination policy.

Invasive meningococcal disease is a rapidly progressive disease that can cause death within 24-48 hours in previously healthy individuals or may lead to serious sequelae among survivors.[3] Additionally, outbreaks of meningococcal disease cause substantial economic cost and societal disruption.[106] Vaccines remain the primary control strategy against meningococcal disease with available meningococcal vaccines targeting either the capsule or outer membrane protein antigens.[26,107] Laboratory surveillance of circulating meningococcal strains is critical for

monitoring trends, informing vaccine policy, timely detection of outbreaks, and understanding the emergence of new IMD strains.[5]

W ST-11 lineage that is associated with the entire epidemiologic spectrum of IMD: sporadic disease, smaller outbreaks, high endemic cases and occasional epidemics. An important contribution of this dissertation to public health is the provision of useful genomic markers that reliably distinguish epidemic from sporadic W ST-11 strains. FHbp genotyping, when combined with standard molecular methods such as MLST and PorA antigen typing could provide additional resolution in discriminating W ST-11 strains.[21]

Successful rollout in 2010-2014 of a highly effective and inexpensive group A vaccine have led to the elimination of group A epidemics in the meningitis belt and ushered a new paradigm for the vaccine prevention of IMD in the region.[108] Group W followed by capsular groups C and X are now the leading cause of IMD in the meningitis belt. Data presented in this work demonstrate the need for a group W vaccine disease in the meningitis belt. Introduction of a polyvalent conjugate vaccine that covers groups A, C, Y, W and ideally X is warranted across the meningitis belt in place of the current A conjugate vaccine. In order for such a vaccine to be introduced, significant commitment of multiple stakeholders will be needed. A cost-effective polyvalent conjugate vaccine can be potentially used beyond the meningitis belt in South Africa, parts of Latin America, and Europe that are facing the emergence of W ST-11 and in China where group C ST-4821 strain has caused several outbreaks.[28,109] In fact, vaccination of teenagers with quadrivalent (A,C,W,Y) vaccine recently occurred in the U.K. because of the an increase in W ST-11 disease.[110]

Protein conjugate vaccines are effective against meningococcal disease and confer herd immunity through reduction in asymptomatic carriage.[111] ST-11 clonal complex contains group C, B and Y strains and there are concerns that global use of conjugate vaccines could potentially lead to selection of capsular switch variants. This is particularly concerning for group B given that there is no vaccine that specifically targets group B capsule. However, use of conjugate vaccines in the UK[112] and United States[13] have not led to increased incidence of capsular switch strains. Outer membrane vesicle vaccines licensed for the prevention of group B disease retain the potential to prevent against non-B capsular groups. One of the vaccines (Bexsero®) contains FHbp variant 1, NadA and Nhba antigens. The Hajj clone contains FHbp variant 1 and both the Hajj clone and non-Hajj W ST-11 strains contain NadA and Nhba suggesting that this vaccine could be protective against W cc11 IMD.[67] A second OMV vaccine formulation (Trumenba®) contains both FHbp variants 1 and 2/3 but not NadA and Nhba. While this vaccine would be effective against both the Hajj clone and non-Hajj W ST-11 strains, there are concerns that vaccination with FHbp alone could lead selection of strains with non-functional FHbp genes observed in a minority of group C ST-11 strains in Canada.[113]

5.2.2 Implication of WGS on molecular epidemiology of IMD.

Molecular epidemiology, defined as the use of molecular approaches to study the distribution and determinants of infectious diseases, complements classic epidemiology in determining disease distribution and determinants.[114] Molecular techniques provide a framework[115] for classification of infectious pathogens based on genetic relationships leading to more reliable measures of associations between host or agent factors and disease. Molecular epidemiologic approaches have helped resolve two global group A pandemics[44,64] and elucidated instances

where increased IMD burden was caused by expansion of endemic strains[116] or introduction of new virulent strains.[63]

Whole genome sequencing is well suited for detailed typing of meningococcal isolates by reference laboratories.[117] Riley[118] described four key attributes of molecular strain typing methods: simplicity, high throughput, cost and appropriateness. WGS has high throughput, providing complete genome sequences for hundreds of isolates while generating data that are shared in a way that subsequent studies are able to compare their isolates to a global strain collection.[117] Also, declining costs are making WGS an even more attractive substitute for conventional typing of surveillance isolates. Drawbacks of WGS are the need for technical expertise and technological infrastructure that is limited in most low resource settings.

In the last five years, use of WGS for *N. meningitidis* has expanded beyond a few reference strains to sequencing of hundreds of isolates from surveillance and carriage studies.[119] Application of WGS has revised our understanding of several hypervirulent meningococcal lineages. In a WGS study, global ST-32 clonal complex was found to consist of three genetically distinct sub-lineages that shared different OMP profiles and geo-temporal characteristics.[120] Notably, this study confirmed that differences in OMP profiles within ST-32 complex observed between isolates from North Europe and Latin America correspond to deeper underlying phylogenetic differences.[120] Likewise, genomic comparison of Y ST-23 strains causing endemic disease in Sweden discovered that a number of distinct sub-lineages YI-YIII were in circulation in Sweden in 1995-2012.[121] Although all three sub-lineages were in circulation over the study period, increased Y ST-23 disease in Sweden 2007-2012 was largely due to a new YI subtype that may have been introduced into Sweden from the United States.[27,121] Interestingly, both the historically endemic and emergent YI subtypes shared identical OMP profiles making them

indistinguishable by conventional molecular typing.[121] Likewise, this work[99] and a similar WGS study[98] demonstrated that despite similarity by OMP genotyping, W ST-11 strains consisted of phylogenetically and antigenically distinct sub-lineages. Remarkably, strains from South America and the UK were closely related and distinct from the Hajj clone and African epidemic strains.[18,98,99] In summary, WGS studies are providing powerful insight into the complex genetic composition of hyperinvasive meningococcal lineages.

5.2.3 Perspective on *Neisseria* population genetics.

This dissertation provides a good case study of the factors at play in the emergence and persistence of hypervirulent meningococcal lineages. Meningococcal clonal complexes are hypothesized to possess a number of mutually co-adapted set of alleles that persist in a dynamic state of genomic flux over time[25,122] and maintain a discrete population structure despite extensive recombination.[100,115] Minor genetic variants arise from major lineages by recombination with most recombinant strains being less fit than the progenitor strain.[105] Rarely, a genetic variant escapes the fitness constraint and persist alongside or may replace the original dominant strain.[64]

Persistence of a single W ST-11 capsular allelic type suggest that the capsular switching event represented a bottleneck event with selective expansion of a single capsular genotype at the expense of other intermediate variants and strains with alternative group W capsular recombination events. Although, the presence of a capsule was a necessary requirement for meningococcal virulence,[2] in the case of W ST-11 the capsule alone was not a sufficient cause of virulence given that historical W ST-11 strains rarely caused disease. Therefore, capsule acquisition represented one important component in a multi-step process of virulence acquisition.

A second genetic bottleneck was associated with the selection and rapid global dissemination of a single clone during the Hajj epidemic. Strains belonging to the Hajj clone were genetically homogenous with very limited genetic and antigenic diversity compared to extensive diversity among non-Hajj W ST-11 strains from both before and after the Hajj epidemic.[45,58,61,98,99] This work demonstrated that the Hajj clone had acquired a unique surface antigen gene (*fHbp*) and a number of genetic elements such as *aniA* and *nor* genes associated with nasopharyngeal survival of meningococci. These genetic factors may have conferred a fitness advantage that led to rapid clonal expansion of the Hajj clone at the expense of other W ST-11 strains. Subsequently, the Hajj clone was introduced and co-existed within a background of genetically heterogeneous endogenous W ST-11 strains in the meningitis belt and South Africa. Instead of replacing endogenous W ST-11 strains in these countries, the Hajj clone co-circulated and co-existed with these endogenous W ST-11 strains. It is not known whether the Hajj clone predominantly affected a different demographic or existed in different provinces within these countries compared endogenous W ST-11 strains which could explain co-circulation without direct competition among W ST-11 strains.[123]

Factors beyond the pathogen genetic composition play a substantial role in determining the epidemiological characteristics of meningococcal strains. Population susceptibility to virulent meningococcal strains is dependent on immunity acquired through natural exposure to circulating meningococci or through vaccination, crowding, weather conditions, age composition, and smoking.[2,8,124-126] Spread of the Hajj clone provides an interesting example of an identical meningococcal strain that when introduced into different human populations led to epidemics in the meningitis belt,[17,66,73] hyperendemic cases in South Africa,[52] and small localized outbreaks in the UK[14,43] and France.[61] Likewise, introduction of the Hajj clone in the UK in

2000-2001 caused much fewer IMD cases than the introduction of the 'South American' W ST-11 strain in 2010-2014 which required targeted vaccination of teenagers.[110] It is of concern whether the continued spread of the South American W ST-11 strain could lead to meningitis belt epidemics. Therefore, environmental and behavioral factors play a central role in the emergence of virulent meningococci and assessment of virulence must take cognizance of the particular epidemiologic context.

5.3 LIMITATIONS

A limitation of this study is that invasive strains represent a highly biased sub-population of isolates that may not reflect true meningococcal population genetic diversity.[126,127] Also, IMD surveillance in some global regions is weak and only a limited number of historical isolates had WGS. However, our results are supported by extensive WGS data for hundreds of IMD surveillance isolates from UK and South Africa. Another limitation of global comparison of IMD isolates is the difficulty in dis-entangling the contribution of differing socioeconomic, behavioral, and climatic factors to observed differences in disease incidence.

5.4 FUTURE DIRECTIONS

This dissertation explores the molecular and genomic epidemiology of meningococcal disease using emergent W ST-11 strain as a case example. Bioinformatics methods applied in this work can be applied to the study of other emerging meningococcal lineages such as group C ST-4821 causing epidemics in China,[28] group Y ST-23 strains associated with increasing endemic disease in the United States[35] and across Europe[22] and emerging group C ST-10217 in parts of the meningitis belt.[15] *Neisseria* PubMLST database has facilitated several comparative genomic studies including that reported in this dissertation. However, the potential of this database for more extensive genomic comparison between meningococci and among other *Neisseria* species

is yet to be fully realized. Strengthened IMD surveillance on a global scale and wider application of WGS is needed to monitor the continuing evolution of W ST-11. There is also a need for an affordable multivalent protein conjugate vaccine that prevents groups W and C meningococcal disease in the meningitis belt.

APPENDIX A: SUPPLEMENTARY MATERIAL FOR PAPER 2

Supplementary Methods

DNA extraction: Isolates for DNA extraction were subcultured from a single colony and grown overnight at 37°C with 5% CO_2 on chocolate agar plates. DNA for Sanger sequencing was extracted by a NucliSENs easyMAG system (bioMérieux, Durham, NC) and by boiling a heavy suspension of organisms in 0.5mL of PBS. DNA for Illumina and Ion Torrent sequencing was extracted by using the DNeasy blood and tissue kit (QIAGEN, Valencia, CA). DNA for PacBio sequencing was extracted using the Easy-DNA gDNA Purification Kit (Invitrogen, Grand Island, NY).

Multilocus sequene typing (MLST) and outer membrane protein (OMP) gene sequencing: For all 26 sequenced genomes PCR based MLST and OMP gene sequence typing of PorA VR1 and VR2, and FetA VR were performed as described.[128] Sequence alleles and sequence types were determined by querying the PubMLST database. MLST and OMP results were compared to those obtained by querying assembled contigs on PubMLST database. MLST and OMP typing for public genomes were obtained by querying genome sequences against the PubMLST database.

16S rRNA gene sequencing: PCR amplification and sequence analysis of the 16S rRNA genes were performed using modifications of published methods.[42,87] Primers 8F and 1492R were used for amplification. All reactions were carried out in 50-µl volumes containing 1 µl of purified genomic DNA (~20 ng), 1× AmpliTaq Gold PCR buffer (15 mM Tris-HCl, pH 8.0, 50 mM KCl), 2.5 mM $MgCl_2$, 0.2 mM each deoxynucleoside triphosphate, 0.2 µM of each primer, and

1.5 units AmpliTaq Gold DNA polymerase (Applied Biosystems, Foster City, CA). Cycling conditions were an initial denaturation step at 95°C for 5 min followed by 35 cycles of 95°C for 1 min, 50°C for 1 min and 72°C for 1.5 min, followed by a final extension step for 7 min at 72°C. PCR products were purified with Exo-Sap It (Affymetrix, Santa Clara, CA) and sequenced with BigDye Terminator v3.1 Cycle Sequencing Kit according to manufacturer's instructions using previously described primers 8F, 1492R, 357, 530, 790, 981, 1968F and 1083F.[42,87] Sequence contigs were analyzed and compared to 16S type 13 and 31 by using DNAstar Lasergene SeqMan Pro software (v.11.2; DNAstar, Madison, WI).

SNP-based whole genome phylogenetic tree (Figure 24) was generated by aligning sequence reads of 153 *N. meningitidis* genome sequences against *M7124* genome. Aligned reads were sorted and filtered as descrided in Methods. 140,903 SNP positions were concatenated and a maximum likelihood tree generated under General Time Reversible model, gamma distribution of rate variation with invariant sites (GTR+Γ+I).

Quality assessment: Several steps were taken during isolate selection, genome assembly, and data analyses to assure data quality and consistency of results. For all 26 newly sequenced genomes at least three iterations of genome assemblies were run and the the assembly with the best assembly quality metrics (low assembly gaps, high average median contig size and larger total genome size) is selected. Assembly quality was further assessed by comparing newly sequenced genomes to both group C ST-11 reference strain, *FAM18* and *M7124* using multiple genome alignments on Mauve v2.3. In addition, read mapping and SNP analyses were used to compare data obtained from different platforms for re-sequenced isolates. No major discrepancies in the form of insertion-deletion, translocation or inconsistent genome sizes were identified.

Three isolates that originally had IonTorrent data were re-sequenced using PacBio to assess whether the relatively large number of assembly gaps obtained from IonTorrent affected phylogenetic clustering of these isolates. Re-sequencing did not change the phylogenetic relationships between strains in our study (data not shown). To ascertain that only good quality genome assemblies were included from PubMLST database, genomes that had incomplete data at one of six antigen gene alleles were excluded from the study.

For whole genome phylogenetic analyses, isolates were selected to capture geographic and antigen gene allelic diversity of the entire set of 270 isolates. Among isolates sharing similar allelic and geographic profiles, one with the best genome assembly parameters was selected as a representative strain. 'Iterative refinement' option was selected to improve genome alignment accuracy while poorly aligned genomic regions were excluded by including only core genome segments that were conserved in all aligned genomes and were at least 500bp in size. The resultant core genome Maximum Likelihood phylogenetic tree remained unchanged after increasing the minimum alignment block size to 2000bp and/or removing very small alignment gaps not addressed by Mauve using GBlocks (results not shown). SNP analysis was limited to 20 genomes that had short read sequencing data (Illumina, IonTorrent, 454). PacBio reads were not used for SNP calling because the tools we used (bwa, samtools) were not designed to handle long read sequence data.

Finally, we compared results of separate phylogenetic trees constructed using ClonalFrame, SptitsTree and Maximum Likelihood to allelic profiles of antigen encoding genes. Concordance between these data and available epidemiologic information for these isolates provided further reassurance of the validity of these results.

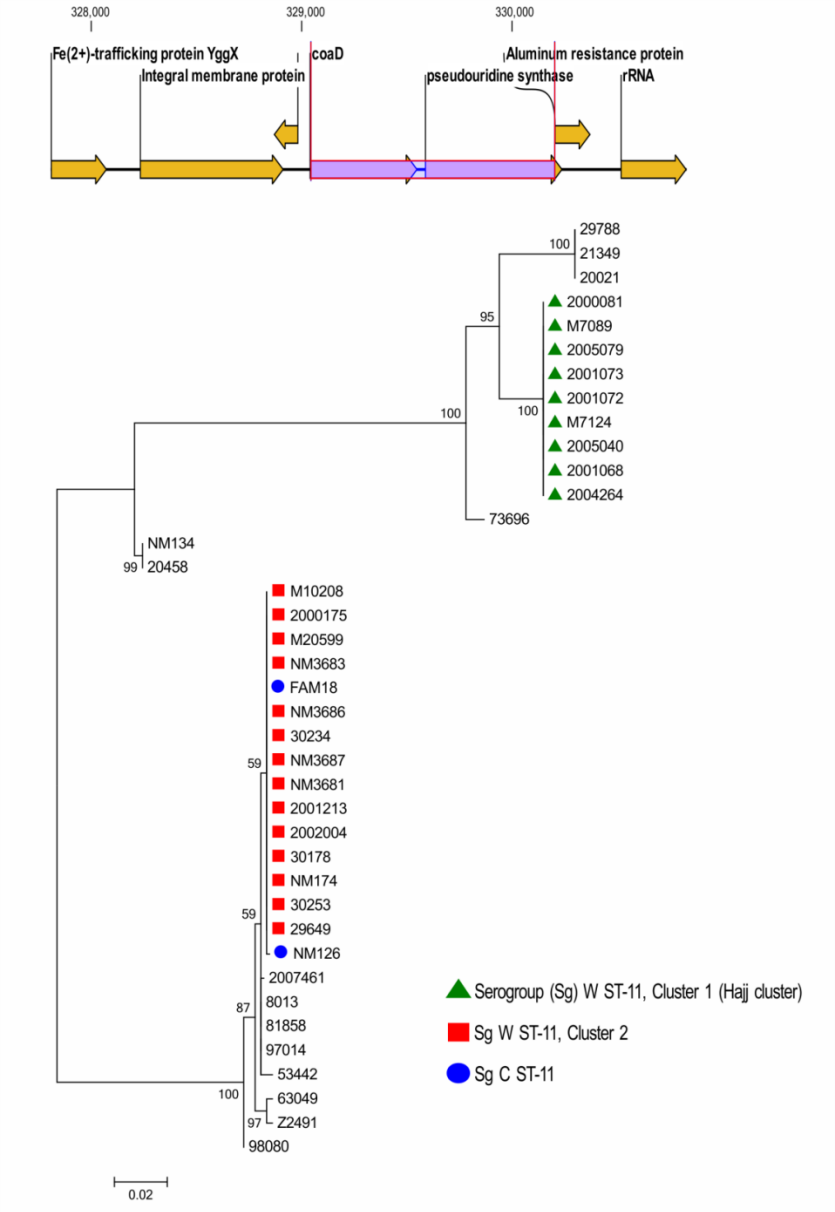

Figure 20: Phylogenetic tree of Hajj cluster recombinant region 1.

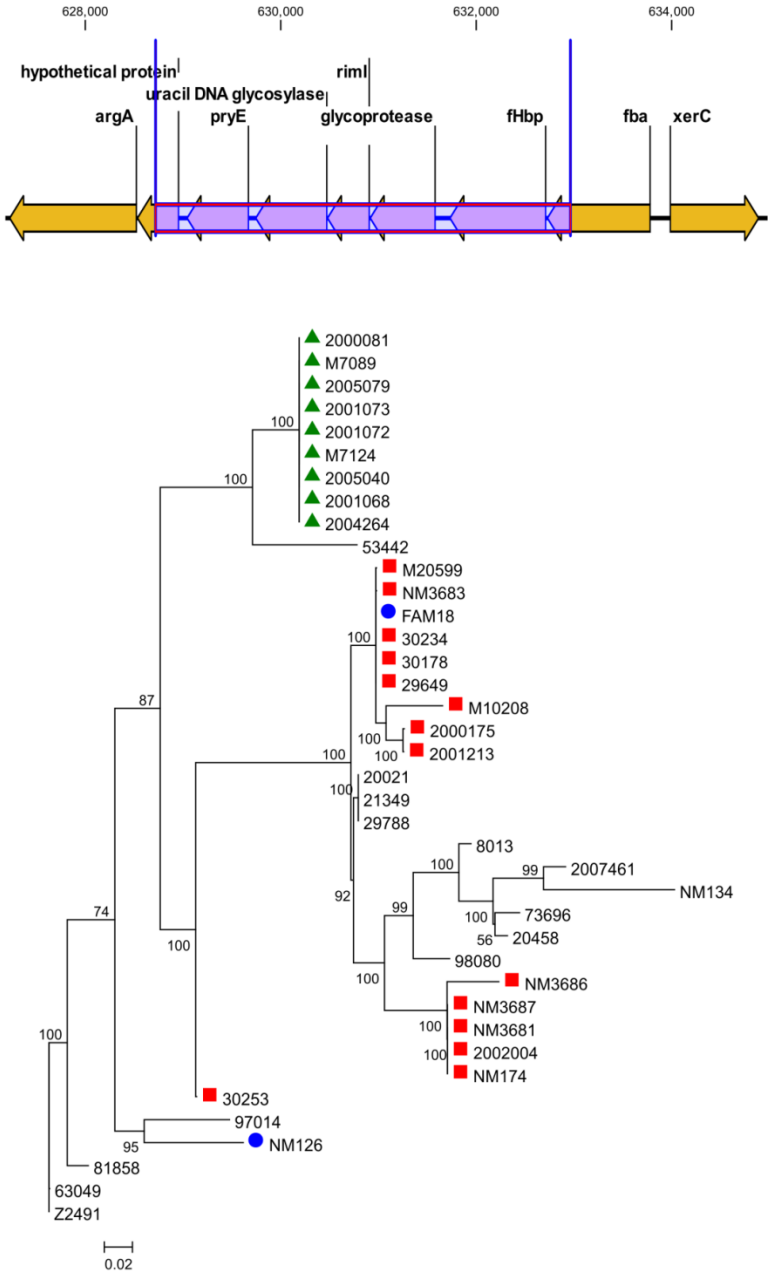

Figure 21: Phylogenetic tree of Hajj cluster recombinant region 2.

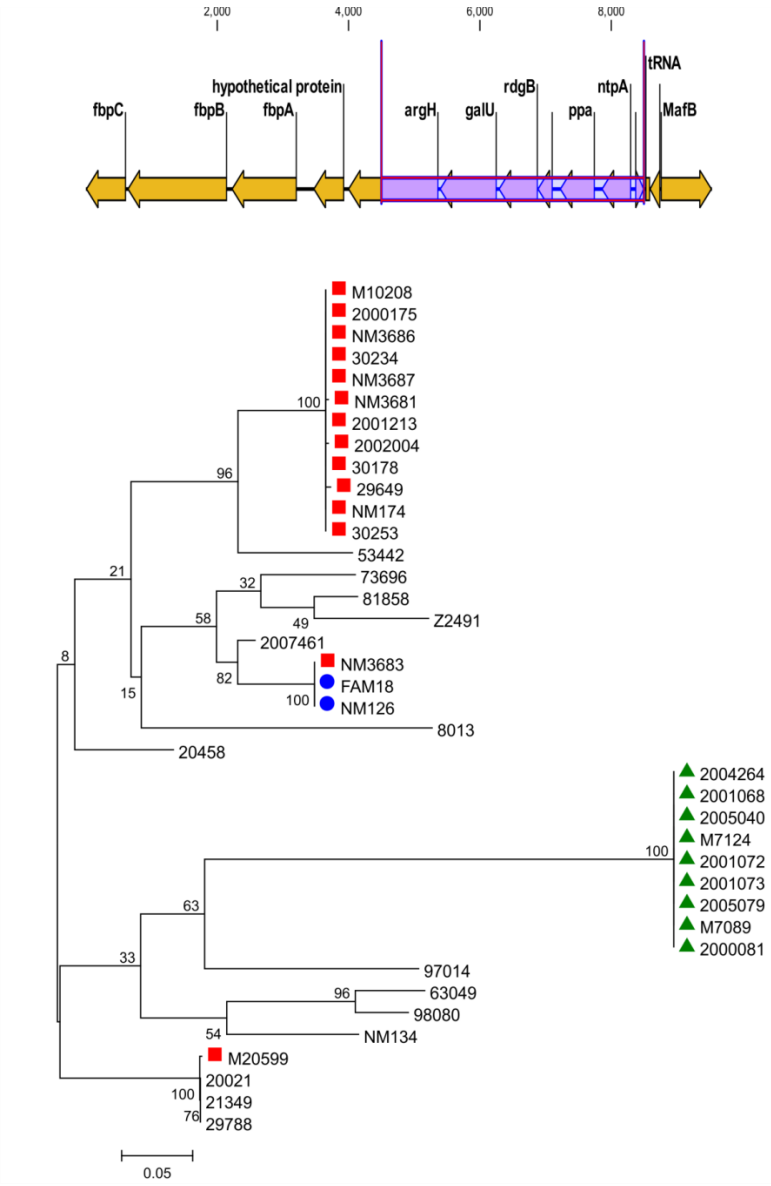

Figure 22: Phylogenetic tree of Hajj cluster recombinant region 3.

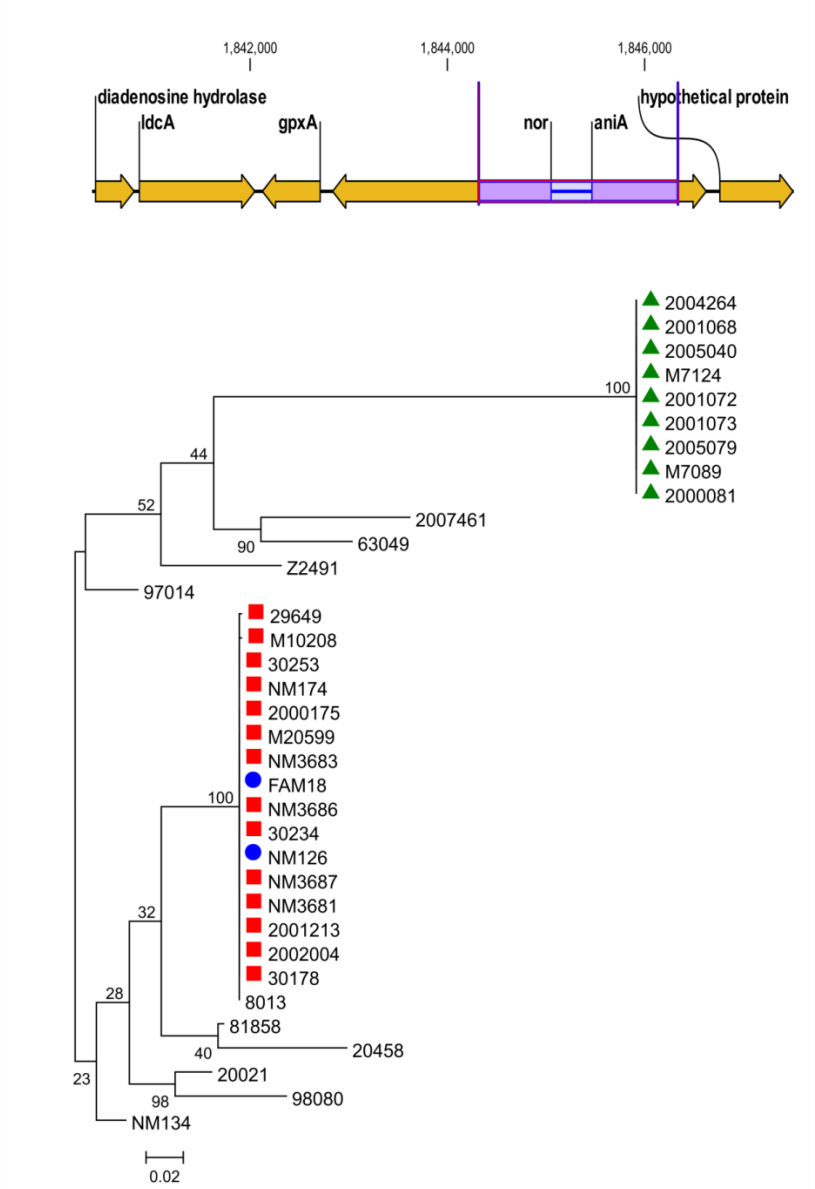

Figure 23: Phylogenetic tree of Hajj cluster recombinant region 4.

Figure 24: Phylogenetic tree of serogroup W, C and B ST-11 strains in relation to other hyperinvasive meningococcal lineages.

Table 2. Characteristics of 26 meningococcal genomes sequenced in this study.

Strain ID	Sequencing technology	GC content	Coverage	Assembly method	Number of contigs	GenBank Accession No.	Genome size	Largest contig size	Median contig size	Smallest contig size	N50[1]	N90[2]
M7124	PacBio[3]	51.7	130.9	HGAP v2.0	1	CP009419	2179483	-	-	-	-	-
NM3686	PacBio	51.7	173.0	HGAP v2.0	1	CP009418	2195266	-	-	-	-	-
NM3687	PacBio	51.6	219.8	HGAP v2.0	10	JRET00000000	2223230	1610344	47059	4073	1609697	103326
NM3681	PacBio	51.7	62.1	HGAP v2.0	2	JREU00000000	2217098	1380815	415099	6085	1380815	12054
NM3682	PacBio	51.7	60.0	HGAP v2.0	1	CP009420	2185359	-	-	-	-	-
M10208	PacBio	51.7	55.0	HGAP v2.0	1	CP009422	2183230	-	-	-	-	-
M20599	PacBio	51.8	60.0	HGAP v2.0	13	JRFG00000000	2258407	1053953	19069	4563	516162	150267
M1412	Ion Torrent[4]	51.7	62.6	Mira v3.9	320	JRFH00000000	2110370	39255	4109	504	12232	3559
M12611	Ion Torrent	51.8	70.8	Mira v3.9	698	JRFI00000000	1914608	13374	2130	10001	3305	1405
M7089	Illumina+454[5]	51.7	32.0	Celera v7.0	31	ANRN00000000	2174132	280052	51414	1714	133850	49971
2001213	Illumina[6]	51.9	54.7	Celera v7.0	125	APUD00000000	2145367	107785	10070	516	29037	8528
2004264	Illumina	52.0	57.6	Celera v7.0	132	APUE00000000	2223810	87880	11580	179	27158	8339
2005079	Illumina	52.0	61.6	Celera v7.0	132	APUG00000000	2121623	78339	10799	136	27213	8691
2005040	Illumina	52.0	62.4	Celera v7.0	133	APUI00000000	2156155	169873	10487	146	27155	9111
2001072	Illumina	51.9	75.5	Celera v7.0	131	APTY00000000	2157792	124299	10473	120	26510	8615
2001073	Illumina	51.9	79.4	Celera v7.0	137	APTZ00000000	2143674	86269	10192	118	28492	8200
2002004	Illumina	51.9	55.9	Celera v7.0	137	APUJ00000000	2256966	68617	12347	133	26724	8625
2000081	Illumina	51.9	60.4	Celera v7.0	140	APUA00000000	2180008	63594	10344	101	26618	8225
NM174	Illumina+454	51.7	43.1	Celera v7.0	19	ANRP00000000	2180278	771479	80215	1914	161674	67157
2000175	Illumina	51.9	51.4	Celera v7.0	134	APUF00000000	2233577	93253	11468	224	25838	8910
2001068	Illumina	51.9	70.2	Celera v7.0	136	APUH00000000	2214688	94883	11879	648	24221	8778
NM3680	Ion Torrent	51.8	63.2	Mira v3.9	250	JSAR00000000	2101022	65014	5616	507	15072	4516
NM3684	Ion Torrent	51.7	64.1	Mira v3.9	261	JSAS00000000	2132590	55229	5131	497	16290	4425
NM3685	Ion Torrent	51.7	47.3	Mira v3.9	280	JSAT00000000	2122125	49178	4782	497	13937	3925
NM3688	Ion Torrent	51.6	63.7	Mira v3.9	249	JSAU00000000	2165310	59342	5196	501	17223	4867
NM3147	Illumina	52.0	81.06	Celera v7.0	136	APTX00000000	2256933	94650	10744	101	29024	8163

[1]N50: 50% of the entire genome is contained in contigs equal to or larger than this value.
[2]N90: 90% of the entire genome is contained in contigs equal to or larger than this value; HGAP: Hierarchical genome assembly process.
[3]Pacific BioSciences RS II sequencing (www.pacificbiosciences.com/)
[4]Life Technologies Ion Torrent PGM sequencing (www.lifetechnologies.com)
[5]Roche 454 sequencing (www.454.com)
[6]Paired end Illumina MiSEQ (www.illumina.com/)

Table 3. List of six Cluster 1 isolates that shared only three out of four recombinant regions with M7124.

Isolate ID	Country	Year	Recombinant region*	Percent identity	Alignment length	Mismatches	Gaps
21583	South Africa	2011	1	82.98	1163	178	5
29326	South Africa	2005	1	82.98	1163	178	5
29336	South Africa	2010	1	82.98	1163	178	5
29402	South Africa	2013	1	82.98	1163	178	5
29387	South Africa	2013	3	99.1	4016	34	1
29393	South Africa	2008	4	98.23	2031	36	0

*Recombinant region numbers correspond to those on Table 1

BIBLIOGRAPHY

1. Halperin SA, Bettinger JA, Greenwood B, et al. The changing and dynamic epidemiology of meningococcal disease. Vaccine 2012;30 Suppl 2:B26-36.
2. Stephens DS. Biology and pathogenesis of the evolutionarily successful, obligate human bacterium *Neisseria meningitidis*. Vaccine 2009;27 Suppl 2:B71-7.
3. Stephens DS, Greenwood B, Brandtzaeg P. Epidemic meningitis, meningococcaemia, and *Neisseria meningitidis*. Lancet 2007;369:2196-210.
4. Rosenstein NE, Perkins BA, Stephens DS, Popovic T, Hughes JM. Meningococcal disease. The New England Journal of Medicine 2001;344:1378-88.
5. Harrison LH, Trotter CL, Ramsay ME. Global epidemiology of meningococcal disease. Vaccine 2009;27 Suppl 2:B51-63.
6. Jafri RZ, Ali A, Messonnier NE, et al. Global epidemiology of invasive meningococcal disease. Population health metrics 2013;11:17.
7. Weekly feedback bulletin on cerebrospinal meningitis. Meningitis Weekly Bulletin. www.who.int/csr/disease/meningococcal/epidemiological/en/index.html; www.meningvax.org/epidemic-updates.php; : World Health Organization; 2002-2015.
8. Lapeyssonnie L. [Comparative epidemiologic study of meningococcic cerebrospinal meningitis in temperate regions and in the meningitis belt in Africa. Attempt at synthesis]. Med Trop (Mars) 1968;28:709-20.
9. Rouphael NG, Stephens DS. *Neisseria meningitidis*: biology, microbiology, and epidemiology. Methods Mol Biol 2012;799:1-20.
10. Harrison OB, Claus H, Jiang Y, et al. Description and nomenclature of *Neisseria meningitidis* capsule locus. Emerg Infect Dis 2013;19:566-73.
11. Teyssou R, Muros-Le Rouzic E. Meningitis epidemics in Africa: a brief overview. Vaccine 2007;25 Suppl 1:A3-7.
12. The epidemiology of meningococcal disease in New Zealand 2008-2013. Institute of Environmental Science and Research Limited, 2014. (Accessed May 1, 2015, 2015, at https://surv.esr.cri.nz/surveillance/Meningococcal_disease.php.)
13. Wang X, Shutt KA, Vuong JT, et al. Changes in the Population Structure of Invasive N*eisseria meningitidis* in the United States After Quadrivalent Meningococcal Conjugate Vaccine Licensure. J Infect Dis 2015.
14. Meningococcal Reference U, Gray SJ, Trotter CL, et al. Epidemiology of meningococcal disease in England and Wales 1993/94 to 2003/04: contribution and experiences of the Meningococcal Reference Unit. J Med Microbiol 2006;55:887-96.
15. Funk A, Uadiale K, Kamau C, Caugant DA, Ango U, Greig J. Sequential outbreaks due to a new strain of *Neisseria meningitidis* serogroup C in northern Nigeria, 2013-14. PLoS Curr 2014;6.
16. Taha MK, Achtman M, Alonso JM, et al. Serogroup W135 meningococcal disease in Hajj pilgrims. Lancet 2000;356:2159.
17. MacNeil JR, Medah I, Koussoube D, et al. *Neisseria meningitidis* serogroup W, Burkina Faso, 2012. Emerg Infect Dis 2014;20:394-9.
18. Ladhani SN, Beebeejaun K, Lucidarme J, et al. Increase in endemic *Neisseria meningitidis* capsular group W sequence type 11 complex associated with severe invasive disease in England and Wales. Clin Infect Dis 2015;60:578-85.

19. Sorhouet-Pereira C, Efron A, Gagetti P, et al. Phenotypic and genotypic characteristics of *Neisseria meningitidis* disease-causing strains in Argentina, 2010. PloS One 2013;8:e58065.
20. Harrison LH, Jolley KA, Shutt KA, et al. Antigenic shift and increased incidence of meningococcal disease. J Infect Dis 2006;193:1266-74.
21. Hedberg ST, Toros B, Fredlund H, Olcen P, Molling P. Genetic characterisation of the emerging invasive *Neisseria meningitidis* serogroup Y in Sweden, 2000 to 2010. Euro Surveillance : Bulletin Europeen sur les Maladies Transmissibles = European Communicable Disease Bulletin 2011;16.
22. Broker M, Jacobsson S, Kuusi M, et al. Meningococcal serogroup Y emergence in Europe: update 2011. Human vaccines & immunotherapeutics 2012;8:1907-11.
23. Materu S, Cox HS, Isaakidis P, Baruani B, Ogaro T, Caugant DA. Serogroup X in meningococcal disease, Western Kenya. Emerg Infect Dis 2007;13:944-5.
24. Delrieu I, Yaro S, Tamekloe TA, et al. Emergence of epidemic *Neisseria meningitidis* serogroup X meningitis in Togo and Burkina Faso. PloS One 2011;6:e19513.
25. Maiden MC, Bygraves JA, Feil E, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 1998;95:3140-5.
26. Pizza M, Rappuoli R. *Neisseria meningitidis*: pathogenesis and immunity. Curr Opin Microbiol 2015;23:68-72.
27. Krauland MG, Dunning Hotopp JC, Riley DR, et al. Whole genome sequencing to investigate the emergence of clonal complex 23 *Neisseria meningitidis* serogroup Y disease in the United States. PloS One 2012;7:e35699.
28. Peng J, Yang L, Yang F, et al. Characterization of ST-4821 complex, a unique *Neisseria meningitidis* clone. Genomics 2008;91:78-87.
29. Marri PR, Paniscus M, Weyand NJ, et al. Genome sequencing reveals widespread virulence gene exchange among human *Neisseria* species. PloS One 2010;5:e11835.
30. Rappuoli R. Reverse vaccinology, a genome-based approach to vaccine development. Vaccine 2001;19:2688-91.
31. Jolley KA, Maiden MC. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 2010;11:595.
32. Bratcher HB, Bennett JS, Maiden MC. Evolutionary and genomic insights into meningococcal biology. Future microbiology 2012;7:873-85.
33. Hao W, Ma JH, Warren K, et al. Extensive genomic variation within clonal complexes of *Neisseria meningitidis*. Genome Biology and Evolution 2011;3:1406-18.
34. Swartley JS, Marfin AA, Edupuganti S, et al. Capsule switching of *Neisseria meningitidis*. Proc Natl Acad Sci U S A 1997;94:271-6.
35. Harrison LH, Shutt KA, Schmink SE, et al. Population structure and capsular switching of invasive *Neisseria meningitidis* isolates in the pre-meningococcal conjugate vaccine era--United States, 2000-2005. J Infect Dis 2010;201:1208-24.
36. Beddek AJ, Li MS, Kroll JS, Jordan TW, Martin DR. Evidence for capsule switching between carried and disease-causing *Neisseria meningitidis* strains. Infection and Immunity 2009;77:2989-94.
37. Mueller JE, Borrow R, Gessner BD. Meningococcal serogroup W135 in the African meningitis belt: epidemiology, immunity and vaccines. Expert Rev Vaccines 2006;5:319-36.
38. Rosenstein NE, Perkins BA, Stephens DS, et al. The changing epidemiology of meningococcal disease in the United States, 1992-1996. J Infect Dis 1999;180:1894-901.
39. Squires SG, Deeks SL, Tsang RS. Enhanced surveillance of invasive meningococcal disease in Canada: 1 January, 1999, through 31 December, 2001. Can Commun Dis Rep 2004;30:17-28.
40. Guibourdenche M, Hoiby EA, Riou JY, Varaine F, Joguet C, Caugant DA. Epidemics of serogroup A *Neisseria meningitidis* of subgroup III in Africa, 1989-94. Epidemiol Infect 1996;116:115-20.
41. Kwara A, Adegbola RA, Corrah PT, et al. Meningitis caused by a serogroup W135 clone of the ET-37 complex of *Neisseria meningitidis* in West Africa. Trop Med Int Health 1998;3:742-6.

42. Lemos AP, Harrison LH, Lenser M, Sacchi CT. Phenotypic and molecular characterization of invasive serogroup W135 *Neisseria meningitidis* strains from 1990 to 2005 in Brazil. The Journal of Infection 2010;60:209-17.
43. Aguilera JF, Perrocheau A, Meffre C, Hahne S, Group WW. Outbreak of serogroup W135 meningococcal disease after the Hajj pilgrimage, Europe, 2000. Emerg Infect Dis 2002;8:761-7.
44. Moore PS, Reeves MW, Schwartz B, Gellin BG, Broome CV. Intercontinental spread of an epidemic group A *Neisseria meningitidis* strain. Lancet 1989;2:260-3.
45. Mayer LW, Reeves MW, Al-Hamdan N, et al. Outbreak of W135 meningococcal disease in 2000: not emergence of a new W135 strain but clonal expansion within the electophoretic type-37 complex. J Infect Dis 2002;185:1596-605.
46. Doyle TJ, Mejia-Echeverry A, Fiorella P, et al. Cluster of serogroup W135 meningococci, southeastern Florida, 2008-2009. Emerg Infect Dis 2010;16:113-5.
47. Taha MK, Kacou-N'douba A, Hong E, et al. Travel-related *Neisseria meningitidis* serogroup W135 infection, France. Emerg Infect Dis 2013;19:1030-2.
48. Araya P, Fernandez J, Del Canto F, et al. *Neisseria meningitidis* ST-11 clonal complex, Chile 2012. Emerg Infect Dis 2015;21:339-41.
49. Zhou H, Liu W, Xu L, et al. Spread of *Neisseria meningitidis* serogroup W clone, China. Emerg Infect Dis 2013;19:1496-9.
50. Taha MK, Parent Du Chatelet I, Schlumberger M, et al. *Neisseria meningitidis* serogroups W135 and A were equally prevalent among meningitis cases occurring at the end of the 2001 epidemics in Burkina Faso and Niger. Journal of clinical microbiology 2002;40:1083-4.
51. Kristiansen PA, Diomande F, Ba AK, et al. Impact of the Serogroup A Meningococcal Conjugate Vaccine, MenAfriVac, on Carriage and Herd Immunity. Clin Infect Dis 2012.
52. von Gottberg A, du Plessis M, Cohen C, et al. Emergence of endemic serogroup W135 meningococcal disease associated with a high mortality rate in South Africa. Clin Infect Dis 2008;46:377-86.
53. Efron AM, Sorhouet C, Salcedo C, Abad R, Regueira M, Vazquez JA. W135 invasive meningococcal strains spreading in South America: significant increase in incidence rate in Argentina. Journal of clinical microbiology 2009;47:1979-80.
54. Barra GN, Araya PA, Fernandez JO, et al. Molecular Characterization of Invasive *Neisseria meningitidis* Strains Isolated in Chile during 2010-2011. PloS One 2013;8:e66006.
55. Kilic A, Urwin R, Li H, Saracli MA, Stratton CW, Tang YW. Clonal spread of serogroup W135 meningococcal disease in Turkey. Journal of clinical microbiology 2006;44:222-4.
56. Parent du Chatelet I, Barboza P, Taha MK. W135 invasive meningococcal infections imported from Sub-Saharan Africa to France, January to April 2012. Euro Surveillance : Bulletin Europeen sur les Maladies Transmissibles = European Communicable Disease Bulletin 2012;17.
57. Fonkoua MC, Taha MK, Nicolas P, et al. Recent increase in meningitis Caused by *Neisseria meningitidis* serogroups A and W135, Yaounde, Cameroon. Emerg Infect Dis 2002;8:327-9.
58. Molling P, Backman A, Olcen P, Fredlund H. Comparison of serogroup W-135 meningococci isolated in Sweden during a 23-year period and those associated with a recent hajj pilgrimage. Journal of clinical microbiology 2001;39:2695-9.
59. Kelly D, Pollard AJ. W135 in Africa: origins, problems and perspectives. Travel Medicine and Infectious Disease 2003;1:19-28.
60. Njanpop-Lafourcade BM, Parent du Chatelet I, Sanou O, Alonso JM, Taha MK. The establishment of *Neisseria meningitidis* serogroup W135 of the clonal complex ET-37/ST-11 as an epidemic clone and the persistence of serogroup A isolates in Burkina Faso. Microbes and Infection / Institut Pasteur 2005;7:645-9.
61. Taha MK, Giorgini D, Ducos-Galand M, Alonso JM. Continuing diversification of *Neisseria meningitidis* W135 as a primary cause of meningococcal disease after emergence of the serogroup in 2000. Journal of clinical microbiology 2004;42:4158-63.

62. Buckee CO, Jolley KA, Recker M, et al. Role of selection in the emergence of lineages and the evolution of virulence in *Neisseria meningitidis*. Proc Natl Acad Sci U S A 2008;105:15082-7.
63. Chiou CS, Liao JC, Liao TL, et al. Molecular epidemiology and emergence of worldwide epidemic clones of *Neisseria meningitidis* in Taiwan. BMC Infectious Diseases 2006;6:25.
64. Zhu P, van der Ende A, Falush D, et al. Fit genotypes and escape variants of subgroup III *Neisseria meningitidis* during three pandemics of epidemic meningitis. Proc Natl Acad Sci U S A 2001;98:5234-9.
65. Memish ZA, Shibl AM. Consensus building and recommendations based on the available epidemiology of meningococcal disease in Gulf Cooperation Council States. Travel Medicine and Infectious Disease 2011;9:60-6.
66. Decosas J, Koama JB. Chronicle of an outbreak foretold: meningococcal meningitis W135 in Burkina Faso. The Lancet Infectious Diseases 2002;2:763-5.
67. Pajon R, Fergus AM, Koeberling O, Caugant DA, Granoff DM. Meningococcal factor H binding proteins in epidemic strains from Africa: implications for vaccine development. PLoS Neglected Tropical Diseases 2011;5:e1302.
68. Ibarz-Pavon AB, Morais L, Sigauque B, et al. Epidemiology, molecular characterization and antibiotic resistance of *Neisseria meningitidis* from patients </=15 years in Manhica, rural Mozambique. PloS One 2011;6:e19717.
69. Weidlich L, Baethgen LF, Mayer LW, et al. High prevalence of *Neisseria meningitidis* hypervirulent lineages and emergence of W135:P1.5,2:ST-11 clone in Southern Brazil. The Journal of Infection 2008;57:324-31.
70. Li YA, Tsang R, Desai S, Deehan H. Enhanced surveillance of invasive meningococcal disease in Canada, 2006-2011. Canada Communicable Disease Report CCDR 2014;40.
71. Wilder-Smith A, Chow A, Goh KT. Emergence and disappearance of W135 meningococcal disease. Epidemiol Infect 2010;138:976-8.
72. Australian Meningococcal Surveillance Programme annual reports. 2014. (Accessed April 30, 2015, 2015, at www.health.gov.au/internet/main/publishing.nsf/Content/cda-pubs-annlrpt-menganrep.htm.)
73. Collard JM, Maman Z, Yacouba H, et al. Increase in *Neisseria meningitidis* serogroup W135, Niger, 2010. Emerg Infect Dis 2010;16:1496-8.
74. Mustapha M. Mustapha JWM, Jorge O. Fernandez, Ana Paula S. de Lemos, Xin Wang, Leonard W. Mayer, Lee H. Harrison. Capsular Switching and Global Spread of *Neisseria Meningitidis* Serogroup W, ST-11. International Pathogenic *Neisseria* Conferences, IPNC; 2014; Asheville, North Carolina, USA: IPNC.
75. Katz LS, Humphrey JC, Conley AB, et al. *Neisseria* Base: a comparative genomics database for *Neisseria meningitidis*. Database : The Journal of Biological Databases and Curation 2011;2011:bar035.
76. Chin CS, Alexander DH, Marks P, et al. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nature Methods 2013;10:563-9.
77. Chevreux B, Pfisterer T, Drescher B, et al. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Research 2004;14:1147-59.
78. Aziz RK, Bartels D, Best AA, et al. The RAST Server: rapid annotations using subsystems technology. BMC Genomics 2008;9:75.
79. Darling AE, Mau B, Perna NT. ProgressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PloS One 2010;5:e11147.
80. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology 2010;59:307-21.
81. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Molecular Biology and Evolution 2006;23:254-67.

82. Vos M, Didelot X. A comparison of homologous recombination rates in bacteria and archaea. The ISME Journal 2009;3:199-208.
83. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 2011;28:2731-9.
84. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 2013;30:772-80.
85. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.
86. Field D, Tiwari B, Booth T, et al. Open software for biologists: from famine to feast. Nature Biotechnology 2006;24:801-3.
87. Sacchi CT, Whitney AM, Reeves MW, Mayer LW, Popovic T. Sequence diversity of *Neisseria meningitidis* 16S rRNA genes and use of 16S rRNA gene sequencing as a molecular subtyping tool. Journal of clinical microbiology 2002;40:4520-7.
88. Koumare B, Ouedraogo-Traore R, Sanou I, et al. The first large epidemic of meningococcal disease caused by serogroup W135, Burkina Faso, 2002. Vaccine 2007;25 Suppl 1:A37-41.
89. Novak RT, Kambou JL, Diomande FV, et al. Serogroup A meningococcal conjugate vaccination in Burkina Faso: analysis of national surveillance data. The Lancet Infectious Diseases 2012;12:757-64.
90. Mothibeli KM, du Plessis M, von Gottberg A, et al. Distribution of factor H binding protein beyond serogroup B: variation among five serogroups of invasive *Neisseria meningitidis* in South Africa. Vaccine 2011;29:2187-92.
91. Valenzuela MT, Moreno G, Vaquero A, et al. [Emergence of W135 meningococcal serogroup in Chile during 2012]. Revista medica de Chile 2013;141:959-67.
92. Beernink PT, Caugant DA, Welsch JA, Koeberling O, Granoff DM. Meningococcal factor H-binding protein variants expressed by epidemic capsular group A, W-135, and X strains from Africa. J Infect Dis 2009;199:1360-8.
93. Granoff DM. Review of meningococcal group B vaccines. Clin Infect Dis 2010;50 Suppl 2:S54-65.
94. Harrison LH, Jolley KA, Shutt KA, et al. Antigenic shift and increased incidence of meningococcal disease. J Infect Dis 2006;193:1266-74.
95. Anjum MF, Stevanin TM, Read RC, Moir JW. Nitric oxide metabolism in *Neisseria meningitidis*. Journal of Bacteriology 2002;184:2987-93.
96. Stevanin TM, Moir JW, Read RC. Nitric oxide detoxification systems enhance survival of *Neisseria meningitidis* in human macrophages and in nasopharyngeal mucosa. Infection and Immunity 2005;73:3322-9.
97. Mackinnon FG, Borrow R, Gorringe AR, Fox AJ, Jones DM, Robinson A. Demonstration of lipooligosaccharide immunotype and capsule as virulence factors for *Neisseria meningitidis* using an infant mouse intranasal infection model. Microb Pathog 1993;15:359-66.
98. Lucidarme J, Hill DM, Bratcher HB, et al. Genomic resolution of an aggressive, widespread, diverse and expanding meningococcal serogroup B, C and W lineage. The Journal of Infection 2015.
99. Mustapha MM, Marsh JW, Krauland MG, et al. Genomic Epidemiology of Hypervirulent Serogroup W, ST-11 *Neisseria meningitidis.* EBioMedicine 2015;2:1447-55.
100. Kong Y, Ma JH, Warren K, et al. Homologous recombination drives both sequence diversity and gene content variation in *Neisseria meningitidis*. Genome Biology and Evolution 2013;5:1611-27.
101. Castineiras TM, Barroso DE, Marsh JW, et al. Capsular switching in invasive *Neisseria* meningitidis, Brazil(1). Emerg Infect Dis 2012;18:1336-8.
102. Wang Q, Shao Z, Wang X, et al. Genetic study of capsular switching between *Neisseria meningitidis* sequence type 7 serogroup A and C strains. Infection and Immunity 2010;78:3883-8.
103. Tzeng YL, Datta AK, Strole CA, Lobritz MA, Carlson RW, Stephens DS. Translocation and surface expression of lipidated serogroup B capsular Polysaccharide in *Neisseria meningitidis*. Infection and Immunity 2005;73:1491-505.

104. Spinosa MR, Progida C, Tala A, Cogli L, Alifano P, Bucci C. The *Neisseria meningitidis* capsule is important for intracellular survival in human cells. Infection and Immunity 2007;75:3594-603.
105. Gupta S, Maiden MC, Feavers IM, Nee S, May RM, Anderson RM. The maintenance of strain structure in populations of recombining infectious agents. Nature medicine 1996;2:437-42.
106. Anonychuk A, Woo G, Vyse A, Demarteau N, Tricco AC. The cost and public health burden of invasive meningococcal disease outbreaks: a systematic review. Pharmacoeconomics 2013;31:563-76.
107. Harrison LH. Vaccine prevention of meningococcal disease: making slow progress. Clin Infect Dis 2006;43:1395-7.
108. Collard JM, Issaka B, Zaneidou M, et al. Epidemiological changes in meningococcal meningitis in Niger from 2008 to 2011 and the impact of vaccination. BMC Infectious Diseases 2013;13:576.
109. Abad R, Lopez EL, Debbag R, Vazquez JA. Serogroup W meningococcal disease: global spread and current affect on the Southern Cone in Latin America. Epidemiol Infect 2014;142:2461-70.
110. Campbell H, Saliba V, Borrow R, Ramsay M, Ladhani SN. Targeted vaccination of teenagers following continued rapid endemic expansion of a single meningococcal group W clone (sequence type 11 clonal complex), United Kingdom 2015. Euro Surveill 2015;20.
111. Cohn AC, Harrison LH. Meningococcal vaccines: current issues and future strategies. Drugs 2013;73:1147-55.
112. Maiden MC, Ibarz-Pavon AB, Urwin R, et al. Impact of meningococcal serogroup C conjugate vaccines on carriage and herd immunity. J Infect Dis 2008;197:737-43.
113. Tsang RS, Hoang L, Tyrrell G, et al. Genetic and antigenic characterization of Canadian invasive *Neisseria meningitidis* serogroup C (MenC) case isolates in the post-MenC conjugate vaccine era, 2009-2013. J Med Microbiol 2015;64:174-9.
114. Foxman B, Riley L. Molecular epidemiology: focus on infection. Am J Epidemiol 2001;153:1135-41.
115. Achtman M. A Phylogenetic Perspective on Molecular Epidemiology. In: Sussman M, ed. Molecular Medical Microbiology. London, UK: Academic Press; 2002:485-509.
116. Tyler S, Tsang R. Genetic analysis of Canadian isolates of C:2a:P1.2,5 and B:2a:P1.2,5 *Neisseria meningitidis* strains belonging to the hypervirulent clone of ET-15. Canadian Journal of Microbiology 2004;50:433-43.
117. Bratcher HB, Corton C, Jolley KA, Parkhill J, Maiden MC. A gene-by-gene population genomics platform: de novo assembly, annotation and genealogical analysis of 108 representative *Neisseria meningitidis* genomes. BMC Genomics 2014;15:1138.
118. Riley LW. Principles and Approaches. In: Riley LW, ed. Molecular Epidemiology of Infectious Diseases. Washington, DC: ASM Press; 2004:1-28.
119. Hill DM, Lucidarme J, Gray SJ, et al. Genomic epidemiology of age-associated meningococcal lineages in national surveillance: an observational cohort study. The Lancet Infectious Diseases 2015.
120. Harrison OB, Bray JE, Maiden MC, Caugant DA. Genomic Analysis of the Evolution and Global Spread of Hyper-invasive Meningococcal Lineage 5. EBioMedicine 2015;2:234-43.
121. Toros B, Hedberg ST, Unemo M, et al. Genome-Based Characterization of Emergent Invasive *Neisseria meningitidis S*erogroup Y Isolates in Sweden from 1995 to 2012. Journal of clinical microbiology 2015;53:2154-62.
122. Gupta S, Maiden MC. Exploring the evolution of diversity in pathogen populations. Trends in microbiology 2001;9:181-5.
123. Cohan FM. What are bacterial species? Annu Rev Microbiol 2002;56:457-87.
124. Bruce MG, Rosenstein NE, Capparella JM, Shutt KA, Perkins BA, Collins M. Risk factors for meningococcal disease in college students. JAMA : the journal of the American Medical Association 2001;286:688-93.
125. Stuart JM, Cartwright KA, Dawson JA, Rickard J, Noah ND. Risk factors for meningococcal disease: a case control study in south west England. Community Med 1988;10:139-46.
126. Caugant DA, Maiden MC. Meningococcal carriage and disease--population biology and evolution. Vaccine 2009;27 Suppl 2:B64-70.

127. Trotter CL, Gay NJ, Edmunds WJ. The natural history of meningococcal carriage and disease. Epidemiol Infect 2006;134:556-66.
128. Feavers IM, Gray SJ, Urwin R, et al. Multilocus sequence typing and antigen gene sequencing in the investigation of a meningococcal disease outbreak. Journal of clinical microbiology 1999;37:3883-7.