Genome-based characterization of *Neisseria meningitidis* with focus on the emergent serogroup Y disease
Cogito, ergo sum
- Descartes
Genome-based characterization of Neisseria meningitidis with focus on the emergent serogroup Y disease
Abstract


*Neisseria meningitidis*, also referred to as meningococcus, is one of the leading causes of epidemic meningitis and septicaemia worldwide. Despite modern treatment, meningococcal disease remains associated with a high mortality (about 10%). Meningococcal disease is mainly restricted to specific hypervirulent lineages and specific capsular groups (serogroups), which have a changing global distribution over time. At the end of the 2000s, the previously unusual serogroup Y emerged, corresponding to half of all of the invasive meningococcal disease (IMD) cases in Sweden by the beginning of the 2010s. The aim of this thesis is to describe the emergence of serogroup Y meningococci genetically in an effort to understand some of the factors involved in the successful spread of this group throughout Sweden. In addition, genetic typing schemes were evaluated for surveillance and outbreak investigation.

Our results indicate that the currently recommended typing for surveillance of meningococci could be altered to include the factor H-binding protein (fHbp). A highly variable multilocus variable number tandem repeat analysis (HV-MLVA) was able to confirm connected cases in a suspected small outbreak. In addition, a strain type sharing the same *porA*, *fetA*, *porB*, *fHbp*, *penA* and multilocus sequence type was found to be the principal cause of the increase in serogroup Y disease. However, a deeper resolution obtained from the core genomes revealed a subtype of this strain, which was mainly responsible for the increase. Finally, when the Swedish serogroup Y genomes were compared internationally, different strains seemed to dominate in different regions. This indicates that the increase was probably not due to one or more point introductions of a strain previously known internationally but more probably multifactorial.

*Keywords*: *Neisseria meningitidis*, meningococcal disease, serogroup Y, molecular characterization, epidemiology, genome sequencing.

Bianca Törös, School of Health and Medical sciences, Örebro University, SE-701 82 Örebro, Sweden, nora-bianka.torosvig@orebroll.se
**Sammanfattning**

*Neisseria meningitidis* (meningokocken) är en bakterie som bärs asymptomatiskt i övre luftvägarna av ungefär 10 % av befolkningen. I vissa fall undkommer dock bakterien människans immunförsvar och når blodomloppet där den orsakar blodförgiftning och/eller korsar blodhjärrbarriären och orsakar epidemisk hjärnhinneinflammation. Dödligheten är ungefär 10 % och kan till följd av det snabba sjukdomsförloppet inträffa inom ett dygn från första symptomet. Meningokocksjukdom behandlas med antibiotika och kan också i viss utsträckning förebyggas med vaccin.


I denna avhandling undersöks olika genetiska metoder för att övervaka cirkulerande meningokocker och för att snabbt kunna identifiera kopplade fall i ett sjukdomsutbrott. Resultaten indikerar att genen som kodar för ett yttermembranprotein som används i vaccin mot serogrupp B meningokocker: faktor H-bindande protein (FHbp), skulle kunna användas för meningokockövervakning. Vid utbrott skulle en genetisk metod som särskiljer olika typer av meningokocker baserat på antalet repetitiva sekvenser i vissa högvariabla regioner av genomet, kunna användas.

List of publications


Reprints have been made with the permission of the publisher.
## List of abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>Adjusted Rand coefficient</td>
</tr>
<tr>
<td>AW</td>
<td>Adjusted Wallace coefficient</td>
</tr>
<tr>
<td>BIGSdb</td>
<td>Bacterial Isolate Genome Sequence Database</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BURST</td>
<td>Based Upon Related Sequence Types</td>
</tr>
<tr>
<td>CC</td>
<td>clonal complex</td>
</tr>
<tr>
<td>CSF</td>
<td>cerebrospinal fluid</td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>D-index</td>
<td>diversity index</td>
</tr>
<tr>
<td>DUS</td>
<td>DNA uptake sequence</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxynucleoside triphosphate</td>
</tr>
<tr>
<td>ddNTP</td>
<td>dideoxynucleoside triphosphate</td>
</tr>
<tr>
<td>FetA</td>
<td>ferric enterobactin transport protein A</td>
</tr>
<tr>
<td>fHbp</td>
<td>factor H-binding protein</td>
</tr>
<tr>
<td>HV-MLVA</td>
<td>highly-variable multilocus variable-number tandem repeat analysis</td>
</tr>
<tr>
<td>IMD</td>
<td>invasive meningococcal disease</td>
</tr>
<tr>
<td>LOS</td>
<td>lipooligosaccharides</td>
</tr>
<tr>
<td>LPS</td>
<td>lipopolysaccharides</td>
</tr>
<tr>
<td>MLST</td>
<td>multilocus sequence typing</td>
</tr>
<tr>
<td>MLVA</td>
<td>multilocus variable-number tandem repeat analysis</td>
</tr>
<tr>
<td>MST</td>
<td>minimum-spanning tree</td>
</tr>
<tr>
<td>NadA</td>
<td><em>Neisseria</em> adhesin A</td>
</tr>
<tr>
<td>OMP</td>
<td>outer membrane protein</td>
</tr>
<tr>
<td>Opa</td>
<td>opacity</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PFGE</td>
<td>pulsed-field gel electrophoresis</td>
</tr>
<tr>
<td>PorA</td>
<td>Porin A</td>
</tr>
<tr>
<td>PorB</td>
<td>Porin B</td>
</tr>
<tr>
<td>PubMLST</td>
<td>public multilocus sequence typing</td>
</tr>
<tr>
<td>rep-PCR</td>
<td>repetitive sequence-based PCR</td>
</tr>
<tr>
<td>rMLST</td>
<td>ribosomal multilocus sequence typing</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>serogroup</td>
<td>serological group</td>
</tr>
<tr>
<td>SLV</td>
<td>single locus variant</td>
</tr>
<tr>
<td>ST</td>
<td>sequence type</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>UPGMA</td>
<td>unweighted pair group method with arithmetic average</td>
</tr>
<tr>
<td>VNTR</td>
<td>variable-number tandem repeat</td>
</tr>
<tr>
<td>VR</td>
<td>variable region</td>
</tr>
</tbody>
</table>
# Table of Contents

INTRODUCTION ................................................................................... 13  
Meningococcal disease ............................................................................. 13  
   Clinical presentations ........................................................................... 13  
   Risk factors .......................................................................................... 14  
   Treatment and prevention .................................................................... 15  
Epidemiology ........................................................................................... 16  
   History ................................................................................................. 19  
   Serogroup Y ......................................................................................... 20  
Biology of *Neisseria meningitidis* ............................................................. 22  
   Genome dynamics ................................................................................ 22  
   Host-microbe interaction ..................................................................... 24  
   *Adhesion and invasion* ..................................................................... 25  
   *Evasion from host immunity* ............................................................. 26  
Identification and characterization of meningococci ................................ 26  
   Basic characterization .......................................................................... 28  
   Multilocus sequence typing (MLST)..................................................... 29  
   *Ribosomal multilocus sequence typing (rMLST)*............................... 30  
   Pulsed-field gel electrophoresis (PFGE) ................................................ 30  
   Repetitive sequence-based PCR (rep-PCR) ......................................... 30  
   Multilocus variable-number tandem repeat analysis (MLVA) .............. 31  
   High-throughput sequencing ................................................................ 32  
   *Sample preparation* ......................................................................... 32  
   *Amplification* .................................................................................... 33  
   *Sequencing* ....................................................................................... 34  
   *Genome assembly* ............................................................................ 35  
   Inference networks ............................................................................... 35  
   *Dendrograms* ................................................................................... 36  
   *Minimum spanning tree* ................................................................... 36  
   *Neighbour-Net networks* ................................................................ 37  
AIMS ........................................................................................................ 38  
MATERIALS AND METHODS .............................................................. 39  
Bacterial isolates ....................................................................................... 39  
   Isolation of DNA (paper I-IV) ............................................................. 40  
   Repetitive sequence-based PCR (paper I) ......................................... 41  
   Pulsed-field gel electrophoresis (paper I) ........................................... 41
Highly variable multilocus variable-number tandem repeat analysis (paper I and III) .......................................................................................... 41
Sequence-based typing (paper I-III) .......................................................... 42
  Real-time PCR ...................................................................................... 42
  DNA sequencing .................................................................................. 45
  Sequencing analysis .............................................................................. 46
Whole-genome sequencing (paper IV) ...................................................... 46
  Sample preparation .............................................................................. 46
  Library preparation and sequencing ..................................................... 46
  Data analysis ........................................................................................ 47
Inference networks (paper I-IV) ................................................................ 47
Statistical analyses .................................................................................... 48
  Discriminatory power .......................................................................... 48
  Congruence between methods ............................................................ 49
RESULTS AND DISCUSSION ................................................................... 50
Molecular typing schemes (Paper I-III) ..................................................... 50
  Surveillance (Paper II and III) ............................................................... 50
  Outbreak investigations (paper I and III) .............................................. 51
  Limitations ........................................................................................... 56
Genetic characteristics of invasive N. meningitidis isolates (paper II-IV) .. 57
  MLST, penA and antigen genes ............................................................ 57
  Serogroup Y ......................................................................................... 57
  Serogroups A, B, C, E, W and Y .......................................................... 61
Whole-genome comparisons of serogroup Y isolates (paper IV)......... 63
  Sweden ................................................................................................. 63
  Global population structure ................................................................ 64
  Limitations ........................................................................................... 64
CONCLUSIONS AND FUTURE PERSPECTIVES .................................. 66
ACKNOWLEDGEMENTS ...................................................................... 67
REFERENCES ......................................................................................... 70
Introduction

Meningococcal disease

Invasive meningococcal disease (IMD) is caused by the bacterium *Neisseria meningitidis*, often referred to as meningococcus. The bacterium exclusively infects humans, and its natural reservoir is the upper respiratory tract (1-3). Transmission of meningococci between hosts occurs by person-to-person direct contact or through upper respiratory oral secretions (2, 3). IMD usually occurs within 1-14 days after exposure. During endemic situations, approximately 10% (4-7) of all healthy individuals will, at a given point, carry this bacterium asymptomatically (8), which acts as an immunizing event (9). The carriage rates increase sharply in teenagers and peak in young adults. In addition, the carriage prevalence is increased in household contacts of IMD cases (10-14) and closed or semi-closed populations such as military recruits and universities (7, 15-17). Meningococci can cause epidemic meningitis and severe sepsis, usually with a rapid and fatal outcome (2). During the first decades of the 20th century, before the introduction of antiserums or antibiotic treatment, the mortality rates were as high as 75-80% (18, 19). In 1919, Herrick described meningococcal disease as “no other infection so quickly slays” (20). Despite the high sensitivity to antibiotics used for treatment, the mortality rate remains approximately 10% (21-27).

Clinical presentations

The clinical response to IMD may range from a relatively benign form to death in only a few hours (28). Diagnosing IMD is difficult because clinical presentations may resemble the symptoms of other less serious diseases. The most common clinical presentation of IMD is meningitis without shock, developed by approximately 60% of patients in industrialized countries (29, 30). Meningitis is characterized by a sudden onset of headache, fever, stiffness of the neck, sometimes also in combination with nausea, vomiting, photophobia and altered mental status (2). The mortality rate is relatively low (approximately 5%). However, long-term complications are present in about 10-20% of all survivors of meningitis, which may include brain damage, hearing loss or some form of learning disability (30-34).

The more severe and second most common form of IMD is septicaemia: about 10% develop sepsis alone and 40% have meningitis and sepsis...
combined (35). Symptoms of meningococcal sepsis include an abrupt onset of fever, flu-like symptoms and a petechial or purpuric rash that may progress to purpura fulminans and hypotension, acute adrenal haemorrhage (Figure 1) and multiorgan failure (2, 36, 37). The reported mortality rate in patients with fulminant septicemia may vary from 20 to 80% because of the diversity in the natural course of the disease, the quality of medical treatment and different disease definitions (1). Amputation or plastic surgery is performed in 10-20% of patients because of skin and limb necrosis (1, 37).

Rare forms of invasive disease include chronic meningococcemia (38-40) and septic arthritis (41, 42). In addition, local infections due to meningococci can cause sinusitis, otitis (43), conjunctivitis (44, 45) and lower respiratory tract infections, including pneumonia (43, 46).

**Figure 1.** Left, fulminant meningococcal septicaemia with ecchymoses, i.e. subcutaneous purpura larger than 1 cm. Right, thrombosis and gangrene of the fingers of a child surviving fulminant meningococcal septicaemia. Reprinted from The Lancet (29) with permission from Elsevier.

**Risk factors**

Among the different factors involved in a higher risk of acquiring IMD, lack of protective bactericidal antibodies is considered the most important (47, 48). Defects in the complement systems and other pathways of the immune system caused by genetic factors (49), as well as anatomical asplenia (50), also infer predisposition to IMD. Finally, active or passive smoking (51) and concurrent viral infection of the upper respiratory tract (52, 53) may increase the formation and spread of respiratory droplets or damage the mechanical integrity of the respiratory mucosa, which is a barrier to invasion.
**Treatment and prevention**

To halt the proliferation of *N. meningitidis* antibiotic treatment should be administered as soon as possible in patients with suspected IMD (54, 55). Many antimicrobial agents are active against this bacterium. Until the causative agent is determined, the recommended treatment of patients with suspected bacterial meningitis in industrialized countries is initially an extended-spectrum cephalosporin combined with vancomycin or ampicillin (56, 57). When *N. meningitidis* is identified, treatment with penicillin alone should be considered or alternatively with a third-generation cephalosporin. Seven days of antibiotic treatment is recommended (58).

Because the primary cause of death in industrialized countries is circulatory collapse, aggressive fluid treatment to increase circulating blood volume may reduce the fatality rates (1, 36, 58-62). Different fluids may be administered intravenously but the volume infused seems more important than the type of fluid (63).

Because the risk of IMD is over 100-fold higher in household contacts than in the normal population (10, 64, 65), antimicrobial prophylaxis is recommended for household members, or anyone exposed to an infected patient’s oral secretions (2, 27). The prophylactic treatment should be administered as soon as possible and at least within 14 days after the onset of disease in the index case to have an optimal effect (66). Because of the cost and risk of drug resistance, a broader prophylaxis is only recommended to control localized outbreaks (e.g., in daycare centres) (29). Sulphonamide drugs were once highly effective but because of many meningococcal strains having developed resistance (associated with a mutation in the *folP* gene) (67, 68), these can no longer be used. Instead, rifampicin, ciprofloxacin and ceftriaxone are used to eradicate *N. meningitidis* from the upper respiratory tract (2).

Although the increased antibiotic resistance in many bacterial pathogens is a serious health threat and the closely related gonococci have reached the status of a super-bug, the meningococcal population remains generally susceptible (69). However, intermediate susceptibility to penicillins has been increasing in many countries (70-73).

To control epidemics and prevent IMD in a longer perspective polysaccharide, conjugate or protein-based vaccines have been developed. Polysaccharide vaccines may provide immunity for up to 3 years; however, the conjugate vaccines introduced in 1999 offer longer protective effects (74). The poor immunogenicity of the serogroup B polysaccharide delayed the development of a broad protective serogroup B vaccine until 2013, when
the first protein-based vaccine covering this serogroup was approved in Europe (75). This vaccine, 4CMenB (Bexsero, Novartis, MA, USA), includes the factor H-binding protein (fHbp), Neisseria adhesin (NadA) allele 3, Neisseria heparin-binding antigen (NHBA) and an outer membrane vesicle (OMV) of a New Zealand serogroup B outbreak strain (75-77). In Sweden, vaccines against IMD are not included in the general vaccination program because of the relatively low IMD incidence.

**Epidemiology**

The incidence of IMD varies with season, in endemic or epidemic situations and is cyclical with peaks and troughs (78, 79). The disease incidence in endemic situations is approximately 1-3/100,000 population (80), but may be as high as almost 1,000/100,000 during severe epidemics in the countries of sub-Saharan Africa, in the “meningitis belt” (81, 82). Outbreaks in the meningitis belt usually occur during the dry season, whereas in Europe and Northern America or other similar temperate regions the incidence is highest during winter season (3, 83-85). Generally, the incidence of IMD has decreased in recent years, from 1.9 per 100,000 population in 1999 to 0.7 per 100 000 in 2010 in Europe (86) and similarly in the US (27, 87). This reduction in the incidence of IMD is partly due to the introduction of routine vaccination with conjugated serogroup C vaccine in some countries (88). Additionally, administration of serogroup A vaccines in some of the countries in the meningitis belt has almost eradicated IMD in these regions (89, 90). The incidence of IMD in Sweden from 1970 to 2013 is shown in Figure 2. IMD incidence is highest among small children and second highest among young adults, whereas the case fatality ratio is highest among the elderly (Figure 3).
Figure 2. The number of cases and incidence of invasive meningococcal disease per 100,000 population in Sweden from 1970 to 2013.

Figure 3. Age-specific incidence per 100,000 population and the corresponding case fatality rates of meningococcal disease in Europe, 2006. The data are from 27 European countries in the European Invasive Bacterial Infections Surveillance Network. Reprinted and modified from Vaccine (29, 78) with permission from Elsevier.
The epidemiology of IMD changes constantly. During circulation in populations, new virulent clonal groups evolve and spread against a background of developing host immunity (91-95). Meningococci can be either encapsulated or unencapsulated and the primary classification system for meningococci is based on the surrounding polysaccharide capsule (see section “Biology of Neisseria meningitidis”). There are 13 serogroups, but other serogroups than A, B, C, W, X and Y rarely cause disease (29). The global distribution of different serogroups in the last decade is shown in Figure 4. The serogroup distribution is in constant change and does not reflect the disease pattern of the past; nor does it predict the future pattern (96). The age distribution of different serogroups is also different: IMD caused by serogroup B is more common among infants and serogroup Y among elderly (Figure 5). This fluctuating epidemiology must be considered in planning disease control, such as vaccination decisions, which has generated different continent-wide surveillance systems (e.g. the European Centre for Disease Prevention and Control, ECDC).

Figure 4. The global distribution of major meningococcal serogroups. The sub-Saharan meningitis belt in Africa is marked in black. Reprinted and modified from Vaccine (29, 78) with permission from Elsevier.
History

The report by Vieusseux in Geneva in 1805 (97) is thought to be the first definitive identification of IMD. However, descriptions similar to IMD have been found dating back to the 16th century (2). The first time the causative agent was linked to epidemic cerebrospinal meningitis was when Weichselbaum cultured the bacteria from a patient in 1887, designating it *Diplococcus intracellularis meningitidis* (98).

In the early 20th century serogrouping of isolates was not developed yet, limiting the amount of epidemiological information. The world wars and inter-war period were associated with large epidemics. The most extended outbreaks in Europe occurred during the Second World War, believed to be caused by serogroup A (2, 79, 96, 99, 100). Meanwhile, Sweden, which was relatively isolated from the Second World War, did not report any epidemics during this period (101). After the end of the world wars, the outbreaks declined and the incidence of IMD returned to endemic levels throughout Europe (102-106).

In the 1960s routine surveillance systems for IMD were operating in most of Europe and serogrouping was performed routinely from the 1970s and onwards (96). Somewhere in the 1970s, serogroups B and C increased. Until the 1990s, virtually all epidemics were caused by serogroups A, B and C (96).

During the end of the 20th century, new strains of serogroups B and C raised the IMD incidence in the Americas (78) and many European coun-
tries (107-111). The age distribution started to change and more cases were reported in the young adult group, probably because of the lack of pre-existing immunity (103, 109, 110, 112). During the same period, serogroups A and C predominated in Asia and Africa (81), except for local outbreaks caused by serogroup X in parts of the meningitis belt (113). In addition, in the mid-1990s serogroup Y caused increased rates of IMD in the USA (114).

In the beginning of the 21st century generally the same pattern has been seen; however, serogroup W emerged as a cause of outbreaks in association with the Hajj pilgrimage which caused large epidemics in the meningitis belt (115, 116). Further, serogroup Y started to increase in Europe.

**Serogroup Y**

Until recently, serogroup Y has accounted for approximately 2% or less of reported IMD cases in Europe (117). The highest increase of IMD caused by serogroup Y was first noted in Sweden (Figure 6), as well as in some other Scandinavian countries in the end of the 2000s, and in the rest of Europe in the beginning of the 2010s (Figure 7) (118, 119).

![Figure 6. The incidence of meningococcal disease caused by serogroups B, C, Y, W and all other serogroups per 100,000 population in Sweden, 1995-2013.](image)
Serogroup Y has generally been associated with a rather benign clinical outcome; Meningococcal pneumonia is considered to be caused mainly by serogroup Y meningococci (27, 120, 121). A review of the literature from 1974 to 1998 (122) showed that, out of 58 patients with meningococcal pneumonia, serogroup Y meningococci were the most commonly recovered and accounted for 44% of identified isolates. IMD caused by serogroup Y is generally believed to be more prevalent in young adults with complement component deficiencies (49, 106, 123-126). Yet, one study showed no differences between isolates from complement-deficient and complement-sufficient patients (127). The literature on IMD caused by serogroup Y is also inconsistent regarding the prevalence of pulmonary infections. In some reports presentation of pneumonia was more common than meningitis or septicemia (128, 129), and in some reports not so common (114, 128, 130-132). The conflicting findings may be due to inconsistencies in the diagnosis. In the case series of serogroup Y disease in a group of US Air Force recruits in 1971-1974 by Koppes et al. (128) meningococcal pneumonia was documented by transtracheal aspirates in 94% of the cases. However, only 4 (6%) of the 68 patients with pneumonia had positive blood cultures.

In a 1997-1998 study of meningococcal carriage among university students in Nottingham, UK serogroup B (24%) was the most prevalent and serogroup Y corresponded to 8% (72/904) as the fourth most prevalent
serogroup of all carriage isolates (133). Similarly, serogroup Y stood for 9% of all carriage isolates in 2010 in teenagers in a region in Germany (134) and was the third most prevalent serogroup among carriage isolates in the Czech Republic, Greece and Norway in 1991-2000 (135). However, serogroup Y carriage has recently increased in the UK. In Nottingham, UK in 2008-2009 serogroup Y represented about half of all carriage isolates (136, 137), which was significantly higher than the rates in 1999-2001 where serogroup Y carriage was 5-6%.

**Biology of *Neisseria meningitidis***

The bacterium *N. meningitidis*, a member of the β-proteobacteria class and the Neisseriaceae family, is a Gram-negative aerobic diplococcus. The Neisseria genus includes the two human pathogens *N. meningitidis* and *N. gonorrhoeae*, and the usually non-pathogenic commensal species such as *N. lactamica*, *N. sicca*, *N. sublava*, *N. mucosa*, *N. flavescens*, *N. cinerea*, *N. polysaccharea* and *N. elongata* (138). The best recognized among the commensals, *N. lactamica*, is very similar to *N. meningitidis* as they both share antigenic structures and colonize the respiratory tract. Carriage of *N. lactamica* in infants and children has even been associated with the development of a cross-protective immunity against *N. meningitidis* (139).

**Genome dynamics**

The three first meningococcal genomes sequenced were those of serogroups A, B and C, namely the strains Z2491, MC58 and FAM18 (140-142). The three genomes, as well as the *N. gonorrhoeae* genome, differ from each other by 9-10% (29). The meningococcal genome consists of a single circular chromosome with approximately 2.2 million base pairs (bp) that encode at least 1,337 genes (143).

The meningococcus has many mechanisms to rapidly vary and diversify its genome in order to avoid the host immune defences and adapt to new environments and selective pressures (Figure 8). In addition to this genetic variation, genome maintenance is necessary to balance the genome dynamics. Thus, the genetic variation is balanced by many DNA repair pathways (144). The plasticity of the meningococcal genome is largely due to its natural competence for transformation throughout its entire life cycle (145-147).

Transformation is one of the three major genetic mechanisms for genetic exchange among bacteria, along with conjugation and transduction. Using transformation, the bacterium can take up and incorporate extracel-
lular DNA from different bacteria by homologous recombination (a type of genetic recombination in which two similar DNA strands exchange genetic material). DNA uptake sequences (DUS), which are repetitive elements of 10 bp involved in the recognition and uptake of DNA (132), are crucial for efficient neisserial transformation. The meningococcal genome has a high abundance of repetitive sequences, where approximately 20% of the chromosome is included in repeats (144, 148), and DUS are the most common of the different types of repetitive sequences (approximately 1900 copies in the genome) (149).

**Figure 8.** Examples of different strategies used by meningococci to obtain genetic variation, DNA repair and selective pressures, which define the genetic diversity and fitness of the population. Reprinted and modified from Trends in Microbiology (150) with permission from Elsevier.

Other mechanisms related to the virulence of meningococci include rapid doubling time, phase and antigenic variation, release of outer membrane vesicles (blebs), molecular mimicry and the possible release of toxins (29). Phase and antigenic variation is used to allow immune escape by variation in expression or structure of the components of the outer membrane: pili, lipooligosaccharides (LOS), outer membrane proteins (OMP)
and the surrounding polysaccharide capsule (Figure 9). These components are all major contributors to the virulence of \textit{N. meningitidis}. The genetic switches used in phase and antigenic variation are due to transformation of homologous DNA, slipped-strand mispairing of repetitive nucleotides, regulation of promoter regions, intergenic recombination events and insertion sequence (IS) element movement. Blebs, which contain OMP and LOS, are thought to rapidly initiate the inflammatory cascades of sepsis and meningitis, as well as facilitate transformation through autolysis that results in the release of DNA. Molecular mimicry can be achieved through expression of host antigens that down-regulate the human immune response (151). One example is the $(\alpha 2\rightarrow 8)$-linked polysialic acid serogroup B capsule, which is identical to structures on the human neural cell adhesion molecule, N-CAM (152).

\textbf{Figure 9.} Major virulence factors in \textit{N. meningitidis}. The bacterium outer membrane contains the outer membrane proteins Opa and Opc, porins, lipooligosaccharides (LOS) and pili (encoded by pilE and pilC, which encodes the tip adhesin PilC). Most of the invasive \textit{N. meningitidis} are surrounded by a polysaccharide capsule.

\textbf{Host-microbe interaction}

The main stages of the pathogenesis of \textit{N. meningitidis} are outlined in Figure 10.
Adhesion and invasion
Among the putative adhesins identified so far, pili, and opacity-associated proteins Opa and Opc, are expressed in the greatest abundance in meningococci (153).

Pili are complex OMP organelles of several thousand micrometers stretching out from the bacterial cell surface. Meningococcal pili are Type IV (Tfp), the same as for numerous other pathogenic bacterial species (154, 155). Their function is to allow for the meningococci (i) to adhere to the epithelial cell surface and thereby colonize the host (156-158), (ii) to interact with each other (aggregation) to form microcolonies (159), (iii) to take up DNA during transformation (160, 161) and (iv) to move through the mucus layer and over epithelial surfaces by using twitching motility (154, 162, 163).

Figure 10. Stages in the pathogenesis of N. meningitidis. Meningococci are transmitted by aerosol or secretions and establish intimate contact with the nonciliated epithelial cells of the nasopharynx, where they multiply, i.e. colonize. In a small number of cases the bacteria cross over the epithelium to the bloodstream, causing systemic disease. Tumour necrosis factor (TNF) from phagocytes and lipopolysaccharide (LPS) cause toxic damage to ciliated epithelial cells of mucosal surfaces. In addition, the meningococci may reach the brain and pass through the blood-brain barrier and infect the meninges and cerebrospinal fluid (CSF). ECM, extracellular matrix. Reprinted from Nat Rev Microbiol (153) with permission from Macmillan Publishers Ltd.

After the initial attachment using pili, a more intimate adherence and internalization is driven by Opa and Opc (164, 165), as well as the major porin B (PorB) (166). Several other new adhesins have been identified, one
of them, NadA, mediates adhesion to and entry into epithelial cells and is expressed in several hyper-virulent lineages (167). NadA is therefore a vaccine component of the new vaccine covering serogroup B (168, 169). During the intimate adhesion, bacterial surface structures that hinder intimate adhesion, such as the pili and capsule, are downregulated (157, 170-172). It is not completely clear how meningococci gain access to the circulation but possible mechanisms are transcytosis, damaging the monolayer integrity, or by phagocytes carrying them like “Trojan horses” over to the blood stream (Figure 10) (153, 173). LOS is a potent endotoxin that has an important role in septic shock. Septic shock is induced by LOS through a cascade of inflammatory responses leading to disseminated intravascular coagulopathy and circulatory collapse (37). The concentration of LOS in the blood is believed to have a direct correlation to the severity of the disease (174-176).

Survival in the blood stream is dependent on iron acquisition from the host and the capsule, preventing the bacteria from complement-mediated bacteriolysis and phagocytosis. The iron in the human host is bound to iron-binding proteins and the meningococcus has therefore developed complex iron acquisition systems. Iron-acquiring proteins include haemoglobin-binding proteins (HmbR), transferrin-binding proteins A and B (TbpA/B), lactoferrin-binding proteins (LbpA/B) and haptoglobin-haemoglobin-binding proteins (HpuA/B) (177-179).

Evasion from host immunity
Many components are involved in protecting the meningococci from phagocytosis by the host immune system. The surrounding capsule shields the meningococcus from opsonization, phagocytosis, antibody and complement deposition, as well as aiding survival of the bacteria in the blood (153). The capsules of serogroups B, C, W and Y contain sialic acid, which is important for immune evasion (180). fHbp binds to complement factor H to prevent complement-mediated killing (181). In addition, co-signalling of human endothelial cells by the pili and LOS leads to bacterial uptake by non-phagocytic cells (182). PorA helps evade the protective immune host response through phase variation by switching the expression on and off or by more graduated variations (183).

Identification and characterization of meningococci
*N. meningitidis* identification and typing are essential for different reasons (184, 185):
• Identification: to understand the clinical problems and possible complications.
• Antibiotic susceptibility testing: to assign the correct treatment of cases.
• Rapid identification of clusters/outbreaks: for effective intervention (chemoprophylaxis or vaccination) of both cases and close contacts.
• Epidemiological surveillance: for local, national and international disease control.
• To assess vaccine coverage.

The gold standard of diagnosis of IMD is using isolation of meningococci by culture from sterile body fluids such as CSF, joint fluid or blood. Subsequently, microscopy after Gram-staining of the colonies and sugar degradation and antibiotic susceptibility testing should be performed. In some cases direct Gram-staining or antigen detection with, for example, agglutination directly on CSF can be used as a rapid and accurate identification of *N. meningitidis* (186-188). Methods that do not require culturing are especially useful when antibiotic treatment has been performed before sample collection, which may result in a false-negative result. Therefore, DNA-based techniques that can be used on non-viable bacteria have been developed (189-191).

The classical immunological methods for typing of OMPs with monoclonal antibodies have been almost completely replaced by genetic methods because of their limitations (such as masking of epitopes on the cell surface and lack of protein expression) (192-194). In addition, the results from DNA methods can be performed on non-culture specimens, are unambiguous and highly portable. Characterization of meningococci is dependent on the question at hand (185), different targets are appropriate for long- or short-term epidemiology (Table 1) and there is probably no one single solution for both.
Table 1. Examples of methods or gene targets used for meningococcal characterization.

<table>
<thead>
<tr>
<th>Aim</th>
<th>Method/target</th>
<th>Function</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient management</strong></td>
<td>16S rRNA</td>
<td>Species identification</td>
<td>(195, 196)</td>
</tr>
<tr>
<td></td>
<td>ctrA</td>
<td></td>
<td>(197, 198)</td>
</tr>
<tr>
<td></td>
<td>crgA</td>
<td></td>
<td>(199)</td>
</tr>
<tr>
<td></td>
<td>sacC/mynB</td>
<td>Serogroup A</td>
<td>(200, 201)</td>
</tr>
<tr>
<td></td>
<td>siaD</td>
<td>B, C, Y, W</td>
<td>(202, 203)</td>
</tr>
<tr>
<td></td>
<td>xcbA</td>
<td>X</td>
<td>(204)</td>
</tr>
<tr>
<td></td>
<td>ctrA</td>
<td>E, X, Z</td>
<td>(205)</td>
</tr>
<tr>
<td></td>
<td>cnl</td>
<td>null locus</td>
<td>(206)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Surveillance</strong></td>
<td>porA</td>
<td>OMP7 Porin A</td>
<td>(209-211)</td>
</tr>
<tr>
<td></td>
<td>fetA</td>
<td>OMP7 FetA</td>
<td>(212)</td>
</tr>
<tr>
<td></td>
<td>MLST²</td>
<td>7 housekeeping genes</td>
<td>(92)</td>
</tr>
<tr>
<td></td>
<td>MLVA³</td>
<td>8 VNTR8 loci</td>
<td>(213-215)</td>
</tr>
<tr>
<td><strong>Outbreak Investigation</strong></td>
<td>porB</td>
<td>OMP7 Porin B</td>
<td>(216)</td>
</tr>
<tr>
<td></td>
<td>HV-MLVA⁴</td>
<td>4 VNTR8 loci</td>
<td>(215), paper I, III</td>
</tr>
<tr>
<td></td>
<td>rep-PCR⁵</td>
<td>Repetitive sequences</td>
<td>(217), paper I</td>
</tr>
<tr>
<td><strong>Antibiotic susceptibility</strong></td>
<td>penA⁶</td>
<td>Penicillin G</td>
<td>(218, 219)</td>
</tr>
<tr>
<td></td>
<td>rpoB</td>
<td>Rifampicin</td>
<td>(220, 221)</td>
</tr>
<tr>
<td></td>
<td>gyrA</td>
<td>Ciprofloxacin</td>
<td>(222)</td>
</tr>
<tr>
<td></td>
<td>catP</td>
<td>Chloramphenicol</td>
<td>(223)</td>
</tr>
<tr>
<td><strong>Vaccine coverage</strong></td>
<td>fhbP</td>
<td>OMP7 fhbP</td>
<td>(224-227)</td>
</tr>
<tr>
<td></td>
<td>nadA</td>
<td>OMP7 NadA</td>
<td></td>
</tr>
</tbody>
</table>

¹Outbreak investigations may include the targets for surveillance
²Multilocus sequence typing
³Multilocus variable-number tandem repeat analysis
⁴Highly variable multilocus variable-number tandem repeat analysis
⁵Repetitive sequence-based PCR
⁶Encoding penicillin-binding protein 2
⁷Outer membrane protein
⁸Variable-number tandem repeat

**Basic characterization**
The current recommended characterization of meningococci includes serogroup designation and DNA sequencing (Figure 11) of the variable regions (VRs) of the genes porA (VR1 and 2) and fetA (coding for Ferric enterobactin transport protein A) in combination with multilocus sequence typing (MLST, see section “Multilocus sequence typing”). The nomenclature should be of the form: serogroup: PorA type: FetA type:
sequence type (ST) (clonal complex, CC), for example, C:P.1-19,15: F5-1: ST-33 (cc32).

For rapid investigations of disease outbreaks, it is suggested that the antigen encoding genes *porA* and *fetA* should be targeted, as well as *porB* when additional resolution is required (228).

**Figure 11.** Principle of the Sanger method for DNA sequencing. 1) The amplified DNA template is elongated using labelled ddNTPs (different dyes for each type of ddNTP) that randomly terminates the elongation when incorporated. 2) The subsequent DNA fragments of different lengths are separated by size using capillary electrophoresis. Each fragment is registered by its laser-excited dye. This signal is converted by a computer into a chromatogram in which the order of the bases is shown.

**Multilocus sequence typing (MLST)**
MLST is a method for characterizing bacteria based on DNA sequencing of 6-8 housekeeping genes that are under stabilizing selection (229). In *N. meningitidis* seven loci (*abcZ, adk, aroE, fumC, gdh, pdhC* and *pgm*) are
used. Each sequence is assigned an allele number based on its sequence. Different allele combinations of the seven genes in the MLST are used to divide the isolates into ST. If four or more of the seven alleles are shared, the isolates are designated to the same CC (230). The MLST CCs have been shown to correspond to different hyper-invasive lineages (92, 135, 231).

Ribosomal multilocus sequence typing (rMLST)  
rMLST is a new extension of MLST because of the sometimes insufficient resolution among very closely related bacteria provided with MLST (232). It may be used as a universal characterization of bacteria from domain to strain. rMLST indexes 53 genes encoding the bacterial ribosomal protein subunits (rps). The rps genes are particularly appropriate targets because they exist in all bacterial genomes, are distributed over the chromosome and are under stabilizing selection (233).

Pulsed-field gel electrophoresis (PFGE)  
PFGE uses “rare site” restriction enzymes that digest the DNA into large fragments that can be used as DNA fingerprints. The fragments are separated on a gel by applying an electric field that regularly changes direction, which is particularly suitable for higher molecular weight DNA molecules (234, 235). PFGE has proven to be a good method because it is highly discriminatory and able to discriminate isolates indistinguishable with other techniques in several bacterial species (236-238), including N. meningitidis (239-241). However, the technique is limited by its difficulties with resolving bands of similar size (242) and issues with interlaboratory reproducibility.

Repetitive sequence-based PCR (rep-PCR)  
A rep-PCR is a microbial typing method that assesses outbreaks in real-time by using multiple non-coding repetitive sequences dispersed throughout the genome as a DNA fingerprint (243, 244). The DiversiLab system (bioMérieux) is an automated rep-PCR that can be performed using a species specific kit with quality-controlled reagents. Subsequently, the amplicons with varying lengths are separated by electrophoresis on a microfluidics chip (Figure 12). The data are automatically collected and reports that include dendrograms, electropherograms, virtual gel images and scatter plots are generated (217).
Figure 12. An overview of the two main steps in a repetitive sequence-based PCR using the DiversiLab system (bioMérieux). The first step is PCR amplification in which different regions between repeat sequences are amplified generating fragments of different length. The second step, fragment detection, is performed by separating the fragments by size using gel electrophoresis creating a fingerprint pattern. Modified with permission from Sabina Davidsson.

Multilocus variable-number tandem repeat analysis (MLVA)

A large part of all genomes consists of repeats with multiple copies. These repeats vary in size, location, complexity and repeat mode. MLVA uses the variability in the number of short tandem repeated sequences to create DNA fingerprints for many different bacterial pathogens (Figure 13) (245), including *N. meningitidis* (213, 214). Rapid alterations, such as slipped-strand mispairing in short DNA tandem repeats situated in coding or promoter regions of genes controlling the expression of meningococcal surface antigens, have been shown to improve survival in *N. meningitidis* (246).

Figure 13. An example of the fingerprints obtained from two strains using a multilocus variable-number tandem repeat (MLVA) analysis.
Different VNTR loci can vary in their stability over time. Some loci are more stable, such as the eight loci described by Shouls et al. (215), which were used in a MLVA scheme that yielded clusterings similar to those of MLST. Other VNTR loci that are more highly variable can be used in a highly variable MLVA (HV-MLVA) more suitable for the analysis of outbreaks where MLST is not discriminatory enough.

The number of tandem repeats in each specific VNTR loci is determined by amplifying each locus and subsequently separating the fragments by electrophoresis. The number of repeats (copy number) is determined by the amplicon size (from the electrophoresis), an offset size (the number of basepairs between the primers and the repeats) and the number of basepairs in each repeat (Equation 1). The offset size and repeat size are determined when designing a specific MLVA protocol, usually by performing in silico tandem repeat searches.

\[
\text{copy number} = \frac{\text{amplicon size} - \text{offset size}}{\text{repeat size}}
\]  

High-throughput sequencing

The genome of Haemophilus influenzae was the first sequenced genome of a free-living organism (247). It was sequenced using the Sanger method in 1995 and it took years of effort requiring six-digit budgets. The machines that have been used for Sanger sequencing have since then maximized their capacity to approximately 1 million DNA bases per day. Since the introduction of high-throughput (or next-generation) sequencing in about 2005, a bacterial genome can be sequenced in a matter of hours and thousands of times cheaper (248). There are currently a wide variety of high-throughput sequencing platforms, each with their own throughput, read length, errors and bias patterns. A brief overview of the high-throughput sequencing workflow is shown in Figure 14.

Sample preparation

Each platform starts with a fragmentation of the genomic DNA, either enzymatic or mechanical, to generate random, overlapping DNA fragments. The fragments are subsequently tagged with adaptors by ligation. Following ligation, the fragments are size selected using band excision from agarose gels or paramagnetic-bead-based technology depending on the platform and application (248). Fragmentation and tagging can be
combined in a “tagmentation” (249) available only for the Illumina platform, which minimizes the sample loss and hands-on time.

**Figure 14.** A schematic of the sample preparations and template amplifications for the main high-throughput sequencing platforms. PGM, Personal Genome Machine. Reprinted from Nat Rev Microbiol (248) with permission from Macmillan Publishers Ltd.

**Amplification**

Platforms can be divided into two groups based on amplification. The earliest one, which is one of the most widely used, depends on the production of clonally amplified templates made from a single DNA molecule in the initial sample. The second one determines the sequence of single mole-
cules without amplification. In preparation for amplification the fragments are attached to a solid surface. For this, flow cells, solid beads or ion sphere particles are used depending on the platform (Figure 14). In solid-phase bridge amplification each library fragment is clonally amplified using a universal primer. The bridge amplification onto the flow cell creates thousands of unique clusters (Figure 14) (250). For bead-based amplification, the beads are enclosed in an emulsion PCR, i.e. aqueous phase microreactors separated from each other in a water-in-oil emulsion.

Paired-end and mate-pair sequencing are available on some platforms to obtain maximum coverage and longer contiguous sequences. Paired-end reads are short fragments and mate-pairs are long-insert paired-end reads that add valuable information about the location of sequences in difficult regions such as highly repetitive sequences (248).

Sequencing
The details and approach in the sequencing chemistry differ between sequencing platforms. The most widely employed is the Illumina platform that uses Solexa chemistry (250). All the clusters created from the bridge amplification are sequenced in cycles simultaneously base-by-base using four fluorescently labelled “reversible terminator” nucleotides. When a complementary nucleotide is bound to the template, the clusters are excited by a laser emitting a different wavelength for each nucleotide. The fluorescent label and blocking group are subsequently removed to allow for the next nucleotide to be incorporated.

The 454 and Ion Torrent platforms avoid the use of terminators and a single type of dNTP is flowed across the template in each cycle. The bases in the Ion Torrent platform are detected by hydrogen ions released during base incorporation (251). The 454 exploits the pyrosequencing approach using pyrophosphate (252, 253).

Single-molecule sequencing, used mainly by Pacific Biosciences (254, 255), is distinct from all other technologies because it does not require amplification before sequencing and hence is free from amplification artefacts. Using continuous imaging in real-time, it detects the incorporation of each labelled nucleotide by a DNA polymerase molecule. The DNA polymerase is tethered within a zero-mode waveguide detector working on very low detection volumes.
Genome assembly
After sequencing, the generated sequence reads need to be assembled into contiguous DNA sequences (contigs) by merging overlapping sequence reads. There are two approaches for assembly: either to assemble the reads with the use of a reference genome as a guide, or de novo assembly without the use of any reference genome (35, 256). The most efficient assemblers for short-read sequences are usually those based on de Bruijn graphs (257), which are directed graphs representing overlaps between sequences. The most common de Bruijn-based assembler is Velvet (Figure 15) (258).

There are many ways to assess the quality of a de novo assembly, the number of scaffolds and contigs required to represent the genome, the proportion of reads that can be assembled and the absolute length of contigs and scaffolds (259). The most common metric is the N50, which is based on assembly size. N50 is a weighted median statistic such that 50% of the entire assembly is contained in contigs or scaffolds equal to or larger than this value.

![Figure 15. The basic concepts of a de novo assembly using Velvet (258). The reads are first converted into k-mers, which are substrings of a sequence with the length k, using a hash table. Subsequently, overlapping k-mers are assembled into contigs via a de Bruijn graph.](image)

Inference networks
Cluster analysis is a way of visualizing hierarchical relatedness between samples by building a dendrogram, tree or a network based on their DNA sequences. First, a distance matrix needs to be constructed based on a specific similarity coefficient contingent on the type of input variable. Distances are commonly calculated for each pair of isolates in which identical isolates have a distance of 0 and those with nothing in common have a
distance of 1. Similarity coefficients are divided into categorical (qualitative) and numerical and binary (quantitative). One of the numerical coefficients, the Pearson correlation coefficient, is a curve-based coefficient based on densitometric values and can be used to compare different fingerprint patterns such as those of a rep-PCR. Binary coefficients are more suitable for band-based methods such as PFGE. Categorical coefficients construct similarity matrices based on the number of, for example, alleles that differ between two profiles. In systems involving recombination a single genetic event may result in a large number of altered sites. Consequently, allele sequences are not used and instead each allele number difference is treated identically. If two profiles differ in a single allele out of four, the cost will be 1/4; if two loci differ, the cost will be 2/4, and so on.

One important deficiency in all dendrograms, trees and networks is that the result from a clustering analysis is usually not unique and may differ depending on, for example, the algorithm used or the order of the input entries. The second important deficiency is that algorithms always assume that the input data is perfect, minor variations due to experimental errors using the same algorithm may result in different clusters.

**Dendrograms**

Dendrograms use similarity matrices as input. The dendrograms for HV-MLVA and rep-PCR data may be constructed using the unweighted pair group method with arithmetic average (UPGMA), which is a frequently used clustering method. The problem with dendrograms constructed with UPGMA is that a slightly different clustering may be seen when the data are presented to the algorithm in a different order. It is thus important not to draw phylogenetic inferences from the clustering pattern seen with this method. Therefore, UPGMA may be used as a quick guide to identify similar isolates.

**Minimum spanning tree**

A minimum spanning tree (MST) is a subgraph of a connected weighted undirected graph that connects all samples with Prim’s algorithm, i.e. so that the summed distances of all the edges (like the branches in phylogenetic trees) of the tree are minimized. These trees are calculated from distance matrices rather than from the data set directly. There may be many possible MSTs for a given dataset, but the priority rules of Based Upon Related Sequence Types (BURST) clustering (260) is used for the MLST and MLVA cluster analysis. The BURST algorithm first links types that
have the highest number of single- or double-locus variants, the highest number of entries and the most frequent states. BURST uses a model of bacterial evolution in which an ancestral (or founding) genotype increases in frequency in the population and diversifies to produce a cluster of closely related genotypes, a CC, that are all descended from the founding genotype. For MLST data based on seven loci, it has been found that a cut-off point of five identical loci maximizes the inclusion of strains belonging to a single CC.

Neighbour-Net networks
Phylogenetic networks can be divided into split networks, phylogenetic trees and reticulate networks. They are an alternative to phylogenetic trees suitable to be employed on reticulate events such as hybridization, horizontal gene transfer and recombination (261). Phylogenetic networks generalize phylogenetic trees because they permit the representation of conflicting signals or alternative phylogenetic histories (262). Neighbour-Net is a type of split network that works as a consensus network, which is a combinatorial generalization of phylogenetic trees and is designed to represent incompatibilities within and between data sets (263). The incompatible signals are represented by parallel edges instead of single branches and the nodes do not represent ancestral species (261). Like MSTs, these networks are constructed from distance matrices.
Aims

The general aim of this thesis is to investigate the genetics underlying the *N. meningitidis* serogroup Y emergence in Sweden in the end of the 2000s.

The specific aims of this thesis are to:

- Find optimal molecular typing schemes for surveillance and outbreak investigations of *N. meningitidis* (paper I and III).

- Describe the clonal pattern of invasive *N. meningitidis* serogroup Y in Sweden during 2000-2012 using 12 fine typing genes (paper II-III).

- Compare the population structure and phylogeography of serogroup Y genomes in Sweden with those in the USA, England, Northern Ireland and Wales (paper IV).
Materials and methods

Bacterial isolates

All cases of IMD in Sweden were mandatorily notified by clinicians and the corresponding isolates sent to our laboratory, the Swedish Reference Laboratory for Pathogenic Neisseria, Department of Laboratory Medicine, Clinical Microbiology, Örebro University Hospital, Örebro, for confirmation of species, antibiotic susceptibility testing using the Etest method (bioMérieux), serogrouping by co-agglutination (264) and PorA typing (265). MLST and FetA typing are also performed for yearly epidemiological surveillance. All Swedish meningococcal isolates had been cultured and stored at -70°C as part of the routine diagnostics. Culturing was performed using chocolate agar at 37°C in a 5% CO₂ enriched atmosphere overnight. The serogroup B strain MC58 (141) was included in all analyses as a reference. Basic epidemiological data (age, sex, area of residence, clinical site of isolation and date of sample collection) were gathered routinely for all isolates.

In paper I serogroup C isolates from 2001 to 2009 (n=36) in Sweden were analyzed. The sites of collection were blood (n=25) and CSF (n=11). Because the aim of paper II was to evaluate molecular typing methods for meningococci suspected to belong to disease outbreaks, 17 of the isolates were spatiotemporally associated in seven clusters, each cluster consisting of isolates within the same serogroup collected in the same central county and within a timeframe of 1 month. The remaining isolates were chosen because of their common serogroup or PorA type to the isolates under investigation and therefore of interest from a wider epidemiological perspective.

In paper II all invasive \textit{N. meningitidis} serogroup Y isolates (except one in 2009 which was overlooked) between 2000 and 2010 (n=85) in Sweden were examined. The sites of collection were blood (n=73), CSF (n=11) and joint fluid (n=1). The 10-year period investigated was chosen because it contained serogroup Y isolates, both pre- and post-emergence.

In paper III the analysis included all invasive \textit{N. meningitidis} isolates in 2010 (n=57), 2011 (n=61) and 2012 (n=90) in Sweden. These isolates comprised serogroups Y (97/208), C (57/208), B (44/208), W (8/208), A (1/208) and E (1/208). The isolates were collected from clinical specimens of blood (n=159), CSF (n=44), joint fluid (n=4) and tissue (n=1). Of the
208 invasive isolates, 35 were spatiotemporally associated in 16 clusters with the same definition of spatiotemporal clusters as in paper II.

In paper IV the same isolates as in paper I were included, but in addition to this, the missed isolate from 2009 was added as well as invasive serogroup Y isolates from 1995 to 2000 and 2011 to 2012. In total, 186 isolates from 1 January 1995 to 31 November 2012 were included in this study, corresponding to more than 98% (186/188) of all serogroup Y isolates identified during this period. For comparison in the data analysis of this paper, 143 serogroup Y isolates from the Meningitis Research Foundation Meningococcus Genome Library (MRF MGL), which includes all available isolates from the IMD surveillance in England, Wales and Northern Ireland in 2010-2011 (n=523) and 2011-2012 (n=477), were used to represent a region with a relatively low serogroup Y disease increase. Additionally, two US strains collected in Maryland in 1999, which represent the early and late strain type that predominated in the USA and led to serogroup Y being the cause of over one third of all IMD in the USA in the middle of the 1990s (266-268), were included.

Isolation of DNA (paper I-IV)

In paper I the genomic DNA of all isolates for the rep-PCR was prepared using 1 µL loop of plated culture and the UltraClean Microbial DNA isolation kit (Mo Bio Laboratories) following the manufacturer’s instructions. The DNA solutions for the rep-PCR were subsequently diluted to a standard concentration of 35 ng/µL before use.

In paper II the DNA extractions for amplification of the fHbp gene were performed by boiling 1-2 colonies in 100 µl 10 mM TRIS-HCl at 98°C for 15 min. The suspensions were subsequently centrifuged for 5 min at approximately 12,000 × g. An automated extraction was used on all other isolates in paper I and all isolates in paper II and III. For this, 20 colonies from each cultured organism were suspended in 2 mL NaCl (0.85%) and 100 µL was subsequently used in the DiaSorin Bullet instrument with the Bullet BUGS’n BEADS kit (DiaSorin) according to the manufacturer’s description.

In paper IV the genomic DNA (5 µg) was isolated from bacterial cultures using the Wizard Genomic DNA purification kit (Promega) according to the manufacturer’s instructions.

All DNA preparations were stored at 4°C.
Repetitive sequence-based PCR (paper I)
rep-PCR was performed using the DiversiLab DNA Fingerprinting kit Neisseria (Spectral Genomics, Inc.) on a 9700 GeneAmp® PCR System thermal cycler (Applied Biosystems) according to the instructions of the manufacturer. The amplified fragments were separated by electrophoresis with a microfluidics chip (LabChip device; Caliper Technologies, Inc.) on a 2100 Bioanalyzer (Agilent Technologies, Inc.). The fingerprint patterns were determined using the DiversiLab software v3.4. To confirm the conformity of the method all isolates were run in duplicates in separate runs.

Pulsed-field gel electrophoresis (paper I)
PFGE had previously been performed on 26 of the isolates in paper I as described previously (269). Restriction endonucleases SpeI (Bio-Rad Laboratories) and NheI (Roche Diagnostics) were used, as recommended by Bygraves et al. (241). The fingerprints were visually inspected and considered genetically indistinguishable if no bands differed and designated as closely related if their fingerprints differed by no more than three bands (270).

Highly variable multilocus variable-number tandem repeat analysis (paper I and III)
Four highly variable variable-number tandem repeat (VNTR) loci (VNTR4-4, VNTR9-2, VNTR4-2, and VNTR4-3) were analyzed in a HV-MLA as previously described by Schouls et al. (215). The PCR was slightly modified: the multiplex reactions (total volume of 20 µL) contained 1 µL template DNA, 10 µL HotStarTaq mastermix (Qiagen) and the final primer concentrations were 0.4 µM for loci VNTR9-2 and VNTR4-2, 1.6 µM for VNTR4-4 and 0.2 µM for VNTR4-3. The PCR was performed on a 9700 or 2720 GeneAmp® PCR System thermal cycler (Applied Biosystems) and the amplification cycles were changed to 30 cycles consisting of 20 s at 95°C, 60 s at 54°C and 60 s at 72°C. After the PCR, the samples were diluted 1:150 in RNase-free water and 2 µL of the diluted samples were mixed with 10 µL of GeneScan LIZ 500 size standard (Applied Biosystems) diluted 1:100 in Hi-Di formamide (Applied Biosystems). The fragments were denatured at 95°C for 5 min before fragment separation. Separation of the fragments was performed on an ABI PRISM 3130xl genetic analyzer (Applied Biosystems) using the GeneScan module with filter set G5. The sizing of the fragments was performed with the Gene-
Mapper software v4.0 (Applied Biosystems). The number of repeats in each VNTR locus was calculated by correlating the results from the fragments in the reference strain MC58 to the results by Schouls et al. (215). All isolates were run in duplicates in separate runs to confirm the stability of the method.

**Sequence-based typing (paper I-III)**
Sequence typing using the genes \( \textit{porA}, \textit{porB}, \textit{fetA}, \textit{fHbp}, \textit{penA} \) and MLST was performed in paper I-III.

**Real-time PCR**
The genes were amplified using a Rotor-Gene Q real-time PCR system (Qiagen), except for the \( \textit{porA} \) gene, which was amplified using the Light-Cycler System PCR (Roche Diagnostics). The PCR primers for the antigen genes and \( \textit{penA} \) gene are listed in Table 2 and for the MLST in Table 3 (Scandinavian Gene Synthesis AB). The real-time PCR programs are shown in Table 4. All PCR programs finished with a melting curve analysis. The MLST genes and the \( \textit{porA} \) and \( \textit{porB} \) genes were amplified using the LightCycler FastStart DNA Master SYBR Green I Kit (Roche Diagnostics) while the \( \textit{penA}, \textit{fetA} \) and \( \textit{fHbp} \) genes were amplified with the Rotor-Gene SYBR Green I kit (Qiagen). Each reaction mixture contained 2 µL of template in a total volume of 20 µL.
Table 2. Primers used in the real-time PCR and DNA sequencing for the antigen sequence typing and *penA* typing.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer</th>
<th>Primer sequence (5’→3’)</th>
<th>Use</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>penA</em></td>
<td>AA1-F</td>
<td>ATCGAACAGGCGACGATGTC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>PCR</td>
<td>(271)</td>
</tr>
<tr>
<td></td>
<td>ModGc-Down3-R</td>
<td>CGGGGATATAACTGCGGCCGTC&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td>(272)</td>
</tr>
<tr>
<td></td>
<td>1R&lt;sup&gt;b&lt;/sup&gt;</td>
<td>GATTAAGACCGGTGTITTTGACGG</td>
<td>PCR/seq</td>
<td>(218)</td>
</tr>
<tr>
<td></td>
<td>F seq</td>
<td>GTTTTCCCCAGTCAGCAGTTGTA</td>
<td>seq</td>
<td>(273)</td>
</tr>
<tr>
<td></td>
<td>R seq</td>
<td>TTGTGAGCGGATAACAAATTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>aUniversal forward sequence: gttttcccagtcacgacgttgta and reverse sequence: ttgtgagcggtaacaatttc were added as adaptors for sequencing.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>bThe <em>penA</em>1R and <em>fHbp</em> 5CE2086mod primers were used as a complement to the <em>penA</em> ModGcDown3-R and <em>fHbp</em> F primer, respectively, when polymorphism in the annealing sites prevented the regular primers from functioning.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>cPB7f2 was used as an extra sequencing primer to cover the final part of the <em>porB</em> gene if necessary. Initially, <em>porB</em> S1, <em>porB</em> S2, 8L and 8U were used for sequencing. seq=sequencing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>seq=sequencing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F1</td>
<td>GGCAGAACATCAAGCCGCGCT</td>
<td>PCR/seq</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R1</td>
<td>TAAAGCCGATACGACGCGAAATC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F2</td>
<td>CGCCACGCAGTCATCGCTTC</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>R2</td>
<td>CAAGCCCGCTTCCTCCATAGC</td>
<td></td>
</tr>
<tr>
<td><em>fetA</em></td>
<td>s12-F</td>
<td>TTCAACTTCCGACAGCGGCCT&lt;sup&gt;a&lt;/sup&gt;</td>
<td>PCR</td>
<td>(274)</td>
</tr>
<tr>
<td></td>
<td>s15-R</td>
<td>TTGCAGCGCGTCRTACAGGCG&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F seq</td>
<td>GTTTTCCCCAGTCAGCAGTTGTA</td>
<td>seq</td>
<td>(273)</td>
</tr>
<tr>
<td></td>
<td>R seq</td>
<td>TTGTGAGCGGATAACAAATTTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>seq=sequencing</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>TGACCTGCCTCATTGATGC</td>
<td>Modified</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R</td>
<td>CGTAAATTATCGTGTTCGAGCGGC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>s2</td>
<td>CAAATCGAATGGACGGGCAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>s3</td>
<td>TGTTCGATTTTTCGCGTTCCCCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>fHbp</em></td>
<td>5CE2086 mod&lt;sup&gt;b&lt;/sup&gt;</td>
<td>TATGACTTAGGAGYAAACCTG</td>
<td>PCR/seq</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>TGACCTGCCTCATTGATGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>CGTAAATTATCGTGTTCGAGCGGC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>s2</td>
<td>CAAATCGAATGGACGGGCAG</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>s3</td>
<td>TGTTCGATTTTTCGCGTTCCCCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>porB</em></td>
<td>S1</td>
<td>GCAGCCCTCTCCTGTCGC</td>
<td>PCR/seq</td>
<td>(274)</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>TTGCAGATTAGAAATTGTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>8L</td>
<td>GGAGAATCTCGTACGCTAGGG</td>
<td>seq</td>
<td></td>
</tr>
<tr>
<td></td>
<td>8U</td>
<td>TCCGTACGCTACGATTTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PB7f2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>TYGGCAACGTAACGC</td>
<td>This thesis</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Universal forward sequence: gttttcccagtcacgacgttgta and reverse sequence: ttgtgagcggtaacaatttc were added as adaptors for sequencing.

<sup>b</sup>The *penA*1R and *fHbp* 5CE2086mod primers were used as a complement to the *penA* ModGcDown3-R and *fHbp* F primer, respectively, when polymorphism in the annealing sites prevented the regular primers from functioning.

<sup>c</sup>PB7f2 was used as an extra sequencing primer to cover the final part of the *porB* gene if necessary. Initially, *porB* S1, *porB* S2, 8L and 8U were used for sequencing.

seq=sequencing
Table 3. Primers used in the real-time PCR and DNA sequencing for the multi-locus sequence typing (MLST).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Primer sequence (5’→3’)</th>
<th>Use</th>
<th>Amplicon length&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>abcZ</td>
<td>P2C</td>
<td>TCCCCGTCGTAACAAAACAATC</td>
<td>PCR</td>
<td>856 bp</td>
</tr>
<tr>
<td></td>
<td>P1C</td>
<td>TGTTCGCCTCGACTGCAAC</td>
<td>PCR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S1A</td>
<td>AATCGTATATGACCGCAGR</td>
<td>seq</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>GAGAACCAGGCCGGATAGGA</td>
<td>seq</td>
<td></td>
</tr>
<tr>
<td>adk</td>
<td>P1B</td>
<td>CCAAGGCGTGAGAAATCGTAAC</td>
<td>PCR</td>
<td>697 bp</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>CAATACCTCGGTTTCACGG</td>
<td>PCR/seq</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S1A</td>
<td>AGGCWGGCACGCCCTTGG</td>
<td>seq</td>
<td></td>
</tr>
<tr>
<td>aroE</td>
<td>P1B</td>
<td>TTTGAACACGCCTGTTGCCG</td>
<td>PCR</td>
<td>835 bp</td>
</tr>
<tr>
<td></td>
<td>P2B</td>
<td>CAGCGTATCCAGTGCGAC</td>
<td>PCR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S1A</td>
<td>GCGGTCAAYACCGTGRTK</td>
<td>seq</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>ATGATGTGCCGTACACATA</td>
<td>seq</td>
<td></td>
</tr>
<tr>
<td>fumC</td>
<td>S1</td>
<td>TCCGCTGTCGCTTTTGTCAG</td>
<td>PCR/seq</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td>TTGAGGCGGTTTTGCGAC</td>
<td>PCR/seq</td>
<td></td>
</tr>
<tr>
<td>gdh</td>
<td>F</td>
<td>CTGCCCTCGGTTTTTCATCT</td>
<td>PCR</td>
<td>677 bp</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>TGTTCGCGTTATTTCAAGAAGG</td>
<td>PCR</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S4C</td>
<td>RGCACCGGATTCAATYGG</td>
<td>seq</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td>CCTTGCAAAGAAAGCCTGC</td>
<td>seq</td>
<td></td>
</tr>
<tr>
<td>pdhC</td>
<td>F</td>
<td>CCGGCGTGACCGTGCGAA</td>
<td>PCR</td>
<td>818 bp</td>
</tr>
<tr>
<td></td>
<td>R</td>
<td>GATGTGGACGGAATGGGCCAAA</td>
<td>PCR/seq</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S1</td>
<td>TCTACTACATCACCCTGATG</td>
<td>seq</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S2A</td>
<td>GGTGATGATTTCGTYGCRCC</td>
<td>PCR/seq</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S1</td>
<td>GCGCGATGCGCCGACCCTTGG</td>
<td>seq</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>All primers were designed according to the MLST guidance protocols (http://pubmlst.org/neisseria/info/primers.shtml).

<sup>b</sup>The amplicon lengths in MC58.

seq=sequencing
Table 4. Protocol for amplification of the antigen typing genes, the multilocus sequence typing genes and the penA gene.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer concentration (µM)</th>
<th>Activation</th>
<th>Denaturation</th>
<th>Annealing x 40</th>
<th>Elongation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F&lt;sup&gt;a&lt;/sup&gt;</td>
<td>R&lt;sup&gt;b&lt;/sup&gt;</td>
<td>F&lt;sup&gt;a&lt;/sup&gt;</td>
<td>R&lt;sup&gt;b&lt;/sup&gt;</td>
<td>F&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>abcZ</td>
<td>0.5</td>
<td>0.5</td>
<td>95°C 10 min</td>
<td>95°C 10 s</td>
<td>59°C 10 s</td>
</tr>
<tr>
<td>adk</td>
<td>0.3</td>
<td>0.5</td>
<td>95°C 10 min</td>
<td>95°C 10 s</td>
<td>59°C 10 s</td>
</tr>
<tr>
<td>aroE</td>
<td>0.5</td>
<td>0.5</td>
<td>95°C 10 min</td>
<td>95°C 10 s</td>
<td>59°C 10 s</td>
</tr>
<tr>
<td>gdh</td>
<td>0.5</td>
<td>0.5</td>
<td>95°C 10 min</td>
<td>95°C 10 s</td>
<td>59°C 10 s</td>
</tr>
<tr>
<td>pdhC</td>
<td>0.2&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.12&lt;sup&gt;d&lt;/sup&gt;</td>
<td>95°C 10 min</td>
<td>95°C 10 s</td>
<td>59°C 10 s</td>
</tr>
<tr>
<td>pgm</td>
<td>0.1</td>
<td>0.12&lt;sup&gt;d&lt;/sup&gt;</td>
<td>95°C 10 min</td>
<td>95°C 10 s</td>
<td>59°C 10 s</td>
</tr>
<tr>
<td>porB</td>
<td>0.3</td>
<td>0.9</td>
<td>95°C 10 min</td>
<td>95°C 10 s</td>
<td>59°C 10 s</td>
</tr>
<tr>
<td>porA</td>
<td>0.2&lt;sup&gt;e&lt;/sup&gt;</td>
<td>0.12&lt;sup&gt;f&lt;/sup&gt;</td>
<td>95°C 10 min</td>
<td>95°C 10 s</td>
<td>59°C 10 s</td>
</tr>
<tr>
<td>penA</td>
<td>0.6</td>
<td>0.6</td>
<td>95°C 10 min</td>
<td>95°C 10 s</td>
<td>59°C 10 s</td>
</tr>
<tr>
<td>fetA</td>
<td>0.3</td>
<td>0.9</td>
<td>95°C 10 min</td>
<td>95°C 10 s</td>
<td>59°C 10 s</td>
</tr>
<tr>
<td>fHbp</td>
<td>0.6</td>
<td>0.6</td>
<td>95°C 10 min</td>
<td>95°C 10 s</td>
<td>59°C 10 s</td>
</tr>
</tbody>
</table>

<sup>a</sup>Forward primer
<sup>b</sup>Reverse primer
<sup>c</sup>Primer F1
<sup>d</sup>Primer F2
<sup>e</sup>Primer R1
<sup>f</sup>Primer R2

**DNA sequencing**

The PCR products were purified before cycle sequencing in a MultiScreen® PCRμ96 Plate (Millipore) consistent with the manufacturer’s instructions. Subsequently, the cycle sequencing PCR reactions were performed in 96-well plates with the Big Dye Terminator v3.1 Cycle sequencing Kit (v1.1 was used for the cycle sequencing of the porA gene) (Applied Biosystems) in accordance with the instructions of the manufacturer. The primers used for the cycle sequencing PCR are shown in Table 2 and 3. A 2720 or a 9700 Gene-Amp® PCR System thermal cycler (Applied Biosystems) was used to perform 25 cycles of denaturation at 96°C for 10 s, annealing at 50°C for 5 s (59°C for the porA gene) and elongation at 60°C for 4 min. The subsequent products from the cycle sequencing PCR were purified using an ethanol/sodium acetate precipitation and resuspended in 10 µL of formamide (Applied Biosystems) just before capillary electrophoresis. Finally, the nucleotide sequences of each gene were determined by capillary electrophoresis using an ABI PRISM 3130xl genetic analyzer (Applied Biosystems).
Sequencing analysis
The sequences were assembled and manually corrected using the ChromaPro software v1.33 (Technelysium Pty Ltd.). Allele numbers were assigned to each sequence using the *N. meningitidis* sequence query database (pubmlst.org/neisseria/). The PorA VR3 was determined using the PorA VRs database (http://exon.niaid.nih.gov/meningitidis/index.html#home).

Whole-genome sequencing (paper IV)

Sample preparation
The DNA concentrations from the genomic DNA preparations were measured using a PicoGreen® (InvitrogenTM) and the quality of the sample was assessed using a 1% E-Gel® EX (InvitrogenTM). Fragmentation was performed on 500 ng of DNA using an EpisonicTM system (Epi-gentek) with a process time of 200 sec and 10 duty cycles. The distribution of fragments was determined using a Tapestation D1200 system (Agilent/Lab901).

Library preparation and sequencing
Libraries were constructed using the NEBNext® DNA Sample Prep Master Mix Set 1 Kit (New England BioLabs®) according to the manufacturer’s instructions. Ligation of adapters was performed using Illumina® Adapters (Multiplexing Sample Preparation Oligonucleotide Kit). Ligated libraries were size-selected using AMPure® magnetic beads (Agencourt®) and subsequently PCR enriched. The PCR enrichment and adapter extension of each preparation were obtained using 5 μL of size selected library in a 50 μL PCR reaction. After 10 cycles of amplification, as per the instructions of the manufacturer, the reactions were purified with AMPure® XP beads (Agencourt Bioscience Corporation). The final size distribution was determined using a Tapestation 1DK system (Agilent/Lab901). The concentrations used to generate the multiplex pool were determined by PicoGreen® (InvitrogenTM). The library resulting from the pooling was quantified using Agilent qPCR Library Quantification Kit and a MX3005PTM instrument (Agilent). The sequencing was then performed on a HiSeq 2000 to generate 100 bp paired-end reads.
Data analysis

*de novo* assembly was performed using an automated assembly pipeline developed by James Bray (Oxford University, Department of Zoology) with the Velvet assembly package (258). Contigs were then uploaded to the PubMLST database (pubmlst.org/neisseria/), which runs on the Bacterial Isolate Genome Sequence Database (BIGSdb) platform (276). Scanning against alleles defined in the PubMLST sequence definition database was performed automatically using the PubMLST BIGSdb software. New alleles were manually curated and assigned new allele numbers.

Inference networks (paper I-IV)

The different strategies used for the cluster analyses in all papers are summarized in Figure 16.

**Figure 16.** Workflow of the cluster analyses used in papers I-IV. Each method or coefficient used in each experiment is shown in dashed ovals.

In paper I cluster analysis of the HV-MLVA typing data was displayed in a MST created with the Bionumerics software version 6.6 (Applied Maths). The distance matrix was created using a categorical coefficient; however, for HV-MLVA, differences in copy numbers were used instead
of allele differences. The rep-PCR results in paper II were displayed in a dendrogram created with Pearson’s correlation coefficient and UPGMA. Isolates with a similarity of <98% were considered different and isolates with a similarity of >99% were considered indistinguishable. Similarities between 98-99% were assessed manually using the pattern overlay option in the software. There is no general consensus on how to assess outbreaks using DiversiLab; however, the manufacturer recommends that the sample relationships be defined similarly to those of PFGE. According to the manufacturer, isolates with a similarity of >95% should be considered similar; however, the population structure and inherent diversity of the isolates dictate the interpretation criteria.

In paper II a MST was created (using Prim’s algorithm) based on the STs with a MLST analysis tool incorporated into the pubMLST database. The different porA, porB, penA, fetA and fHbp types were then distributed over the different STs in the tree to describe how the different types varied within each ST.

In paper III the MLST data were displayed in a MST created with BioNumerics software version 7.1 (Applied Maths) using a categorical coefficient. Isolates belonging to the same CCs were partitioned into different halos. The dendrogram in paper III that was constructed using the HV-MLVA results was based on a similarity matrix determined by a categorical coefficient and clustered according to the UPGMA. A similarity over 75% (one single-locus variant) was used for cluster designation.

In paper IV a splits network using the Neighbour-Net algorithm (Neighbour-Net is available as part of SplitsTree4) (261) was used to resolve the isolates into networks. The distance matrix was created based on a categorical coefficient according to the number of variable alleles, as in the MLST approach.

**Statistical analyses**

**Discriminatory power**

To assess the discriminative ability of different genes or methods in paper I-III Simpson’s index of diversity (D) (277) was applied. The D-index determines the probability that two randomly picked strains are separated into different groups. With this index, 1 means that all isolates are divided into different groups and 0 indicates that all isolates belong to the same group.
**Congruence between methods**

To compare the different microbial typing methods in paper I adjusted Rand (AR) and adjusted Wallace (AW) coefficients were calculated. The AR coefficient measures the congruence between two typing methods, whereas the AW coefficient provides an estimate of how much information is obtained from another typing method, given a particular typing method (278, 279).
Results and discussion

Molecular typing schemes (Paper I-III)

Surveillance (Paper II and III)
In paper II the discriminatory capacity of porA, fetA, porB, penA, fHbp and MLST was assessed using 85 serogroup Y isolates from Sweden (2000-2010). The highest level of discrimination between strains was achieved using the porA gene (discrimination index of 0.65 [95% CI 0.57-0.74]) and lowest level with the fHbp gene (discrimination index of 0.16 [95% CI 0.05-0.26]). When also including invasive isolates of serogroups A, B, C, E and W in paper III, the Simpson’s index of diversity of the 208 isolates was highest for porA and lowest for serogrouping (Table 5). The highest discriminatory power was found by combining serogrouping, MLST, porA and fetA. Overall, all combinations of different typing targets generated similar indices of diversity, and none of the combinations of three or four targets were significantly more diversifying. Consequently, the same combination, but with fHbp instead of fetA, was found to yield a statistically insignificant difference in the diversity index. The present results are in accord with previous findings regarding efficient routine characterization, such as the study of Lucidarme et al. (280) on 613 invasive isolates, where fHbp had significantly better resolving power than fetA. Thus, in light of the current vaccine against serogroup B disease (BexSero) that includes fHbp as a vaccine component the current routine typing target fetA could be replaced by fHbp.
Table 5. The discriminative ability of serogrouping, MLST and the \textit{porA}, \textit{porB}, \textit{fetA}, \textit{fHbp} and \textit{penA} genes.

<table>
<thead>
<tr>
<th>Typing scheme</th>
<th>Number of types</th>
<th>Simpson's ID</th>
<th>CI (95%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serogroup-\textit{porA}-MLST-fetA</td>
<td>102</td>
<td>0.956</td>
<td>(0.940-0.972)</td>
</tr>
<tr>
<td>Serogroup-\textit{porA}-MLST-porB</td>
<td>100</td>
<td>0.953</td>
<td>(0.937-0.970)</td>
</tr>
<tr>
<td>Serogroup-\textit{porA}-fetA</td>
<td>94</td>
<td>0.952</td>
<td>(0.935-0.969)</td>
</tr>
<tr>
<td>Serogroup-\textit{porA}-MLST-penA</td>
<td>95</td>
<td>0.952</td>
<td>(0.936-0.968)</td>
</tr>
<tr>
<td>Serogroup-\textit{porA}-porB</td>
<td>92</td>
<td>0.950</td>
<td>(0.934-0.967)</td>
</tr>
<tr>
<td>Serogroup-\textit{porA}-MLST-\textit{fHbp}</td>
<td>93</td>
<td>0.950</td>
<td>(0.933-0.967)</td>
</tr>
<tr>
<td>Serogroup-\textit{porA}-MLST</td>
<td>91</td>
<td>0.949</td>
<td>(0.932-0.966)</td>
</tr>
<tr>
<td>Serogroup-\textit{porA}-\textit{penA}</td>
<td>84</td>
<td>0.946</td>
<td>(0.929-0.964)</td>
</tr>
<tr>
<td>\textit{porA}</td>
<td>39</td>
<td>0.849</td>
<td>(0.811-0.886)</td>
</tr>
<tr>
<td>\textit{porB}</td>
<td>39</td>
<td>0.826</td>
<td>(0.786-0.867)</td>
</tr>
<tr>
<td>MLST</td>
<td>48</td>
<td>0.784</td>
<td>(0.735-0.832)</td>
</tr>
<tr>
<td>\textit{fetA}</td>
<td>27</td>
<td>0.782</td>
<td>(0.741-0.824)</td>
</tr>
<tr>
<td>\textit{fHbp}</td>
<td>25</td>
<td>0.738</td>
<td>(0.687-0.789)</td>
</tr>
<tr>
<td>\textit{penA}</td>
<td>22</td>
<td>0.717</td>
<td>(0.673-0.762)</td>
</tr>
<tr>
<td>Serogroup</td>
<td>6</td>
<td>0.664</td>
<td>(0.628-0.700)</td>
</tr>
</tbody>
</table>

ID=index of diversity
CI=confidence interval

Outbreak investigations (paper I and III)

In paper I, 36 invasive serogroup C isolates were compared using three methodologies which can be used for disease outbreak investigations: the current recommended sequencing of the \textit{porA}, \textit{fetA} and \textit{porB} VRs, HV-MLVA (215) and the commercially available rep-PCR from DiversiLab (bioMérieux). Seventeen of these isolates were spatiotemporally linked in seven clusters (Table 6). Two of these had a confirmed connection between cases and the other five were considered highly suspected outbreaks by clinicians and the County Medical Officer for Communicable Disease Control.

The sequencing results were generally congruent with the spatiotemporal connections (Figure 1, paper I). However, the \textit{porA}, \textit{fetA} and \textit{porB} types found among the possible small outbreaks were also common in the general sporadic meningococcal population. This typing method was therefore not able to identify outbreak-associated isolates. The results from the rep-PCR were in concordance with the PFGE previously performed for outbreak investigations. Nevertheless, the rep-PCR seemed to have difficulties differentiating isolates within different CCs (Figure 1, paper I), with
34/36 isolates belonging to cc11 or cc32, which led to a low level of discrimination. The HV-MLVA showed groupings confirmed by the spatiotemporal connections; however, the high level of discrimination excluded some of the spatiotemporally connected isolates. Thus, a single locus variant (SLV) was allowed for the grouping of this method, which led to a higher correlation with the spatiotemporal connections. However, this resulted in the grouping of isolates that did not share a spatiotemporal link, i.e. the isolates from the spatiotemporal clusters “Eskilstuna 1” and “Eskilstuna 2” (Figure 17) and isolates 02-206 and 03-115 (Figure 2, paper I). Still, these isolates were genetically very similar: only PorA VR3 separated the isolates in spatiotemporal cluster “Eskilstuna 2” (Table 6) and only the \textit{porB} gene separated isolates 02-206 and 03-115 (Figure 1, paper I). Nonetheless, “Eskilstuna 1” and “Eskilstuna 2” were approximately 3 or 4 months apart and isolates 02-206 and 03-115 were from separate counties. Hence, they would normally have never been subjected to an outbreak investigation.

The significantly highest level of diversity (non-overlapping 95% CI) was obtained with the HV-MLVA and the lowest with the rep-PCR (Table 1, paper I). Further, the AR coefficients (Table 1, paper I) suggest that the highest congruence was found between the rep-PCR and the sequencing of \textit{porA}, \textit{fetA} and \textit{porB} and the lowest between HV-MLVA and the rep-PCR. The values of \( AW_{HV-MLVA \text{ with SLV} \rightarrow \text{rep-PCR}} \) and \( AW_{\text{porA-fetA-porB} \rightarrow \text{rep-PCR}} \) were both 1 indicating that HV-MLVA with SLV and \textit{porA-fetA-porB} typing predicted the rep-PCR types at 100% level. These congruencies, however, were unidirectional (Figure 18), partly due to the low number of partitions created with the rep-PCR but also possibly because fingerprint-based methods such as rep-PCR provide no specific DNA sequence information.
Table 6. 17 and 35 spatiotemporally linked isolates from invasive meningococcal disease cases in paper I and III, respectively, and their corresponding genetic profiles (agreeing profiles within the same spatiotemporal cluster are marked in grey).

<table>
<thead>
<tr>
<th>Spatiotemporal cluster</th>
<th>SG</th>
<th>DL</th>
<th>PFGE</th>
<th>HV-MLVA</th>
<th>PorV R1, 2, 3</th>
<th>FetA VR</th>
<th>ST</th>
<th>porB</th>
<th>frhtp</th>
<th>penA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Växjö</td>
<td>C</td>
<td>6</td>
<td>A</td>
<td>39-13-11-18</td>
<td>P1.7,16-29,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Växjö</td>
<td>C</td>
<td>6</td>
<td>A</td>
<td>39-13-11-17</td>
<td>P1.7,16-29,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Göteborg</td>
<td>C</td>
<td>1</td>
<td>C</td>
<td>25-14-27-19</td>
<td>P1.5,2,36-2</td>
<td>F1-1</td>
<td>11</td>
<td>2-2</td>
<td>127</td>
<td>3</td>
</tr>
<tr>
<td>Göteborg</td>
<td>C</td>
<td>1</td>
<td>C</td>
<td>25-14-27-19</td>
<td>P1.5,2,36-2</td>
<td>F1-1</td>
<td>11</td>
<td>2-2</td>
<td>127</td>
<td>3</td>
</tr>
<tr>
<td>Eskilstuna 1</td>
<td>C</td>
<td>6</td>
<td>A</td>
<td>14-16-18-25</td>
<td>P1.7,16-29,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Eskilstuna 1</td>
<td>C</td>
<td>6</td>
<td>A</td>
<td>14-16-18-25</td>
<td>P1.7,16-29,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Eskilstuna 2</td>
<td>C</td>
<td>6</td>
<td>A1</td>
<td>11-16-18-25</td>
<td>P1.7,16-29,34</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Eskilstuna 2</td>
<td>C</td>
<td>6</td>
<td>A</td>
<td>11-16-18-25</td>
<td>P1.7,16-29,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Jönköping</td>
<td>C</td>
<td>6</td>
<td>A3/a2</td>
<td>15-13-10-27</td>
<td>P1.7-2,16-29,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Jönköping</td>
<td>C</td>
<td>6</td>
<td>A/2</td>
<td>15-12-10-27</td>
<td>P1.7-2,16-29,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Jönköping</td>
<td>C</td>
<td>6</td>
<td>A/a2</td>
<td>15-13-10-26</td>
<td>P1.7-2,16-29,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Jönköping</td>
<td>C</td>
<td>6</td>
<td>A/a2</td>
<td>15-13-10-27</td>
<td>P1.7-2,16-29,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Västerås</td>
<td>C</td>
<td>6</td>
<td>A/a1</td>
<td>17-9-11-17</td>
<td>P1.7-2,16-29,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Västerås</td>
<td>C</td>
<td>6</td>
<td>A/a1</td>
<td>17-9-11-17</td>
<td>P1.7-2,16-29,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Falun</td>
<td>C</td>
<td>4</td>
<td>-</td>
<td>20-4-28-14</td>
<td>P1.5-1,10-8,36-2</td>
<td>F3-6</td>
<td>11</td>
<td>2-2</td>
<td>649</td>
<td>3</td>
</tr>
<tr>
<td>Falun</td>
<td>C</td>
<td>4</td>
<td>-</td>
<td>21-8-28-14</td>
<td>P1.5-1,10-8,36-2</td>
<td>F3-6</td>
<td>11</td>
<td>2-2</td>
<td>648</td>
<td>3</td>
</tr>
<tr>
<td>Falun</td>
<td>C</td>
<td>4</td>
<td>-</td>
<td>20-8-28-14</td>
<td>P1.5-1,10-8,36-2</td>
<td>F3-6</td>
<td>11</td>
<td>2-2</td>
<td>648</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>21-10-11-23</td>
<td>P1.7-2,16-29,35</td>
<td>F3-3</td>
<td>8876</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>25-10-16-8</td>
<td>P1.5-2,10-1,36-2</td>
<td>F4-1</td>
<td>23</td>
<td>3-36</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>13-7-6-32</td>
<td>P1.5-1,2-2,36-2</td>
<td>F5-8</td>
<td>23</td>
<td>2-55</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>25-10-16-11</td>
<td>P1.5-2,10-1,36-2</td>
<td>F4-1</td>
<td>23</td>
<td>3-36</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>21-5-10-14</td>
<td>P1.7-16-66,35</td>
<td>F5-69</td>
<td>2796</td>
<td>3-24</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>20-9-23-14</td>
<td>P1.5-2,10-12,36-2</td>
<td>F3-3</td>
<td>32</td>
<td>3-1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>10-13-26-22</td>
<td>P1.5-2,36-2</td>
<td>F3-3</td>
<td>11</td>
<td>2-2</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>10-11-11-38</td>
<td>P1.7,16-29,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>32-14-15-27</td>
<td>P1.7-2,16-29,35</td>
<td>F1-49</td>
<td>11</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>15-13-17-7</td>
<td>P1.5-2,10-1,36-2</td>
<td>F5-8</td>
<td>10209</td>
<td>2-224</td>
<td>380</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>13-14-6-14</td>
<td>P1.5-1,2-2,36-2</td>
<td>F1-18</td>
<td>10210</td>
<td>2-55</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>6</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>13-12-9-16</td>
<td>P1.5-1,2-19,36-2</td>
<td>F5-8</td>
<td>10211</td>
<td>3-36</td>
<td>25</td>
<td>1</td>
</tr>
<tr>
<td>Spatio-temporal cluster</td>
<td>SG</td>
<td>DL</td>
<td>PFGE</td>
<td>HV-MLVA</td>
<td>PorA VR1, 2, 3</td>
<td>FetA VR</td>
<td>ST</td>
<td>porB</td>
<td>fHbp</td>
<td>penA</td>
</tr>
<tr>
<td>-------------------------</td>
<td>----</td>
<td>----</td>
<td>------</td>
<td>---------</td>
<td>---------------</td>
<td>---------</td>
<td>----</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>5-17-24-38</td>
<td>P1.5,2,36-2</td>
<td>F3-3</td>
<td>11</td>
<td>2-2</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>19-14-11-32</td>
<td>P1.7,16-29,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>C</td>
<td>-</td>
<td>-</td>
<td>10-9-25-19</td>
<td>P1.5,1,10-8,36-2</td>
<td>F3-6</td>
<td>11</td>
<td>2-23</td>
<td>669</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>16-9-16-9</td>
<td>P1.5-2,10-12,36-2</td>
<td>F4-1</td>
<td>23</td>
<td>3-36</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>8</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>20-10-18-10</td>
<td>P1.5-2,10-1-36-2</td>
<td>F4-1</td>
<td>23</td>
<td>3-36</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>9</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>33-30-16-8</td>
<td>P1.5-2,10-1,36-2</td>
<td>F4-1</td>
<td>23</td>
<td>3-35</td>
<td>25</td>
<td>392</td>
</tr>
<tr>
<td>9</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>22-10-15-15</td>
<td>P1.5-2,10-1,36-2</td>
<td>F4-1</td>
<td>10098</td>
<td>3-455</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>10</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>18-7-7-20</td>
<td>P1.7,16,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>23-8-7-15</td>
<td>P1.7,16,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-24</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>19-5-14-16</td>
<td>P1.5-1,10-1,36-2</td>
<td>F4-1</td>
<td>1655</td>
<td>3-117</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>11</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>24-13-17-13</td>
<td>P1.5-2,10-1,36-2</td>
<td>F4-1</td>
<td>23</td>
<td>3-36</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>11</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>9-11-17-11</td>
<td>P1.5-2,10-1,36-2</td>
<td>F4-1</td>
<td>23</td>
<td>3-36</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>12</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>21-13-18-10</td>
<td>P1.5-2,10-1,36-2</td>
<td>F4-1</td>
<td>23</td>
<td>3-36</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>12</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>19-5-21-22</td>
<td>P1.5-2,10-1,36-2</td>
<td>F4-1</td>
<td>23</td>
<td>3-36</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>13</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>12-13-14-4</td>
<td>P1.5-2,10-1,36-2</td>
<td>F4-1</td>
<td>23</td>
<td>3-36</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>13</td>
<td>Y</td>
<td>-</td>
<td>-</td>
<td>37-10-15-14</td>
<td>P1.5-2,10-1,36-2</td>
<td>F4-1</td>
<td>23</td>
<td>3-36</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>14</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>14-19-8-15</td>
<td>P1.7-20,..37</td>
<td>F1-5</td>
<td>41</td>
<td>3-64</td>
<td>4</td>
<td>368</td>
</tr>
<tr>
<td>14</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>14-19-8-15</td>
<td>P1.7-20,..37</td>
<td>F1-5</td>
<td>41</td>
<td>3-64</td>
<td>4</td>
<td>368</td>
</tr>
<tr>
<td>14</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>14-19-8-14</td>
<td>P1.7-20,..37</td>
<td>F1-5</td>
<td>41</td>
<td>3-64</td>
<td>4</td>
<td>368</td>
</tr>
<tr>
<td>15</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>21-8-9-15</td>
<td>P1.7,16-32,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-82</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>15</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>20-8-9-15</td>
<td>P1.7,16-32,35</td>
<td>F3-3</td>
<td>32</td>
<td>3-82</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>12-12-7-12</td>
<td>P1.17-1,23-3,37</td>
<td>F1-34</td>
<td>1127</td>
<td>3-37</td>
<td>14</td>
<td>391</td>
</tr>
<tr>
<td>16</td>
<td>B</td>
<td>-</td>
<td>-</td>
<td>12-12-7-12</td>
<td>P1.17-1,23-3,37</td>
<td>F1-34</td>
<td>1127</td>
<td>3-37</td>
<td>14</td>
<td>391</td>
</tr>
</tbody>
</table>

SG=serogroup  
DL=DiversiLab  
PFGE=pulsed-field gel electrophoresis (NheI or SpeI)  
HV-MLVA=highly variable multilocus variable-number tandem repeat analysis  
VR=variable region  
ST=sequence type
Figure 17. A minimum spanning tree of the 17 and 35 spatiotemporally linked isolates from invasive meningococcal disease cases in paper I and III, respectively. The city names and the numbers correlate to the spatiotemporal connections in paper I or paper III, respectively. Each circle represents a particular HV-MLVA type and is coloured in relation to the spatiotemporal connection between the isolates. The size of the circles is correlated to the number of isolates belonging to that specific HV-MLVA type. Lines connecting the HV-MLVA types denote a single-locus variant and belong to the same cluster (surrounded by halos).
Figure 18. Congruence between three molecular typing methods used for outbreak investigations determined by the Adjusted Wallace coefficient (95% CI). The arrows represent Adjusted Wallace coefficients > 0.60.

Because of the high discriminatory capacity, cost- and time-efficiency and the overall congruence with the spatiotemporal connections achieved with the HV-MLVA (allowing for a SLV), its usefulness was further evaluated on an extended number of isolates and serogroups in paper III. The HV-MLVA results on 35 spatiotemporally connected N. meningitidis isolates (spatiotemporal clusters 1-16) belonging to serogroups B (n=11), C (n=9) and Y (n=15) in paper III and their corresponding genetic profiles are shown in Table 6. The HV-MLVA clustered seven isolates in three groups when a SLV was allowed for cluster designation (Figure 17). The same three groups, spatiotemporal clusters 14-16, were the only spatiotemporal clusters that were suspected outbreaks by the clinicians and the County Medical Officer for Communicable Disease Control. Spatiotemporal clusters 14 and 16 had a confirmed connection between cases and the cases in spatiotemporal cluster 15 were probably connected because of their similar ages and because they came from the same small community. Figure 17 shows how the HV-MLVA clustered all suspected connected cases from paper I and III. Therefore, in suspected outbreaks where isolates have identical serogroups HV-MLVA can be used as a rapid tool for outbreak investigations.

Limitations
Because of the low frequency of outbreaks and because of the overall low incidence of IMD, the methods could only be tested on a limited amount of isolates (the largest disease outbreak consisted of four isolates). This limitation may even question the use of the expression “outbreak” and thus the term connected cases might be more fitting. However, the WHO
definition of disease outbreak is “the occurrence of cases of disease in excess of what would normally be expected in a defined community, geographical area or season.”

**Genetic characteristics of invasive *N. meningitidis* isolates (paper II-IV)**

**MLST, penA and antigen genes**

Serogroup Y

Because of the emergence of serogroup Y disease in Sweden in the end of the 2000s, the genetic characteristics of *N. meningitidis* serogroup Y in Sweden from 2000 to 2010 were investigated in paper II. Initially, the isolates were typed using the 12 genes: *porA, fetA, porB, fHbp* and *penA* as well as the genes included in the MLST. In paper III-IV, the same typing was performed on serogroup Y isolates from 2011-2012 as well as from 1995-1999, respectively. The collective data of serogroup Y from 1995-2012 (n=190) showed three main dominating strains, all ST-23, in the serogroup Y population. As seen in Figure 19, the dominating strain type, YI, was partly responsible for the increase serogroup Y disease. Although epidemic meningococcal disease is usually associated with clonal strains (281, 282), the overall incidence of IMD in Sweden was not elevated during this period (Figure 6), which is usually a common trait in epidemics.

**Table 7.** Genetic characteristics of three dominating invasive *N. meningitidis* serogroup Y strains in Sweden, 1995-2012.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Frequency (%)</th>
<th>PorA VR1,VR2,VR3</th>
<th>FetA</th>
<th>ST(cc)</th>
<th>porB</th>
<th>fHbp</th>
<th>penA</th>
</tr>
</thead>
<tbody>
<tr>
<td>YI</td>
<td>87/190 (46)</td>
<td>P1.5-2,10-1,36-2</td>
<td>F4-1</td>
<td>ST-23 (cc23)</td>
<td>3-36</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>YII</td>
<td>20/190 (11)</td>
<td>P1.5-1,2-2,36-2</td>
<td>F5-8</td>
<td>ST-23 (cc23)</td>
<td>2-55</td>
<td>25</td>
<td>22</td>
</tr>
<tr>
<td>YIII</td>
<td>7/190 (3.8)</td>
<td>P1.5-1,2-2,36-2</td>
<td>F5-8</td>
<td>ST-23 (cc23)</td>
<td>3-36</td>
<td>25</td>
<td>1</td>
</tr>
</tbody>
</table>

**VR=variable region**

**ST=sequence type**

**cc=clonal complex**
Figure 19. The incidence of the three most common serogroup Y strain types in Sweden and all other serogroup Y isolates per 100,000 population, 1995-2012 (n=190).

The antigenic profiles of the most frequent strain types dominated among the Swedish serogroup Y isolates (Figure 20 and 21). ST-23 represented 82% of all serogroup Y isolates. The different STs found in Sweden 1995-2012 and the dominance of cc23 are further outlined in Figure 22.

Figure 20. Frequency of different PorA and FetA types and MLST sequence types (STs) among invasive N. meningitidis serogroup Y isolates, 1995-2012 (n=190).
**Figure 21.** Distribution of PorA types associated with different FetA types (represented by more than one isolate) of serogroup Y isolates, 1995-2012 (n=176).

**Figure 22.** Minimum spanning tree from MLST profile data for invasive N. meningitidis serogroup Y isolates in Sweden from 1995-2012 (n=190). Each circle represents a MLST sequence type and the size of the circle represents the number of isolates belonging to that particular type. Thick solid lines denote single-locus variants, thin solid lines denote double-locus variants; and dashed and dotted lines indicate types that differ in more than three and four loci, respectively. Isolates belonging to the same CC are surrounded by halos.
As seen in the geographical distribution of the dominating strains in Figure 23, strain type YI has circulated throughout Sweden, as have the other serogroup Y isolates. Strain type YIII isolates were only identified in the middle-east parts of Sweden. The geographical pattern of spread of disease caused by the different strain types indicates that these isolates are circulating and causing disease quite randomly and are neither a result of a temporary localized outbreak nor have they been introduced in a specific part of Sweden.

Other European countries with an increase in serogroup Y disease reported that the median age of IMD caused by this serogroup Y had decreased. The average patient ages in Portugal, Spain, Denmark and Italy was 16-27 years in 2011, which can be compared with 57 years in Sweden (119). In paper II it was shown that the median age of patients with IMD caused by strain type YI in Sweden was 47 years, which was a significantly younger age than the median age of 72 years in patients with IMD caused by isolates with other genetic profiles. However, when the strain type YI
isolates are separated by the approximate time of the increase in serogroup Y, the median age has actually increased in recent years from 20 years to 63 years (Figure 24). This observation further strengthens the belief that the increase in serogroup Y disease is not an epidemic in which the age distribution shifts and the proportion of IMD in adolescents and young adults usually rises (283-285).

![Figure 24](image.png)

**Figure 24.** The median age of IMD cases caused by strain type YI and all other serogroup Y isolates as a function of two periods (n=190).

In paper II sulfamethoxazole was included as a phenotypic marker in the different strain types, and strain type YI was defined as Clone YI (which was sulfamethoxazole resistant). However, in paper III a shift in sulfamethoxazole susceptibility was found in 2010 (Figure 2, paper III); some isolates genetically identical with strain type YI were instead sulfamethoxazole susceptible. Interestingly, no differences in the gene associated with sulphonamide resistance, *folP* (67, 68), were found among the isolates within strain YI (data not shown). Therefore, for a more correct name and to distinguish that these were genetically identical but did not necessarily share the same sulfamethoxazole susceptibilities the clones in paper IV were referred to as strain types.

Serogroups A, B, C, E, W and Y
In paper III isolates from 2010-2012 belonging to other serogroups, as well as serogroup Y, were characterized. The four dominating STs during
this period (Figure 25) were, in decreasing order, ST-23 (serogroup Y), ST-11 (serogroup C), ST-32 (serogroups B and C) and ST-41 (serogroup C). The main antigenic profiles in each serogroup is depicted in Figure 2, paper III.

Figure 25. Distribution of sequence types (STs) associated with different N. meningitidis serogroups (represented by more than one isolate), 1995-2012 (n=163).

Figure 26 shows that serogroup B was most common among the younger patient group, the pattern for serogroup C is somewhat unclear and serogroup Y dominated the older patient group, although 38% of all cases with IMD among 10-19-year-olds were caused by serogroup Y. The pattern for strain YI was also quite indistinct; the highest percentage was found among patients in the 70-79 age group but dominated among serogroup Y isolates in the 0-9 age group. However, there were only two cases of IMD caused by serogroup Y in this age group. Moreover, compared with the European serogroup distribution in different age groups (Figure 5), serogroup B seemed less common among the elderly in Sweden. The trends for the age distribution among serogroups C and Y were seemingly similar.
Whole-genome comparisons of serogroup Y isolates (paper IV)

Sweden

In paper IV 185 draft genomes were produced with approximately 20-50× coverage of each nucleotide (one isolate failed to sequence because of poor DNA quality from the extraction). Of the 1600 loci representing the core genome of *N. meningitidis*, 1387 loci among the strain YI isolates successfully assembled. These loci revealed that strain type YI consists of two subtypes (Figure 2, paper IV), separated by allelic differences in approximately 100 loci. Moreover, the demographic data from the isolates in the two subtypes (referred to as subtype 1 and subtype 2) showed that there was no geographic or age association in the two subtypes; however, the isolates in subtype 1 were restricted to collection dates from 2006 and onwards. Consequently, the increase in serogroup Y disease was particularly due to subtype 1 (Figure 3, paper IV).

The genomes of strain types YI and both YII and YIII differed in more than 500 loci among the 1241 of the 1600 core genes that assembled, indicating that these strains are clearly distinct. Although the clusters obtained from the core genomes, as well as the rMLST, were generally concordant with the previous genetic characterization using 12 genes, the isolates in strain type YII seemed to cluster separately (Figure 1A and C, paper IV). These clusters were separated by approximately 300 genes out of the 1241 core genes, suggesting that the previous characterization may have underestimated the genetic diversity for this particular strain type.

**Figure 26.** Age distribution of the three most common invasive *N. meningitidis* serogroups and the most common serogroup Y strain type, YI, in Sweden, 2010-2012 (*n*=198).
Known or hypothesized virulence genes were compared between different strain types and subtypes. We found that 25 loci had alleles that were distinctly associated with strain type YI and 12 loci differed between the two subtypes of strain YI. Especially lpxL, NEIS1965 (NMB1990) and opcA, which have a role in immune evasion, iron transport and adhesion, respectively, had internal stop codons. This may have had an impact on the virulence or transmission of strain (sub)type YI. In addition, because not all of the genes potentially contained in the Neisseria pathogenome have been discovered, other possible virulence genes outside the core genome or among the incompletely assembled genes could have had an impact on pathogenicity.

**Global population structure**

Different strains dominated in different regions. The dominating strain in each region seemed to be present among isolates in all investigated regions, but at different frequencies (Table 1, paper IV). In Sweden, ST-23 dominated; in England, Wales and Northern Ireland it was ST-1655 (131), and in the USA ST-23 dominated in the beginning of the 1990s, but was replaced by ST-1621 later that decade. Generally, the population structures of both the rMLST and the core genes in Sweden were similar to England, Wales and Northern Ireland (Figure 1, paper IV). The strains from the USA also clustered among strain types YI and YII; however, the late strain type NM233 did not cluster with either of the subtypes. This observation indicates that all strain types are continuously circulating in all populations and the increase in serogroup Y IMD is probably not due to one or more point introductions, but possibly due to other factors within the strain or the host. If the changed serogroup Y incidence pattern is not the result of an introduction of a new strain to which the Swedish population has little immunity, it is possible that something within the genetics of strain type YI changed about the middle of the 2000s. This change could have given rise to subtype 1, which more successively spread through the population and led to the increase in serogroup Y. A genome-wide association study could help explore the genetic mechanisms underlying the strains. The gene expressions may also have to be taken into account, however.

**Limitations**

The present results are limited by the lack of a serogroup Y reference genome in the PubMLST database, the loci that were incompletely assem-
bled and therefore excluded from the Neighbour-Net graphs and the relatively low depth of which the genomes were sequenced. The MLST-like gene-by-gene approach used in paper IV can only show the number of loci with allelic differences (minor or major), which is applicable to large scale population studies with diverse genotypes (276). Accordingly, the resulting Neighbour-Net graphs cannot be interpreted as true phylogenetic trees to predict the evolutionary relatedness between isolates.

Moreover, the serogroup Y isolates in PubMLST were mainly from Sweden, England, Wales and Northern Ireland, and the regions were investigated during different periods.
Conclusions and future perspectives

For routine surveillance of circulating isolates, a molecular typing scheme could include the serogroup B vaccine component fHbp instead of (or in addition to) the currently recommended FetA (paper III). Outbreak investigations require rapid typing methods and enough discriminatory power to identify connected cases, making HV-MLVA a suitable method for small outbreaks (paper I, III).

Molecular typing using 12 genes (MLST, porA, fetA, porB, fHbp and penA) on all invasive N. meningitidis serogroup Y isolates from 1995 to 2012 revealed that the emergence of this serogroup in about 2006 and subsequent spread was due to the expansion of a particular strain type belonging to ST-23. This strain type, YI, represented nearly half of all serogroup Y isolates during this time (paper II-III). Further, genome-wide studies showed that the increase of serogroup Y IMD after 2006 was mainly caused by a specific subtype within strain YI (paper IV).

Genome comparisons with isolates from England, Wales and Northern Ireland, which have had a relatively low serogroup Y IMD, and the USA, which has had a similar serogroup Y expansion as Sweden, showed that the population structures are similar internationally. The most predominant strains in each region, however, are dissimilar; although all belong to cc23, they pertain to different STs. Therefore, the serogroup Y IMD increase in Sweden is not likely due to the introduction of an international strain type but most likely due to multiple factors, including increased virulence or transmission or changes in the host adaptive immune system (paper IV). In addition, no notable increases in overall IMD incidence were noted and the geographical distribution of strain YI was not locally delimited, indicating that this was probably not the result of an epidemic (paper II-III).

Future studies, such as genome-wide association studies, could further reveal genes involved in the success of strain type YI in the Swedish meningococcal population. However, the answer may lie beyond genetics, i.e. the epigenetics and gene expressions may also play an important role. Furthermore, the clinical presentations associated with serogroup Y has not been studied to the extent of determining whether this serogroup is more benign than others.
Acknowledgements

This thesis would never have been possible without the help and support of many people. I would especially like to express my sincere gratitude to all of you who contributed to this thesis, in particular to:

Associate professor **Hans Fredlund**, my first main supervisor, without your enormous kindness I would have never ended up doing my master thesis at the Department of Laboratory Medicine, Microbiology and been introduced to the world of meningococci. Thank you for your expert guidance, encouragement and never-failing enthusiasm.

Associate professor **Paula Mölling**, my second main supervisor, for always inspiring me to jump into the unknown and for your friendly support and all the fun times we have had on all of our many work trips. Thank you for always being available to answer all my burning questions and critically reviewing nearly everything I have ever written.

**Sara Thulin Hedberg**, my assistant supervisor, for your friendly support and always being in such a contagiously good mood. Thank you so much for your endless patience in the lab when I first came to the Department of Laboratory Medicine and for reviewing all the manuscripts.

Professor **Per Olcén**, my former assistant supervisor, for your good mood and inspiring love of meningococci as well as expert guidance, especially in the tricky business of the immunological part of the meningococcal world. Thank you for taking me on as your last PhD student.

Associate professor **Sören Andersson**, Director of the Department of Laboratory Medicine, for giving me the opportunity to perform my PhD studies at the department.

Associate professor **Magnus Unemo**, for all your expert guidance in the field of antibiotic resistance and phylogeny. Thank you for all the intoxicating discussions, for a fruitful collaboration in paper IV, critical review of parts of this thesis and for the PFGE results in paper I.

**Susanne Jacobsson**, for a nice collaboration in paper I and III-IV. Thank you especially for helping me make sense of all the lists and records of
meningococcal isolates, and for your contribution of good mood to our
group of ”princesses” and of course for being such a great roommate at
conferences.

Professor Martin Maiden and his research group at the Department of
Zoology, University of Oxford, for welcoming me to your lab and for a
fruitful collaboration in paper IV.

Anna Önnberg, my colleague and partner in crime in the world of PhD
studies, for all your help in the lab (especially with DOS while I was busy
writing this thesis) and for being a true friend who I can talk to about
everything.

Daniel Golparian, for all your help with my many computer problems and
understanding the new world of whole-genome sequencing and genomic
data analysis. Thanks for all the insightful discussions and of course the
many laughs.

Karin Johansson, my colleague and office roommate, for all the help with
my numerous questions and for being an excellent running coach.

Bengt Hellmark, for your excellent computer support and for installing
the various programs needed to analyse the data included in the present
thesis.

All the PhD students (both current and former): Ronza, Gabriella, Lovisa,
Kerstin, Isak and Sabina, with whom I shared both good times and hard
times. Thank you so much for all your support.

All my colleagues at the Department of Laboratory Medicine for all your
help, support and encouragement, and providing a very conductive at-
mosphere where one looks forward to go to work.

Above all and everything, Mathias, my love, for always making me laugh
and feel so loved. Thank you for being my steady rock and shoulder to cry
on when times were tough. We truly are like two minds ;)

My mother, father, sister and of course my tremendously sweet nephew
Theo, for all your endless love and support. Vă mulțumesc din tot sufletul
pentru încurajări, ajutor și toată dragostea pe care mi-o oferiți. Vezi mamsi, eu m-am dus tot înainte...

All my friends spread all over Sweden, for all the laughs and support, especially Dipti, for our daily calls and all the philosophical and humorous discussions. Thank you for always lifting my spirits and being my go-to person. And Erika, Malin and My for our weekly “fika” which has helped me not to lose myself totally in the world of meningococci.

This thesis was supported by:
The Örebro Council Research Committee
The Foundation for Medical Research “Nyckelfonden”, Örebro University Hospital
The European Disease Society
Federation of European Microbiological Societies
Capio’s Research Foundation
School of Health and Medical Sciences, Örebro University
References


173. Brandtsaeg P. Pathogenesis and pathophysiology of invasive meningococcal disease In: Frosch M, Maiden MCJ, editors.


178. Mickelsen PA, Sparling PF. Ability of Neisseria gonorrhoeae, Neisseria meningitidis, and commensal Neisseria species to obtain iron from transferrin and iron compounds. Infect Immun 1981;33(2):555-64.


BIANCA TÖRÖS  Genome-based characterization of N. meningitidis  83


219. Thulin S, Olcen P, Fredlund H, Unemo M. Combined real-time PCR and pyrosequencing strategy for objective, sensitive, specific, and high-throughput identification of reduced susceptibility to


88 BIANCA TÖRÖS Genome-based characterization of N. meningitidis


35. Söderqvist, Fredrik (2009). Health symptoms and potential effects on the blood-brain and blood-cerebrospinal fluid barriers associated with use of wireless telephones.


41. Gustafsson, Sanna Aila (2010). The importance of being thin – Perceived expectations from self and others and the effect on self-evaluation in girls with disordered eating.

42. Johansson, Bengt (2010). Long-term outcome research on PDR brachytherapy with focus on breast, base of tongue and lip cancer.

43. Tina, Elisabet (2010). Biological markers in breast cancer and acute leukaemia with focus on drug resistance.


46. de Leon, Alex (2010). Effects of Anesthesia on Esophageal Sphincters in Obese Patients.


52. Loiske, Karin (2011). Echocardiographic measurements of the heart. With focus on the right ventricle.


64. Nordin Olsson, Inger (2012). Rational drug treatment in the elderly: "To treat or not to treat".


67. Thuresson, Marie (2012). The Initial Phase of an Acute Coronary Syndrome. Symptoms, patients’ response to symptoms and opportunity to reduce time to seek care and to increase ambulance use.


75. Gustavsson, Anders (2012): Therapy in Inflammatory Bowel Disease.


83. Lönn, Johanna (2013): The role of periodontitis and hepatocyte growth factor in systemic inflammation.


96. Sundh, Josefin (2013): *Quality of life, mortality and exacerbations in COPD.*


98. Palmetun Ekbäck, Maria (2013): *Hirsutism and Quality of Life with Aspects on Social Support, Anxiety and Depression.*


102. Söderström, Ulf (2014): *Type 1 diabetes in children with non-Swedish background – epidemiology and clinical outcome*

103. Wilhelmsson Göstas, Mona (2014): *Psychotherapy patients in mental health care: Attachment styles, interpersonal problems and therapy experiences*


105. Demirel, Isak (2014): *Uropathogenic Escherichia coli, multidrug-resistance and induction of host defense mechanisms*


109. Törös, Bianca (2014): *Genome-based characterization of Neisseria meningitidis with focus on the emergent serogroup Y disease*