GENOMIC ANALYSIS OF SEROGROUP Y *NEISSERIA MENINGITIDIS* ISOLATES

REVEALS EXTENSIVE SIMILARITIES BETWEEN CARRIAGE AND DISEASE-ASSOCIATED ORGANISMS

Neil J. Oldfield,¹ # Odile B. Harrison,² Christopher D. Bayliss,³ Martin C. J. Maiden,² Dlawer A. Ala’Aldeen,¹ and David P. J. Turner¹

¹School of Life Sciences, University of Nottingham, Nottingham NG7 2RD, UK
²Department of Zoology, University of Oxford, Oxford OX1 3PS, UK
³Department of Genetics, University of Leicester, Leicester LE1 7RH, UK

Running Head: Genomic analysis of MenY isolates

Type of article: Major article

Abstract word count: 196

Text only word count: 3499

References: 50

# Correspondence: Centre for Biomolecular Sciences, University of Nottingham, Nottingham NG7 2RD, UK. E-mail neil.oldfield@nottingham.ac.uk; Tel. (+44) 115748 6122; Fax (+44) 115846 8002
ABSTRACT

Background. *Neisseria meningitidis* is a frequent colonizer of the human nasopharynx with asymptomatic carriage providing the reservoir for invasive, disease-causing strains. Serogroup Y (MenY) strains are a major cause of meningococcal disease. High resolution genetic analyses of carriage and disease isolates can establish epidemiological relationships and identify potential virulence factors.

Methods. Whole genome sequence data were obtained from UK MenY carriage isolates from 1997-2010 (n=99). Sequences were compared to those from MenY invasive isolates from 2010 and 2011 (n=73) using a gene-by-gene approach.

Results. Comparisons across 1,605 core genes resolved 91% of isolates into one of eight clusters containing closely related disease and carriage isolates. Six clusters contained carried meningococci isolated in 1997-2001 suggesting temporal stability. One cluster of isolates, predominately sharing the designation Y: P1.5-1,10-1: F4-1: ST-1655 (cc23), was resolved into a sub-cluster with 86% carriage isolates and a second with 90% invasive isolates. These sub-clusters were defined by specific allelic differences in five core genes. Extraction of sequences encoding Bexsero vaccine antigens predicts coverage of 15% of MenY isolates.

Conclusions. High resolution genetic analyses detected long-term temporal stability and temporally-overlapping carriage and disease populations for MenY clones but also evidence of a disease-associated clone.

Keywords: *Neisseria meningitidis*; whole genome sequencing; carriage; serogroup Y; epidemiology
BACKGROUND

*Neisseria meningitidis*, an obligate nasopharyngeal commensal, is carried asymptomatically by 10 to 30% of the adult human population, although these carriage rates are setting dependent and generally higher in young adults and amongst close-contact populations [1, 2]. Occasionally meningococci become invasive and enter the bloodstream potentially leading to the development of septicemia and meningitis. Invasive meningococcal disease (IMD) results in substantial mortality and morbidity despite effective antibiotic treatment [3].

A key virulence factor is the polysaccharide capsule, which allows the bacterium to resist complement-mediated lysis and opsonophagocytosis [4]. Twelve serogroups are recognized based on the biochemical structure of the capsular polysaccharide and genetic analyses [5], with serogroups A, B, C, W, X and Y being responsible for the majority of disease worldwide [6]. DNA sequence-based approaches have been extensively applied to the analysis of the population structure of meningococci [7]. Multilocus sequence typing (MLST), using sequences of seven representative housekeeping genes, has detected a highly structured population with most strains belonging to groups of closely related genotypes referred to as clonal complexes (ccs) [8]. Some of these clonal complexes correspond to ‘hyper-virulent lineages’, which are responsible for most cases of disease worldwide [9, 10]. In addition, clonal complexes are often associated with specific combinations of antigenic proteins, such as Porin A (PorA) and Ferric enterobactin transport protein A (FetA), as well as a limited number of serogroups [11, 12].

Much of the IMD in Europe and North America is caused by a limited range of serogroup/genotype combinations, for example serogroup B (MenB) ST-41/44, ST-32 and ST-269 isolates and serogroup C (MenC) isolates from ST-11 and ST-8 complexes [6, 13]; however, in recent decades the incidence of IMD due to MenY organisms, often belonging to cc23, has
increased in several countries, notably including the USA, Sweden and the United Kingdom [14-18]. In the UK, several carriage studies performed between 2008 and 2012 detected evidence of recent alterations in MenY carriage epidemiology in young adults [19-22]. For example, MenY meningococci were found in only 1-2% of participants and constituted only ca. 10% of recovered isolates when carriage was assessed in 1997-8 in first-year university students at the University of Nottingham, UK and during 1999-2001 in >48,000 15-17 year-old school students throughout the UK [23, 24]. In contrast, in 2008-9 and 2009-10, significantly higher rates of overall carriage, principally resulting from the high prevalence of MenY strains, were detected in university students in Nottingham [19, 20]. These observations were supported by subsequent multisite studies undertaken to investigate carriage in UK school and university students [21, 22]. Identification of isolates in the 2008-9 and 2009-10 Nottingham carriage studies relied on PCR amplification of capsule genes and, while some further typing information was generated for a subset of the 2008-9 isolates [19], only limited information was available on the numbers and genetic background of the different MenY-associated clonal complexes carried in 2009-10.

High resolution analyses of the genome-wide genetic relationships among large numbers of representative carriage and invasive isolates have the potential to determine the prevalence of disease-causing isolates among collections of carriage isolates and to detect specific disease-associated loci. The PubMLST.org/neisseria database, which employs the Bacterial Isolate Genome Sequence database (BIGSdb) platform, is a scalable, open-source web-accessible database, to identify, index and extract genetic variation data from whole genome sequence (WGS) data [25]. This approach was utilized to resolve an outbreak of ST-11 disease [26] and to investigate the evolution and global spread of the ET-5/ST-32 lineage [27], with a recent publication describing MenY disease isolates in Sweden [28]. Additionally, a genealogical
analysis of 108 representative meningococcal genomes led to the proposal of a new ‘lineage’
nomenclature reflecting the increased resolution of WGS typing compared to MLST [29].

Here we investigated the population structure of MenY invasive and carriage isolates in
the UK using WGS data generated from 99 carriage isolates obtained from school or university
students (typically 16 to 20 years old) between 1997 and 2010 and compared this genomic data
with 73 publically available genomes from invasive MenY strains isolated in 2010-11. Extensive
genetic similarities were revealed between invasive and carriage isolates, with isolates forming
distinct clusters, with evidence of temporal stability of these clusters. Notably, discrete invasive-
and carriage-associated sub-clusters were identified within one cluster consistent with distinct
genomic variation occurring within these isolates. WGS data were also analyzed to determine the
potential for coverage of MenY isolates by the newly licensed 4CMenB/Bexsero™ and
rLP2086/Trumenba™ vaccines using a gene-by-gene analysis of all relevant loci.

METHODS

Isolate Selection and Genomic DNA Extraction

A total of 99 MenY isolates, all obtained from nasopharyngeal carriers in Nottingham (East
Midlands), UK, were included in the WGS analysis (Supplementary Table 1). Of these, 77 were
isolated from students attending the University of Nottingham in 2009 [20] and were chosen as
follows: (i) 20 obtained in September 2009 from first-year students; (ii) 18 obtained in
September 2009 from second-year students; (iii) 19 obtained in December 2009 from first-year
students; (iv) 20 obtained in December 2009 from second-year students [20]. To provide context,
10 isolates were chosen randomly from a collection of MenY meningococci isolated from sixth-
form school students in Nottingham in 1999-2001 [24] and six isolates were chosen from MenY
carried isolates obtained from first-year students at the University of Nottingham during 1997-8 [23]. All of these isolates were chosen as known MenY organisms based on PCR or serological typing methods, without prior knowledge of their clonal complex. Six additional MenY carriage isolates were chosen as representative examples of the predominant MenY lineages circulating in a 2008-9 cohort of first-year students at the University of Nottingham [19].

Meningococci were grown overnight on heated horse-blood (‘chocolate’) agar (Oxoid) at 37°C in an atmosphere of air plus 5% CO₂ and genomic DNA extracted using the Wizard Genomic DNA Purification Kit (Promega).

**Illumina Sequencing, Assembly and Accession Numbers**

Genomic DNA was sequenced as described previously [29]. Short-read sequences were assembled using the VelvetOptimiser de novo short-read assembly program optimization script after which resultant contiguous sequences (contigs) were uploaded to the PubMLST.org/neisseria database. Sequence reads were deposited in the European Nucleotide Archive (Supplementary Table 1). Genome sequences of the 73 MenY disease isolates for the epidemiological year 2010-11 in England, Wales and Northern Ireland (Supplementary Table 2) were accessed via the Meningitis Research Foundation Meningococcus Genome Library database (http://pubmlst.org/perl/bigsdb/bigsdb.pl?db=pubmlst_neisseria_mrfgenomes; last analyzed September 2015).

**Genomic Analyses**

The genome assemblies deposited in the database are automatically curated and annotated for all loci currently defined in the database thus identifying alleles with ≥98% sequence identity. Over
2,600 loci were defined at the time of analysis. These have a ‘NEIS’ prefix and are organized into schemes which enables, for example, the rapid identification of isolate genogroup, clonal complex, and PorA and FetA antigen types. Further analysis was undertaken using the BIGSdb Genome Comparator tool implemented within the database using the *N. meningitidis* cgMLST v1.0 core genome scheme (1,605 loci) [29]. Output distance matrices (Nexus format) were used to generate NeighborNet graphs with SplitsTree4 (v4.13.1).

**RESULTS**

**General Features of Sequenced MenY Carriage Genomes**

WGS data were obtained from 99 MenY carriage isolates. After de novo assembly, the 100-bp paired Illumina reads produced contiguous sequences between 2,018,731 bp to 2,214,168 bp in size, consistent with expectations for meningococcal genomes (Supplementary Table 1). Genome assemblies were automatically annotated in a ‘gene-by-gene’ approach using the BIGSdb platform and strain designation data extracted (Supplementary Table 1). Isolates from cc23 predominated (57 of 99), followed by cc174 (18 of 99), cc167 (11 of 99) and cc22 (7 of 99). The most prevalent strain designations were Y: P1.5-1,10-1: F4-1: ST-1655 (cc23), Y: P1.5-1,2-2: F5-8: ST-23 (cc23) and Y: P1.21,16: F3-7: ST-1466 (cc174), which collectively accounted for 48 of these 99 carriage isolates (Table 1). Of the 16 carriage strains isolated in 1997-2001, 11 shared identical strain designations with 2008-10 carriage isolates suggesting persistence of these strain designations over this 7-13 year time period (Table 1).

To investigate the occurrence of these carriage strain designations amongst invasive MenY isolates, identical typing information was extracted from the WGS data of 73 invasive UK MenY meningococci isolated during 2010-11 available via the MRF Meningococcus Genome
Library database (Supplementary Table 2). Isolates from cc23 predominated (58 [79%] of 73), followed by cc174 (7 [10%] of 73), cc167 (4 [5%] of 73) and cc22 (2 [3%] of 73). The most prevalent strain designations among the invasive isolates matched those found in the carriage collection (Table 1). Ten designations were present in both carriage and invasive isolates: these designations accounted for 74% of carriage and 73% of invasive isolates, respectively (Table 1).

WGS Analysis of MenY Isolates Identifies Clusters of Highly Related Isolates

To allow higher resolution genealogical analyses, comparison of all 172 MenY genomes was undertaken using the BIGSdb Genome Comparator tool, the principal output of which is a distance matrix based on the number of variable loci within those loci selected for analysis; these differences were then resolved into a network using standard algorithms [30]. Comparison of the genomes using the core *N. meningitidis* cgMLST v1.0 scheme [29] identified 1,157 loci which varied in at least one isolate and resolved isolates into two distinct groups comprising 56 and 116 isolates, respectively (Figure 1). Only thirteen loci were found to be identical between these two groups: these included loci encoding ribosomal and hypothetical proteins. Within the two groups, distinct clusters of isolates containing multiple examples of both carriage and invasive isolates were evident. Group 1 comprised three clusters, containing isolates belonging to cc167, cc22 and cc174. Group 2 contained only cc23 meningococci, which formed five distinct clusters of carriage and invasive organisms (Figure 1). Overall 91% (157/172) of isolates localized to one of these eight clusters.

Relationships Between Invasive and Carriage MenY Isolates in Identified Clusters
To visualize the relationships between closely related individual isolates, NeighborNet graphs were generated for each cluster with color-coding of isolate names detailing provenance (Figures 2, 3 and 4). Amongst the 25 isolates in the cc174 cluster (Figure 2A), evidence of extensive genetic similarities between carriage isolates was apparent with, for example, only 6 allelic differences distinguishing isolates 22014 and 23214. Highly-related 2009-10 carriage isolates were often isolated from students in the same year group suggestive of intra-year group transmission. This was also apparent in other clusters of isolates, such as cc22 (e.g. isolates 22667 and 21258; 8 allelic differences) (Figure 2C). Conversely, the cc22 cluster revealed highly related meningococci isolated from individuals in different year groups suggestive of inter-year group transmission (e.g. isolates 23009 and 21513; 3 allelic differences) (Figure 2C).

The cc167 cluster (Figure 2B) and cc23 cluster 4 (Figure 3D) each resolved into distinct sub-clusters. Interestingly, the ST-767 cc167 sub-cluster (Figure 2B) contained carriage isolates from 2001, 2008 and 2009 and a 2011 invasive isolate (M11 240071), suggestive of a long-lived clone capable of causing disease. Only 27 allelic differences distinguished M11 240071 from N117.1; 62 differences distinguished the former from NO01020675 – a carriage isolate obtained in 2001 (Figure 2B).

In some cases, clusters containing isolates with identical designations could also be resolved into distinct sub-clusters on the basis of WGS analysis. Notably, cc23 cluster 1 could be resolved into two sub-clusters (Figure 3A). The first contained a carriage isolate from 2000 (NO0010442), five 2008-10 carriage isolates and two 2010-11 invasive isolates. Since NO0010442 is only 34 allelic differences apart from 21251 (a 2009 carriage isolate) and 42 from the invasive isolate M10 240732, this sub-cluster represents another persistent clone, capable of causing disease.
WGS Analysis Resolves cc23 Cluster 5 into Invasive- and Carriage-Associated Sub-clusters

The cc23 cluster 5 contained the largest number of MenY isolates analyzed. Despite predominantly sharing a common strain designation, WGS-based analysis resolved meningococci in this cluster into two sub-clusters (Figure 4): sub-cluster 1 with 18 carriage isolates and three invasive isolates; and sub-cluster 2 with three carriage and 27 invasive meningococci. A total of 997 loci were identical between all cc23 cluster 5 isolates. Loci differing between the two sub-clusters of cc23 cluster 5 are shown in Table 2.

Vaccine Antigen Diversity

Two recombinant-protein based vaccines have been developed with the intention of protecting against MenB disease; widespread use of these vaccines could in principle impact on MenY populations if they protect against carriage. The distribution and variation of the MenB vaccine antigens was surveyed in all 172 isolates (Figure 5, Supplementary Tables 2 and 3). All isolates harbored alleles encoding Neisseria heparin binding antigen (nhaA). Meningococci in the cc174 and cc167 clusters predominantly encoded sub-variants 6 (23 of 25) and 9 (14 of 15), respectively. All cc22 isolates encoded sub-variant 20. Isolates in cc23 cluster 1 typically encoded sub-variant 6 (sub-cluster 1) or 8 (sub-cluster 2), whilst meningococci in the remaining cc23 clusters almost exclusively encoded sub-variant 7. NHBA sub-variant 2, which is present in Bexsero™, was found in one isolate (isolate 20588).

All three main factor H binding protein variants (fHbp-1, fHbp-2 and fHbp-3, and further divided into sub-variants) were identified, but most isolates (163 [95%] of 172) harbored fHbp-2 variants. Meningococci in the cc23 clusters encoded fHbp-2.25 alleles almost exclusively.
Notably, the cc174 cluster contained some meningococci expressing fHbp-1 alleles (mainly fHbp-1.13; 5 [20%] of 25 isolates); fHbp-1.1 (present in Bexsero™), fHbp-1.55 and fHbp-3.45 (present in Trumenba™) were not found in any isolates in this study. No isolates encoded the PorA P1.4 allele present in Bexsero™. The Neisserial adhesin A gene (nadA) was found exclusively in the cc174 isolates; all harbored alleles encoding variant NadA-3 sub-variant 8 (NadA-3.8), matching that present in Bexsero™.

**DISCUSSION**

Nucleotide sequence-based methods involving small numbers of genes have been invaluable in characterizing the population structure and antigenic repertoires of meningococci [31]. The advent of WGS has greatly enhanced resolution and has begun to provide improved insights into the genetic relationships among bacterial isolates [32]. Since carriage is directly relevant to the epidemiology of IMD, we undertook to resolve the genealogical relationships between carriage and invasive isolates. We focused on MenY lineages due to recent observations of fluctuations in MenY disease and carriage levels in the UK. Although meningococci of this serogroup have been less prevalent globally as causes of disease compared to serogroups A, B and C [33], the proportion of IMD attributable to MenY organisms, predominately those belonging to cc23, increased markedly, a trend first recognized in the mid-1990s in the USA [14, 34], and more recently in other countries including the UK [17, 18] and Sweden [15, 35]. The higher MenY IMD case load in the UK was concomitant with a significant increase in MenY carriage, as first detected in studies of nasopharyngeal carriage in students at the University of Nottingham undertaken from 2008 to 2010 [19, 20].
The automated extraction of strain designation information from WGS data demonstrated the similarity of MenY isolates from carriage and invasive disease. This similarity was confirmed by the enhanced discrimination afforded by core genome analysis of the WGS data which resolved most of the isolates into one of eight defined clusters. While most isolates in a particular cluster shared the same strain designation (i.e. ST, PorA and FetA types), each cluster contained variants, demonstrating the enhanced discrimination afforded by WGS. A key finding was that every cluster contained both invasive and disease isolates, indicating that all MenY lineages have the ability to cause disease.

Bacterial populations are often viewed as unstable collections of rapidly evolving clones with frequent extinctions or replacement of older clones. Temporal shifts are potentially important components of IMD epidemiology. Thus, analysis of IMD cases indicated replacement of an ‘early’ cc23 MenY lineage in the USA by an antigenically and genetically distinct ‘late’ strain type [36, 37]. A parallel shift in carriage of these clones was assumed but not confirmed. A significant finding from the present study was that six out of the eight MenY clusters contained historic carriage isolates (i.e. from 1997-2001). The stability of this association appears to be strong as it was detected with only sixteen historic genome sequences. Thus, these six MenY clusters are long-lived and have been present within the UK for a 7-13 year time period. The uneven distribution (e.g. cc167 and cc23 cluster 1) and apparent outlier position (e.g. cc174 and cc22) of historic isolates in some clusters is suggestive of within-cluster evolution over time. The exception to this generalization was cc23 cluster 5, which was the largest cluster and yet contained no historic strain types, potentially suggesting the arrival of a non-UK associated epidemic lineage or major alterations in the genetic structure of a long-lasting UK MenY clone. The presence of long-lasting clones indicates that the genetic structure of meningococcal clones
is stable and that extinctions of clones are rare events. The presence of a long-lived host-adapted commensal population has importance as introduction of the MenACWY vaccine into the main carrier population has the potential to radically-perturb a long-lasting association with unknown consequences.

Evidence for antigenic shifts comes from consideration of the cc23 isolates. These were distributed into five clusters separated by PorA type but not ST type. Four of the clusters differed in sequence for VR2 of PorA, the major target of bactericidal antibodies while the two clusters with identical PorA VR2 sequences had different PorA VR1 sequences, a variable target of bactericidal antibodies. The differences in the VR2 amino acid sequence are amplification of three amino acids (NKQ) from one copy in P1.10-1, to two in P1.10-4 and three in P1.10-10: a rapid and minor change in protein structure. Notably, this is not a feature of all surface antigens as there was limited variation in FetA with four cc23 clusters having the same FetA variant. Further analysis of WGS data may indicate other antigenic variants or allelic variants of other genes that correlate with this segregation of cc23 isolates; nevertheless the PorA distribution suggests that minor differences in antigenicity may be driving changes in population structure.

Geographic distribution of clones and potential sources of new clones was apparent from comparisons between WGS studies in different countries. Comparison of invasive cc23 isolates from Sweden, UK and USA identified three principal cc23 sub-lineages (designated 23.1, 23.2 and 23.3) with overlapping, but differentially prevalent repertoires in each country [28]. For example, the Swedish ‘strain-type YI’ which was largely responsible for the increase in Swedish MenY disease [16, 35], formed a cluster within the 23.1 sub-lineage, but very rarely caused disease in the UK [28]. Using the overlap in MenY WGS data analyzed, *i.e.* UK invasive cc23 strains isolated in 2010-11 examined previously [28], we further resolved the 23.1 sub-lineage
into four sub-clusters (cc23 clusters 2-5, herein) and found that cc23 cluster 1 corresponds to lineage 23.2. Cluster 5 responsible for most cases of UK IMD was rarely observed in Sweden. The resolution of cc23 cluster 5 into distinct carriage- and disease-associated sub-clusters, (1 and 2, respectively) was surprising as this cluster contained the highest number of MenY disease and carriage isolates. A confounding factor is that the sub-cluster 1 carriage isolates were all isolated in one geographical location, and hence may have a high level of one specific (highly transmissible) clone. However, two of these isolates (20601 and 21619) were isolated in September (first week of term) from first-year students who are presumed to have been colonized prior to arrival at the University. An alternative hypothesis is that the ability of sub-cluster 1 strains to cause disease is associated with rapid within host evolution into a sub-cluster 2 phenotype; however sub-clusters were defined by differences in loci encoding proteins with hypothetical or core enzymic functions not loci more explicitly linked to adaptation to a systemic niche (e.g. survival in blood). A further possibility is that sub-cluster 1 has recently evolved from sub-cluster 2 into a highly transmissible carriage strain with a consequent reduction in virulence. A high-quality assembled cc23 genome is required to detect the effects on virulence mediated by genes outside the core alleles utilized in this study and in order to determine how the transition between these sub-clusters has occurred.

The two recently licensed recombinant protein-based anti-MenB vaccines, 4CMenB/Bexsero™ (Novartis now GlaxoSmithKline) [38, 39] and rLP2086/Trumenba™ (Pfizer) [40-42], are predicted to cover the majority of currently circulating invasive MenB strains in Europe and North America [43-47]. Our data predict that Bexsero™ will cover one of the major MenY lineages (cc174), whereas Trumenba™ will have a wider coverage assuming there is cross-reactivity between fHbp main variants 2 and 3. These predictions are consistent
with previous studies [17, 21, 28, 48], but are likely to underestimate coverage due to cross-reactivities between vaccine antigens and those present in MenY isolates [43, 49, 50]. Nevertheless, the introduction of Bexsero™ into the UK infant immunization schedule may only protect against a limited number of currently-circulating MenY strains.

In summary, high resolution genealogical relationships between MenY isolates highlighted the high degree of genetic similarity between carriage and invasive isolates and evidenced long-term stability of MenY clones. The detection and resolution of a highly prevalent UK clone (Y: P1.5-1,10-1: F4-1: ST-1655 cc23) into invasive- and carriage-associated sub-clusters exemplifies the improved precision of whole genome analysis for separating apparently identical isolates.

POTENTIAL CONFLICTS OF INTEREST

All authors report no potential conflicts.

FUNDING

This work was supported by the Wellcome Trust [grant 087622/Z/08/2 to M.C.J.M].

CORRESPONDING AUTHOR CONTACT INFORMATION

Neil Oldfield, Centre for Biomolecular Sciences, University of Nottingham, Nottingham NG7 2RD, UK. E-mail: neil.oldfield@nottingham.ac.uk. Tel: (+44) (+44) 115748 6122. Fax: (+44) 115846 8002.

ACKNOWLEDGMENT
This publication made use of the Meningitis Research Foundation Meningococcus Genome Library (http://www.meningitis.org/research/genome) developed by Public Health England, the Wellcome Trust Sanger Institute and the University of Oxford as a collaboration and funded by Meningitis Research Foundation.
REFERENCES


456 multicomponent vaccine (4CMenB) in Europe: a qualitative and quantitative assessment. Lancet
459 B isolates and estimated coverage of the 4CMenB meningococcal vaccine. BMC Microbiol
460 2014; 14:111.
464 as vaccine antigens in meningococcal endemic and epidemic strain panels from Norway.
FIGURE LEGENDS

**Figure 1.** NeighborNet graph comparison of 172 UK MenY genome sequences analyzed using the BIGSdb Genome Comparator utilizing the *N. meningitidis* cgMLST v1.0 scheme. 91% of isolates analyzed localized to one of eight clusters. Strain designation(s) represent the most frequently occurring designation(s) in each cluster. Unlabeled nodes represent unassigned invasive (n=9) and carriage (n=6) isolates. Scale bar = number of allelic differences.

**Figure 2.** NeighborNet graphs comparing isolates in the (A) cc174, (B) cc167 and (C) cc22 clusters as defined in Figure 1. Sequences were analyzed using BIGSdb Genome Comparator tool utilizing the *N. meningitidis* cgMLST v1.0 scheme. Isolate names are color-coded as follows: 1997-2001 carriage isolates in fuchsia; 2008-9 carriage isolates in black; 2009-10 carriage isolates from first year students in green; 2009-10 carriage isolates from second year students in blue and invasive isolates from 2010-11 in red. Scale bar = number of allelic differences.

**Figure 3.** NeighborNet graphs comparing isolates in the cc23 cluster nos. 1, 2, 3 and 4 (panels A-D, respectively) as defined in Figure 1. Sequences were analyzed using BIGSdb Genome Comparator tool utilizing the *N. meningitidis* cgMLST v1.0 scheme. Isolate names are color-coded according to the scheme described in the Figure 2 legend. Scale bar = number of allelic differences.

**Figure 4.** NeighborNet graph comparison of isolates in the cc23 cluster 5 defined in Figure 1. Sequences were analyzed using BIGSdb Genome Comparator tool utilizing the *N. meningitidis* cgMLST v1.0 scheme.
cgMLST v1.0 scheme. Isolate names are color-coded according to the scheme described in Figure 2 legend. Scale bar = number of allelic differences.

**Figure 5.** Genetic characterization of MenB vaccine antigens in the 172 MenY isolates. (A) Prevalence of NHBA peptides; (B) Prevalence of fHbp alleles; (C) Prevalence of PorA VR2; (D) pie graph of NadA presence and variant/sub-variant. Alternative naming schemes can be cross-referenced at [http://pubmlst.org/neisseria/](http://pubmlst.org/neisseria/).
### Table 1. Frequency of Strain Designations in the MenY Carriage and Invasive Collections

<table>
<thead>
<tr>
<th>Strain designation</th>
<th>Carriage group</th>
<th>Total carriage</th>
<th>Invasive</th>
<th>Total carriage and invasive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1997-2001 (n=16)</td>
<td>2008-2010 (n=83)</td>
<td>2010-11 (n=73)</td>
<td>(n=172)</td>
</tr>
<tr>
<td>Y: P1.5-1,10-1: F4-1: ST-1655 (cc23)</td>
<td>1</td>
<td>20</td>
<td>21</td>
<td>26</td>
</tr>
<tr>
<td>Y: P1.5-1,2-2: F5-8: ST-23 (cc23)</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>Y: P1.21,16: F3-7: ST-1466 (cc174)</td>
<td>0</td>
<td>12</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>Y: P1.5-2,10-1: F4-1: ST-23 (cc23)</td>
<td>1</td>
<td>4</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Y: P1.5-1,10-4: F4-1: ST-1655 (cc23)</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Y: P1.5-1,10-4: F4-1: ST-6463 (cc23)</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Y: P1.5-1,2-2: F5-1: ST-3651 (cc22)</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Y: P1.5-1,10-10: F4-1: ST-1655 (cc23)</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Y: P1.5-1,10-1: F1-3: ST-767 (cc167)</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Y: P1.5-1,10-4: F4-1: ST-23 (cc23)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Y: P1.5-8,10-4: F5-2: ST-168 (cc167)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Y: P1.5-1,10-22: F5-1: ST-114 (cc22)</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Y: P1.5-1,10-46: F3-9: ST-103 (cc103)</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Y: P1.5-1,10-62: F1-3: ST-767 (cc167)</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Y: P1.22,9: F3-7: ST-1466 (cc174)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Other&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3</td>
<td>12</td>
<td>15</td>
<td>16</td>
</tr>
</tbody>
</table>

<sup>a</sup> Includes all strain designations occurring only once.
Table 2. Loci with Allelic Differences between the Two Sub-clusters of cc23 Cluster 5

<table>
<thead>
<tr>
<th>BIGSdb locus identifier</th>
<th>Neisseria locus identifier</th>
<th>Predicted protein/function (gene)</th>
<th>Allele number (%)</th>
<th>% nucleotide identify</th>
<th>Amino acid differences</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sub-cluster 1</td>
<td>Sub-cluster 2</td>
<td></td>
</tr>
<tr>
<td>NEIS0395</td>
<td>Valine-pyruvate transaminase (<em>avtA</em>)</td>
<td>112 (100)</td>
<td>113 (96.7)</td>
<td>99.9</td>
<td>1</td>
</tr>
<tr>
<td>NEIS0825</td>
<td>Superoxide dismutase (<em>sodB</em>)</td>
<td>155 (100)</td>
<td>22 (96.7)</td>
<td>99.8</td>
<td>1</td>
</tr>
<tr>
<td>NEIS0929</td>
<td>Hypothetical protein</td>
<td>42 (100)</td>
<td>3 (100)</td>
<td>99.6</td>
<td>0</td>
</tr>
<tr>
<td>NEIS1199</td>
<td>Glycerate kinase (<em>glxK</em>)</td>
<td>47 (100)</td>
<td>24 (100)</td>
<td>99.9</td>
<td>1</td>
</tr>
<tr>
<td>NEIS1568</td>
<td>Hypothetical protein</td>
<td>67 (100)</td>
<td>68 (96.7)</td>
<td>99.9</td>
<td>1</td>
</tr>
</tbody>
</table>