The Genomic Epidemiology of Hyperinvasive and Carriage-Associated Lineages of *Neisseria meningitidis*

Dorothea M. C. Hill

Hertford College

University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2015

Abstract

Neisseria meningitidis is the bacterial pathogen responsible for meningococcal disease: 10% of cases result in death. Endemic disease is sporadic, with fewer than three cases per 100,000 population, but elevated incidence also occurs during hyperendemic and epidemic periods. Successful control of meningococcal disease is achieved via immunisation against capsular polysaccharides, however, no vaccine exists against serogroup B polysaccharide due to poor immunogenicity and safety concerns.

Recently, 'serogroup B-substitute vaccines' containing sub-capsular antigens have been developed: one such vaccine is to be introduced to UK infant immunization schedules from September 2015. Meningococcal populations are genetically diverse and arranged into clonal complexes that fluctuate in geotemporal prevalence. It is therefore essential to survey meningococcal populations in given locations to understand molecular evolution following vaccine introduction.

The advent of whole-genome sequencing presents an opportunity to investigate meningococcal populations at high genetic resolution. This thesis introduces the MRF Meningococcus Genome Library, a comprehensive collection of WGS data from national surveillance of meningococcal disease in England and Wales between the 2010/11-2012/13 epidemiological years. The utility of WGS data in routine surveillance was investigated.

Structuring of genomic variation revealed population-level lineages analogous to clonal complexes. Prevalent hyperinvasive lineages were composed of immunologically distinct sub-lineages that, when linked with patient data, were age-group associated: vaccines may affect age-groups differently. Sub-lineage specific core-genome and epigenetic DNA methylation variation was primarily in metabolic loci. Finally, genomic variation of an epidemic clone, generated by recombination, was great, and epidemiological data remains indispensible for investigating transmission. These findings provide insight into meningococcal evolution and are of importance in understanding the effects of new vaccines.

Declaration

The work herein is my own except where explicitly stated in 'Publications and Contributions' and in relevant chapters. No part of this work has been submitted for any other degree or professional qualification.

Dorothea M. C. Hill, Trinity 2015.

Acknowledgements

I would like to thank my supervisor, Prof. Martin Maiden, for his guidance and support during the course of this D.Phil. Thanks also goes to all members of the Maiden research group for their input, and in particular to Dr. Odile Harrison and Dr. Carina Brehony for their encouragement. I am grateful for the enthusiastic advice given by Prof. Angela Brueggemann. Finally, my special thanks are extended to the Meningitis Research Foundation without whose backing this D.Phil. would not have been possible. It has been a privilege to have been part of their mission to see a world free from meningitis and septicaemia.

Publications and Contributions

Publications

The following published papers include work presented in this thesis:

- Hill, D. M. C., Lucidarme, J., Gray, S.J., Newbold, S.L., Ure,R., Brehony, C., Harrison,O.B., Bray,J.E., Jolley,K.A., Bratcher,H.B., Parkhill,J., Tang, C.T., Borrow,R., Maiden,M.C.J. (2015) "Genomic epidemiology of age-associated meningococcal lineages in national surveillance: an observational cohort study" Lancet Infect Dis 15(12): 1420-1428

- Seib, K. L., F. E. Jen, A. Tan, A. L. Scott, R. Kumar, P. M. Power, L. T. Chen, H. J. Wu, A. H. Wang, D. M. Hill, Y. A. Luyten, R. D. Morgan, R. J. Roberts, M. C. Maiden, M. Boitano, T. A. Clark, J. Korlach, D. N. Rao and M. P. Jennings (2015) "Specificity of the ModA11, ModA12 and ModD1 epigenetic regulator N6-adenine DNA methyltransferases of *Neisseria meningitidis*" Nucleic Acids Res 43(8): 4150-4162

- Toros, B., S. T. Hedberg, M. Unemo, S. Jacobsson, D. M. Hill, P. Olcen, H. Fredlund, H. B. Bratcher, K. A. Jolley, M. C. Maiden and P. Molling (2015) "Genome-based characterization of emergent invasive *Neisseria meningitidis* serogroup Y in Sweden, 1995 to 2012" J Clin Microbiol 53(7): 2154-2162

- Lucidarme, J., D. M. C. Hill, H. B. Bratcher, S. J. Gray, M. du Plessis, R. S. Tsang, J. A. Vazquez, M. Taha, M. Ceyhan, A. M. Efron, M. C. Gorla, J. Findlow, K. A. Jolley, M.C. Maiden, R. Borrow (2015) “Geo-temporal distribution of distinct strains of the hyperinvasive meningococcal ST-11 clonal complex” J Infect Dis 10.1016/j.jinf.2015.07.007

- Brehony, C., D. M. C. Hill, J. Lucidarme, R. Borrow, M. C. J. Maiden. “Distribution of meningococcal vaccine antigen diversity in global databases” Euro Surveill 20(49).

The following manuscript, under review for publication, includes work in this thesis:

- Tan, A.*, D. M. C. Hill*, O. B. Harrison, Y. N. Srikhanta, M. P. Jennings, M. C. J. Maiden, K. L. Seib "Distribution of the type III DNA methyltransferases *modA*, *modB* and *modD* in Neisseria meningitidis: implications for gene regulation and virulence" Scientific Reports.

I presented the following talks and posters at scientific conferences during the course of my DPhil:

- "Whole-genome epidemiology of the ET-15 clone in the Czech Republic in 1993", oral presentation at the International Pathogenic *Neisseria* Conference (IPNC), in Wurzburg, Germany, 2012.
- "SMRT sequencing reveals phase-variable methylation in ET-15 clone", poster presentation at the American Society of Microbiology (ASM) general meeting, Denver, USA, 2013.
- "Geographic distribution of meningococcal clonal complexes in England and Wales, 2010/11 and 2011/12", poster presentation at the MRF Symposium, Bristol, 2013.
- "MRF meningococcus genome library: population structure", oral presentation at the European Meningococcal Disease Society (EMGM) conference, Bad Loipersdorf, Austria, 2013.
- "MRF meningococcus genome library: whole-genomes in routine surveillance/sharing genomic data via PubMLST.org", oral presentation at the MRF Conference, Royal Society of Medicine, London, 2013.
- "A whole genome view of the association of meningococcal lineages with the age of meningococcal disease patients", poster presentation at the ASM general meeting, Boston, USA, 2014.

I contributed to the following publications during the course of my DPhil:

- Jolley, K. A., D. M. Hill, H. B. Bratcher, O. B. Harrison, I. M. Feavers, J. Parkhill and M. C. J. Maiden (2012) "Resolution of a meningococcal disease outbreak from whole genome sequence data with rapid web-based analysis methods" J Clin Micro 50(9): 3046-3053
- van Tonder, A. J., S. Mistry, J. E. Bray, D. M. Hill, A. J. Cody, C. L. Farmer, K. P. Klugman, A. von Gottberg, S. D. Bentley, J. Parkhill, K. A. Jolley, M. C. Maiden and A. B. Brueggemann (2014) "Defining the estimated core genome of bacterial populations using a bayesian decision model" PLoS Comput Biol 10(8): e1003788

Contributions

MRF Meningococcus Genome Library data

The Meningitis Research Foundation Meningococcus Genome Library (MRF-MGL) whole-genome sequence data were analysed in Chapters 2, 4, and 5. Generation of these data was a collaborative effort: isolates were collected by the Public Health England Meningococcal Reference Unit (PHE-MRU); culture and DNA extraction were carried out by Jay Lucidarme at PHE-MRU; and whole-genome sequencing was carried out at the Wellcome Trust Sanger Institute, overseen by David Harris and Julian Parkhill. In the Maiden group: the majority of genomes were assembled and uploaded to the PubMLST database by Dr James Bray; Dr Odile Harrison assigned genogroups; and Dr Keith Jolley guided my use of the PubMLST database. All other work was my own.

Chapter 3: 1993 Czech Republic ET-15 epidemic

Isolates collected by the Czech Republic National Reference Laboratory were contributed by Dr. Paula Kriz. Culture, DNA extraction, and assignment of genogroups were performed by Dr. Odile Harrison. The Prokka program and BLASTclust were run by Dr. James Bray.

Chapter 4: Phase-variable Type III methyltransferases

SMRT sequencing, assembly of SMRT reads, and identification of consensus modification sites were performed by Tyson Clark at Pacific Biosciences, California. Plasmid constructs, cloning, and SMRT sequencing were performed by Yvette A. Luyten and Richard D. Morgan at New England Biolabs (NEB), New Jersey. Bioinformatic identification of methyltransferases with SeqWare was carried out by Richard Roberts at NEB.

Table of Contents

Chapter 1. Introduction

Despite more than 40 years of extensive meningococcal vaccine research (Gotschlich *et al.* 1969c, Artenstein *et al.* 1970, Greenwood and Wali 1980) and pioneering work on meningococcal population structure (Caugant *et al.* 1986b, Maiden *et al.* 1998, Feil *et al.* 1999), disease caused by the bacterium *Neisseria meningitidis* continues to be responsible for significant mortality and morbidity worldwide. This results from a lack of comprehensive vaccines against the meningococcal genotypes responsible for most disease. There are, however, various theoretical problems that remain incompletely understood. For instance, why does a bacterium with a commensal life history on occasion cause invasive disease? What are the driving forces of variable disease rates worldwide? Why are certain lineages more likely to be invasive? And, how and why are lineages maintained in the face of extensive horizontal gene transfer? In addition to theoretical interest, the evolutionary biology, evolutionary ecology, population genetics, and molecular epidemiology of *Neisseria meningitidis* provide insights of great importance to vaccine development and implementation (Rappuoli 2007).

Since the 1960s, these fields of study have advanced through rapid maturation of the methods for discrimination among meningococcal variants. Following early serological work on capsular and subcapsular meningococcal antigens (Slaterus 1961, Frasch and Chapman 1972), and the development of multilocus enzyme electrophoresis (Selander *et al.* 1986), the addition of Sanger dideoxy sequencing to meningococcal characterisation permitted antigen sequence typing (Feavers *et al.* 1992, Jolley *et al.* 2007) and multilocus sequence typing (Maiden *et al.* 1998). Subsequent exploration of meningococcal population structures continued at previously unimagined resolution, and in recent years, single gene sequencing has been almost supplanted by whole-genome sequencing: it is hoped that longstanding unanswered questions can now be addressed in novel ways.

Thesis aims and outline

The Meningitis Research Foundation Meningococcus Genome Library (MRF-MGL) was created in 2010 as the first comprehensive epidemiological collection of meningococcal whole-genome sequence (WGS) data, intended for "meningococcal vaccine research […] and investigation of pathogenesis and population biology" by "the national and international research community" (MRF, 2010). It contains the WGS of every meningococcus isolated from cases of meningococcal disease in England, Wales, and Northern Ireland in the epidemiological years 2010/11, 2011/12, and 2012/13. In June 2015, this collection remains the only publically available repository of an epidemiologically relevant collection of WGS data from a bacterial pathogen.

Throughout this D.Phil. I have collaborated with PHE-MRU, the Wellcome Trust Sanger Institute, my supervisor Martin Maiden and his research group, and the MRF to ensure that the MRF-MGL was established with reliable, annotated WGS data. In this thesis, I explore the application of these data to the population biology of hyperinvasive meningococcal lineages and to public health efforts against meningococcal disease.

In Chapter 2, I present the MRF-MGL and demonstrate its utility in public health surveillance: analysis of the temporal distribution of age-associated meningococcal lineages is provided as proof of concept. This chapter provides a baseline description of the contents of the MRF-MGL and an accessible analysis methodology for those interested in using the genomes for meningococcal population research.

In Chapter 3, I extend the use of WGS data in public health to meningococcal epidemiology on a micro-evolutionary scale. The 1993 Czech Republic epidemic was caused by a clone that could not be distinguished from the rest of its lineage (the ST-11 clonal complex) by MLST, and whose variants could not be distinguished by MLEE. The work of an established national reference laboratory constrained the epidemic through targeted vaccination, however, insight into the dissemination of

the clone through the country was unavoidably limited. In light of increasingly prevalent ST-11 disease in the UK, I sought to investigate: (i) concordance between MLEE and whole-genome typing of the epidemic clone; (ii) whether WGS data facilitated discrimination among isolates; (iii) geo-temporal genetic structuring of the clone within the Czech Republic; and (iv) the extent to which WGS data will be of use in identifying short chains of lineage 11 transmission during outbreaks.

In Chapter 4, I move on from the inter-host variation of a meningococcal clone to look at random generation of intra-host variation incurred by epigenetic modification. Restriction-modification systems (RMS) have been suggested to stabilize meningococcal lineage structure, and Mod enzymes associated with some type III RMS mediate phase-variable alterations in meningococcal gene expression. In this study, I utilized WGS data to describe the variation and distribution of three *mod* genes among disease and carried meningococci. These data indicated that associations exist among Mod variants and meningococcal lineages and among Mod variants and meningococcal invasiveness, and that a large proportion of genomic loci epigenetically modified by ModA12 are involved in meningococcal metabolism.

In Chapter 5, I use the MRF-MGL to quantify the potential baseline coverage of different age-groups in England and Wales by meningococcal 'serogroup B-substitute' vaccines. Associations of hyperinvasive meningococcal sub-lineages with hosts of different ages, vaccine antigen variant distributions, and the geotemporal trends of such meningococci were explored. These data indicated that multicomponent/multivariant meningococcal vaccines are most likely to effect broad coverage across age-groups in endemic periods. Comprehensive surveillance with WGS data will be invaluable for design and implementation of both broad-spectrum and targeted vaccines in coming years.

In Chapter 6, insights from Chapters 2-5 are discussed and future directions are proposed. In the remainder of this chapter, I review aspects of the molecular epidemiology and population biology of *N. meningitidis* relevant to the concepts and findings in this thesis.

Neisseria meningitidis

Neisseria meningitidis is one of two human pathogens within the *Neisseria*, a genus belonging to the beta class of gram-negative proteobacteria. Both *N. meningitidis*, commonly referred to as the meningococcus and responsible for severe sepsis and meningitis, and *N. gonorrhoeae,* the agent of gonorrhea, have human restricted life histories, as do the well characterized commensal species *N. lactamica*, *N. sicca*, and *N. flavescens* (Sarafian and Knapp 1989, Bennett *et al.* 2012). An unknown number of *Neisseria* species exist in animals, with *N. macacae* of recent interest for its use as a model for *Neisseria* –host interactions (Vedros *et al.* 1983, Barrett and Sneath 1994, Weyand *et al.* 2013).

Disease caused by the meningococcus was first described in the early 19th century (Vieusseux 1805, Arkwright 1907, North 1980). Although it is unlikely that pathogenic meningococci emerged much before this time given the lack of recorded meningococcal disease incidence, dating the evolutionary divergence of neisserial species has proven complex (Schoen *et al.* 2008). It has been suggested that meningococcal capsule acquisition may have coincided with development of a pathogenic lifestyle, and that *N. gonorrhoeae* is a clone that diverged from the ancestor of *N. meningitidis* and *N. lactamica* following a switch from the nasopharynx to the urogenital niche (Vazquez *et al.* 1993, Claus *et al.* 2002, Vogel *et al.* 2004, Schoen *et al.* 2008, Bennett *et al.* 2010).

Host-meningococcal interactions

The cellular and molecular factors that permit meningococcal survival within its niche have largely been elucidated *in vitro*; given the human specificity of the meningococcus and the lack of an suitable animal model, a great deal of work on meningococcal pathogenicity has been achieved using human cell lines or organ cultures (Stephens and Farley 1991, Gorringe *et al.* 2005, Sjolinder and Jonsson 2007, Melican *et al.* 2013). In coming years, the wealth of whole-genome sequence (WGS) data from many bacterial species will permit prediction of proteins involved in a variety of

meningococcal functions. It is clear, however, that adhesion to host tissues, acquisition of nutrients such as iron, and evasion of immune clearance are key to meningococcal survival. These take place through a series of incompletely understood mechanisms dependent on meningococcal surface structures and proteins (Virji 2009, Hill *et al.* 2010).

Transmission and colonization

Despite its medical significance, the meningococcus, like the human commensal *Neisseria* species, is primarily adapted to harmless colonization of the nasopharyngeal epithelium. Following acquisition, usually through inhalation of respiratory droplets, the most common outcome of infection is transient and asymptomatic carriage. Initial colonization is established through adhesion to epithelial cells to avoid mucus clearance and secretion of IgA protease, which cleaves immunoglobulin (Mulks *et al.* 1980, Virji 2009).

The capsule and LPS

The outermost parts of meningococcal cells consist of a cytoplasmic membrane, a peptidoglycan cell wall, and an outer-membrane. The majority of meningococci are able to produce a polysaccharide capsule surrounding the outer-membrane. Both the capsule and lipopolysaccharide (LPS) structures that extend from the outer-membrane may influence initial adhesion to host epithelia (Jones *et al.* 2003). LPS is composed of an inner and outer oligosaccharide core attached to lipid A and is highly immunogenic; at least 12 structural variants have been identified and classified into LPS immunotypes (Scholten *et al.* 1994a).

The meningococcal capsule protects the cell between and within hosts. Although not essential in transmission, the capsule is highly hydrated and may protect against desiccation when airborne (Diaz Romero and Outschoorn 1994, Tzeng *et al.* 2013). Within hosts the capsule protects against the immune system and the sialic acid composition of capsules of the most invasive meningococci reduces immune system recognition through host cell mimicry (Mandrell *et al.* 1988, Diaz Romero and Outschoorn 1994). Certain LPS immunotypes (L3, L7, L9) may also be sialylated (Mandrell *et al.*

1988, Jones *et al.* 1992). In the respiratory tract capsule expression may be down-regulated to fully expose subcapsular structures involved in host-meningococci interactions (Hammerschmidt *et al.* 1996b, Deghmane *et al.* 2002).

Adhesins

The adhesins are key to initial colonization of non-ciliated epithelial cells (Stephens and Farley 1991, Virji 2009). *In vitro* observations suggest that the 'major adhesins' effect quantifiably greater interactions than the 'minor adhesins', more recently identified using WGS data (Virji 2009). Major adhesins include pili and outer-membrane opacity proteins. Meningococcal pili are filamentous hair-like structures that project from the cell surface and are considered to mediate primary host interactions (Merz and So 2000, Proft and Baker 2009). They are composed of pilin subunits encoded by *pilE* (Morand and Rudel 2006). A number of additional *pil* genes are implicated in pilin biosynthesis and pilin-associated functions: for example, inter- and intra-genomic recombination of *pilE* with silent pilin genes, *pilS*, results in alterations to pilin structure (Segal *et al.* 1986), the product of *pilC* may increase adhesion, and *pilQ* forms pores through which pili are translocated (Virji 1997, Morand and Rudel 2006, Virji 2009). Meningococcal pili also aid DNA uptake through transformation and conjugation and, since they belong to the type IV family, switching motility. The opacity proteins, Opa and Opc, initially known as Class 5 proteins, are encoded by up to four *opa* loci and one *opcA* locus in single isolates: these are involved in both adhesion to and invasion of host epithelial cells (Aho *et al.* 1991, Virji *et al.* 1993).

Minor adhesins include the autotransporter proteins NhhA, App, NadA, and MspA, and a haemagglutinin/haemolysin-related protein, HrpA, that forms part of a two-partner secretion system (Comanducci *et al.* 2002, Serruto *et al.* 2003, Scarselli *et al.* 2006, Turner *et al.* 2006, Schmitt *et al.* 2007). Nhha and App have been considered as potential vaccine candidates and NadA is a component of the newly licensed Bexsero® vaccine (Serruto *et al.* 2003, Bambini *et al.* 2009, Peak *et al.* 2013). There is some diversity among NadA peptide sequences, examined in Chapters 2 and 5,

with peptides belonging to one of four recently updated variants that are classified into two groups (Bambini *et al.* 2014). NadA expression is repressed by the regulator NadR under growth conditions in vitro, and is induced in vivo (Fagnocchi *et al.* 2012, Fagnocchi *et al.* 2013).

Nutrient acquisition

Outer-membrane associated proteins such as porins and iron-binding proteins are crucial for nutrient acquisition. PorA and PorB, formerly class 1 and class 2/3 proteins respectively, are abundant outer-membrane proteins (Tommassen *et al.* 1990) that form β-barrel structures through which hydrophilic ions diffuse (Derrick *et al.* 1999). Both PorA and PorB have eight surface exposed loops; two PorA loops are of particular interest because they contain immunologically important epitopes, variable region (VR) 1 and VR2 (McGuinness *et al.* 1990, Derrick *et al.* 1999). A great deal of work has been carried out on these antigens since, despite their variability, they are highly immunogenic and are vaccine components (van den Dobbelsteen *et al.* 2007). PorB is a less important potential vaccine constituent; the class 3 form, PorB3, may be poorly accessible to antibody binding (Sacchi *et al.* 1998, Michaelsen *et al.* 2001).

Iron is a crucial growth factor during colonization and disease and the meningococcus possesses a variety of Fur-regulated outer-membrane iron transporters (Archibald and DeVoe 1978, Perkins-Balding *et al.* 2004). Since the majority of iron in the human body is stored intracellularly in ferritin and haemoglobin, the major source of iron available to meningococci is complexed to host iron-binding proteins (Perkins-Balding *et al.* 2004). In the nasopharynx, the predominant iron source is lactoferrin, and LbpA and LbpB comprise the meningococcal lactoferrin receptor (Prinz *et al.* 1999). Transferrin is likely to be a particularly important source of iron during invasive disease since it is abundant in human serum; the transferrin-binding proteins TbpA and TbpB comprise the meningococcal transferrin receptor and are induced in the outer-membrane under iron restricted conditions (Ala'Aldeen 1996).

Although all body sites inhabited by meningococci have only trace levels of haeme, two independent meningococcal haeme acquisition systems have been identified. HmbR has specificity for haemoglobin and HpuA/B binds haemoglobin and haptoglobin/haemoglobin complexes (Stojiljkovic *et al.* 1996, Lewis *et al.* 1997). It is likely that the haemoglobin receptors are of greater importance during invasive disease than in asymptomatic colonization (Perkins-Balding *et al.* 2004). Like Lbp and Tbp, Hpu is composed of a Ton-B dependent gated porin and an accessory lipoprotein subunit that may enhance the specificity of ligand binding (Noinaj *et al.* 2012).

Ferric enterobactin transport protein A, FetA, is also expressed under iron-limiting conditions, is involved in iron acquisition and has sequence similarity to other Ton-B dependent outer-membrane transporters (Kortekaas *et al.* 2007). The crystal structure of FetA is a 22-strand β-barrel structure with 11 surface-exposed loops (Kortekaas *et al.* 2007, Saleem *et al.* 2013) although 13 loops were predicted in a protein model (Pettersson *et al.* 1995). Given their role in pathogenicity, the surface receptors that mediate iron acquisition are attractive as potential vaccine components. The FetA iron-binding site is distinct from the surface exposed antigenic region (loop 7 in the protein model and loop 5 of the crystal structure) that elicits an immune response in vaccinees (Wedege *et al.* 1998).

Surface structure modulation

The meningococcus has numerous mechanisms for avoidance of the human immune system, including modulation of surface structures through antigenic- or phase-variation (Kremastinou *et al.* 1999). This renders many surface components unsuitable as vaccine candidates. Antigenic-variation takes place via recombination in one of two ways: (i) since *N. meningitidis* is naturally competent, imported DNA can provide the material for allelic exchange and (ii) possession of multiple gene copies permits intragenomic recombination and concomitant alteration of surface structure variation (van der Woude and Baumler 2004). For example, sequence variation generated through

intra-genomic recombination among *pil* genes alters pilin primary structure (Segal *et al.* 1986) and pilins are not considered suitable vaccine candidates for this reason.

Phase-variable gene expression is achieved via several mechanisms, including slipped strand mispairing (SSM). SSM in tracts of repeating DNA sequence, whether in open reading frames or upstream of genes, leads to insertions or deletions that alter transcriptional efficiency or result in incompletely translated proteins. Phase-variation of the methyltransferase components of Type III restriction-modification systems, the subject of Chapter 4 of this thesis, occurs through SSM of N-terminal region repeats. At the cell surface, phase-variable expression of multiple *opa* genes results in expression of different *opa* repertoires. Although phase-variable expression of surface exposed components mediates escape from bactericidal antibody killing *in vitro* (Tauseef *et al.* 2013), the Opa proteins are considered possible vaccine candidates due to their abundance and functional conservation (Callaghan *et al.* 2011).

Further examples include: phase-variation of PorA due to a homopolymeric tract in its promoter region (van der Ende *et al.* 1995); modulation of LPS through phase-variation of enzymes involved in biosynthesis and through post-translational modifications (Bayliss *et al.* 2008, Hill *et al.* 2010); and phase-variable expression of *nadA* due to an internal homopolymeric tract (Bambini *et al.* 2014). 'Capsule switching' can occur through horizontal gene transfer (HGT) of certain capsule biosynthesis genes and reversible alteration of capsule expression is mediated by phase-variation or through insertion of the mobile element IS1301 to capsule biosynthesis genes (Hammerschmidt *et al.* 1996a, Hammerschmidt *et al.* 1996b, Swartley *et al.* 1997). Capsules are highly successful vaccine components since they are expressed by the majority of meningococci *in vivo*.

Invasion

Rarely, infected hosts develop invasive disease. The exact molecular influencers and sequence of events from harmless colonization to invasive disease remain incompletely understood (Virji 2009, Hill *et al.* 2010). Following colonization, interactions among meningococcal outer-membrane

structures and host receptors can result in internalization of meningococci by epithelial cells (Hill *et al.* 2010); since *N. meningitidis* has been identified in sub-epithelial sites in healthy individuals this may allow avoidance of immune mechanisms (Sim *et al.* 2000). The capsule is down-regulated in the nasopharynx and interactions among Opc proteins and epithelial and endothelial cells may result in internalization of unencapsulated meningococci. However, disseminated infection is usually caused by encapsulated meningococci and it is possible that in circumstances such as prior viral infection encapsulated meningococci penetrate the epithelium and enter the blood due to the actions of Opa and pili (Rowe *et al.* 2007). Some minor adhesins are also likely to support invasion of the mucosal barriers (Hill *et al.* 2010).

Haematogenous spread

In the majority of infected individuals, dissemination is prevented by the immune response, however, in susceptible hosts, meningococci may survive and spread via the vasculature. Capillaries close to mucosal epithelial cells are a possible point of entry to the blood, and it has been suggested that PorA and PorB may be involved in bacterial uptake by endothelial cells via re-arrangement of the cytoskeleton (Nassif *et al.* 1999). Strong inflammatory responses and the complement and coagulation cascades are triggered by meningococci in the blood stream. For example, severe sepsis is caused by LPS induced proinflammatory responses, which lead to endothelial damage and capillary leakage, and ultimately organ failure.

Meningococcal survival in blood is effected through the anti-opsonic and anti-phagocytic characteristics of the capsule that facilitate resistance to complement-mediated lysis. Meningococci are also able to recruit negative regulators of complement in order to avoid lysis via mimicry. For example, factor H-binding protein (fHbp) is an outer-membrane lipoprotein composed of two beta barrels joined by a short linker that recruits human factor H (Madico *et al.* 2006); fHbp is a particularly effective vaccine antigen since it not only elicits serum antibodies that activate classical complement, but antibody-fHbp binding also reduces fH-fHbp binding, making meningococcal cells

more susceptible to bactericidal activity. fHbp is antigenically diverse, and sub-variants have been grouped into sub-families A and B, or into three variant groups (Masignani *et al.* 2003, Fletcher *et al.* 2004, Brehony *et al.* 2009). In addition, PorA is able to bind the complement regulator C4-binding protein (C4bp) and Opc may aid binding to vitronectin to inhibit insertion of the membrane attack complex to the meningococcal membrane (Jarva *et al.* 2005, Hill *et al.* 2010). Neisserial heparin binding antigen (NHBA) also increases serum resistance through the binding of heparin (Serruto *et al.* 2010) and is included in the recently licensed Bexsero® vaccine; NHBA peptides are grouped into a small number of variants.

Meningococci must cross the blood brain barrier in order to reach the central nervous system (Nassif *et al.* 2002). Pili play a major role in internalization by endothelial cells at the blood brain barrier, and it is likely that other adhesins such as Opc proteins are also involved (Unkmeir *et al.* 2002, Mairey *et al.* 2006). This results in the inflammation of the meninges known as meningitis.

Clinical disease and management

Disease presentation and treatment

In the UK, meningococcal disease is a notifiable disease. Clinical diagnosis can be complicated by symptoms that are similar to those of less serious infections and by different presentation in different age-groups (Pace and Pollard 2012). The disease is renowned for its rapid progression from onset to lethality. Meningitis occurs in around 50% of patients with meningococcal disease and symptoms include a sudden headache, fever, stiff neck, and photophobia (Rosenstein *et al.* 1999). Severe sepsis occurs in up to 20% of patients, and results in the characteristic rash; in some cases, sepsis will progress to purpura fulminans (Fig. 1), Waterhouse-Friderichsen syndrome (hypotension and adrenal haemorrhage) and organ failure (Rosenstein *et al.* 2001). More rarely, various syndromes of the respiratory tract (e.g. pneumonia, epiglottitis, otitis media), and focal infections (e.g. conjunctivitis and septic arthritis) occur. A patient may experience one or multiple symptoms.

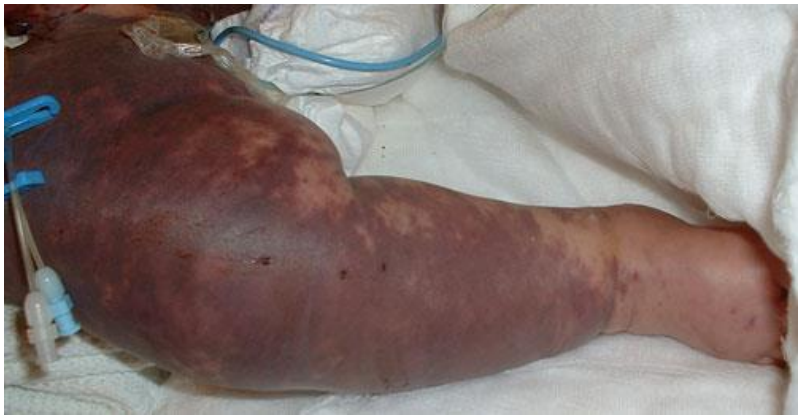


Figure 1. Purpura fulminans in meningococcal septicaemia. (Source: Meningitis Research Foundation)

Prior to the discovery and use of sulphonamide antimicrobials in the early twentieth century, the mortality rate of patients with meningococcal disease was around 70% (Rosenstein *et al.* 2001). Development of sulphonamide resistance, resulting from mutations to the gene encoding their target dihydropteroate synthase, *folP*, led to reduced usage from the late 1960s (Fiebelkorn *et al.* 2005). Fortunately, reports of resistance to commonly used antimicrobials are now relatively rare and β-lactam antibiotics such as penicillin are often administered. Reports of reduced susceptibility to penicillin G have increased slightly in the past few decades due to alterations in the structure of penicillin-binding protein 2 (PBP2), suggested to result from transformation events involving *penA*. Meningococcal resistance to rifampicin has also been reported: decreased susceptibility is mediated by mutation in *rpoB* that encodes the β subunit of RNA polymerase; however, the rarity of such isolates is suggested to result from the decreased evolutionary fitness of such meningococci (Taha *et al.* 2010). Finally, isolates with reduced susceptibility to ciprofloxacin, conferred by mutations in *gyrA* which encodes the fluoroquinolone target gyrase A, are occasionally found.

Case fatality rates are now around 5% in the UK and 10% in Europe, and are higher in patients with severe sepsis (Trotter *et al.* 2007, Ladhani *et al.* 2012a). Although prompt antibiotic administration is necessitated by the speed with which invasive disease progresses, early antibiotic administration has

not been definitively linked with an improved outcome (Rosenstein *et al.* 2001). Up to 20% of those who recover from meningococcal disease are left with long-term sequelae (Pace and Pollard 2012). These range from the relatively mild to severe; for example, around 7% of patients are left with neurological impairment, around 2% of those with meningitis are left with hearing loss, and 3% of those with septicaemia require amputation of, often multiple, digits and limbs (Edmond *et al.* 2010). Ongoing mental stress among the families of patients is also significant.

Case confirmation and reporting

Laboratory-confirmation of meningococcal disease preferably takes place through culture, Gram staining, and visualisation of meningococci via microscopy. Primarily, clinical specimens of cerebrospinal fluid or blood, which are sterile in uninfected individuals, are required for meningococcal isolation. However, culture sensitivity can be low if specimens are collected subsequent to antibiotic administration, and cases of meningococcal pneumonia are harder to confirm since it is not possible to distinguish carried from invasive meningococcal isolates in sputum specimens (Wylie *et al.* 1997). In the UK, many NHS hospital microbiology laboratories carry out culture-confirmation of meningococcal disease and the Public Health England Meningococcal Reference Unit (PHE-MRU) offers a free national service for species confirmation and meningococcal characterisation (Gray *et al.* 2006). These isolates comprise the MRF Meningococcus Genome Library, whose data forms the basis of much of this thesis. In addition to culture-confirmation, PCR testing was introduced in 1996 for identification of meningococcal DNA in specimens from which no meningococci can be isolated; this has the advantage of being quicker and more sensitive than culture techniques (Bryant *et al.* 2004, Gray *et al.* 2006). Today, just over half of all cases of meningococcal disease in England, Wales, and Northern Ireland are confirmed by PCR only (Heinsbroek *et al.* 2013). Since 2010, electronic reporting of meningococcal disease through LabBase2 by NHS hospitals has become a statutory requirement, and comparisons between the number of electronic reports and specimens received at the PHE-MRU indicate that case ascertainment is at 98% (Ladhani *et al.* 2012a, Heinsbroek *et al.* 2013).

Meningococcal characterisation

Enhanced epidemiological surveillance for detection of meningococcal outbreaks and vaccine design necessitates discrimination among meningococcal isolates, and, where culture is not possible, among meningococcal DNA samples: the differentiation of organisms of a species is referred to as 'typing'. In addition to its direct utility in public health, meningococcal typing contributes to theoretical population genetic and mathematical models of meningococcal populations, and choice of typing method is dependent on the practicalities of its implementation and on the level of discrimination required. Here, a very brief outline of the development of typing methodologies is provided with a focus on the sequence-based approaches that are applicable to typing using whole-genome sequence data.

Antigen typing

Original typing methods were serological approaches that distinguished meningococci on the basis of the capsule polysaccharide. From the 1950s a variety of agglutination techniques with CSF or serum and monoclonal or polyclonal antibodies were used to consolidate the systems developed at the start of the 20th century into the capsule serogroup system used today (Gordon 1918, Branham 1953, Slaterus 1961, Eldridge *et al.* 1978). Serological typing of subcapsular antigens developed through the 1970s and 1980s following identification of distinct antigen types within serogroups using bactericidal assays (Goldschneider *et al.* 1969). The serotype, serosubtype, and lipopolysaccharide typing schemes, based on PorB, PorA, and LPS respectively, permitted the first uniform nomenclature that could be used in epidemiological applications (Frasch and Chapman 1972, Abbott *et al.* 1985, Frasch *et al.* 1985, Cartwright *et al.* 1986), and antisera are still routinely used to characterise meningococcal isolates (Gray *et al.* 2006).

Sequence typing

DNA sequence typing permits a highly accurate, reproducible, and electronically portable approach to meningococcal characterisation and PCR without prior culture facilitates characterisation of non-

culture clinical specimens. Capsular genogrouping through PCR and sequencing became possible once the molecular structures of the capsule polysaccharides responsible for serogroup specificity were determined. Genes involved in polysaccharide biosynthesis and cell surface translocation are clustered in the *cps* locus, which is divided into six regions; for example, region A encodes the enzymes responsible for biosynthesis (Harrison *et al.* 2013b). The genogrouping target depends on the serogroup: the *siaD* gene may be used for meningococci of serogroups B, C, W, and Y which have sialylated polysaccharide capsules; the *mynA* gene may be used for serogroup A meningococci; and the capsule null locus (*cnl*) may be used for meningococci that lack the capsule biosynthesis operon (Jolley *et al.* 2006). Twelve capsule types are currently recognized: A, B, C, E, H, I, K, L, W, X, Y, Z (serogroup D was found to be an unencapsulated genogroup C variant) (Harrison *et al.* 2013b) and six serogroups (A, B, C, W, X, Y) cause the majority of invasive disease.

The first *porA* gene sequence was determined in 1989 (Barlow *et al.* 1989). Subsequent adoption of PorA VR1 and VR2 sequencing permitted a database of VR sequence diversity to be established for research and public health (Russell *et al.* 2004): these sequences are stored at http://pubmlst.org/neisseria where they are grouped into a number of families, along with the database of whole *porB* sequences. FetA has also been employed as an epidemiological marker since the adoption of DNA sequencing typing, and FetA VR1 sequences (loops 5 or 7 depending on the structural model) are also recorded in the PubMLST database (Urwin *et al.* 2004).

Multilocus typing

MLEE

Many meningococcal isolates were not serogroupable or serotypeable using immunological methods, in part due to the speed with which new antigenic variants emerged compared to the rate with which new reagents could be developed. The level of antigenic-variation generated through recombination also limited studies of meningococcal population structure. Multilocus enzyme electrophoresis (MLEE) was developed for classification of bacteria by the electrophoretic mobility

of their metabolic enzymes (Caugant *et al.* 1986a, Selander *et al.* 1986). The genes encoding these metabolic enzymes ('housekeeping loci') were more likely to be neutrally evolving than those encoding surface exposed structures, and, therefore, MLEE permitted the genetic relationships among meningococcal isolates to be investigated. Novel combinations of allelic variants are referred to as electrophoretic types (ETs). Clustering of ETs into clonal complexes first demonstrated meningococcal lineage structuring (Caugant *et al.* 1986b, Caugant *et al.* 1987a), and application of MLEE to global meningococcal epidemiology revealed the power of multilocus techniques by identifying the major clones responsible for serogroup A pandemics and clonal complexes associated with serogroup B and C disease in Europe (Caugant *et al.* 1986b, Crowe *et al.* 1987, Olyhoek *et al.* 1987).

MLST

Multilocus sequence typing (MLST) builds on the principles of MLEE with DNA sequencing and is a highly portable, accurate, and reproducible means by which to characterise a wide range of bacterial species (Maiden *et al.* 1998, Maiden 2006). The MLST scheme for the *Neisseria* is based on the fragments of seven housekeeping loci: new sequence variants of each fragment are submitted to the PubMLST database where they are assigned unique allele numbers regardless of the number of substitutions, and new allele profiles are assigned a sequence type (ST). As with ETs, STs can be clustered into clonal complexes (cc) based on a central genotype.

A strain type nomenclature is commonly used which incorporates the capsular group, PorA type, FetA type, MLST sequence type, and MLST clonal complex of an isolate; for example: B: P1.19,15: F5-1: ST-33 (cc32) (Jolley *et al.* 2007). Although MLST is now the primary method for the determination of meningococcal lineages, the ET is often included in lineage descriptions for historical reasons; for example, cc11 may be referred to as ST-11 complex/ET-37 complex.

Characterisation in UK national surveillance

Although MLST is part of multiple public health applications, such as outbreak delineation and formation of national vaccine implementation strategies, it is not routinely carried out at PHE-MRU as part of national surveillance (Feavers *et al.* 1999, Gray *et al.* 2006). This is due to financial and time constraints and because such a level of characterisation is not essential to diagnosis and treatment. Instead, meningococcal isolates undergo a series of carbohydrate utilization tests, serogrouping via agglutination (serogroups A, B, C, X, Y, Z, E, W) and dot-blot ELISA (serogroups A, B, C), and serotyping and serosubtyping via dot-blot ELISA (Gray *et al.* 2006). Non-culture genogrouping takes place via PCR of genes of the capsule biosynthesis operon from CSF, whole blood, serum, plasma, and joint fluids (Gray *et al.* 2006). Characterisation of the outer-membrane proteins is of great use in providing additional resolution in outbreak investigations (Feavers *et al.* 1999).

Epidemiology of meningococcal disease

Disease incidence

Meningococcal disease is characterised by substantial geographical and temporal incidence variation (Harrison *et al.* 2009a). The unique epidemiology of meningococcal disease in the 'African Meningitis Belt', so-called for cyclical epidemics that occur in sub-Saharan countries between Senegal and Ethiopia (Lapeyssonnie 1963), contrasts with endemic patterns common elsewhere. Incidence increases in Belt countries occur as major periodic epidemics every 5-12 years and annual seasonal outbreaks; the largest epidemic occurred in 1996-1997 with over 150,000 cases (Teyssou and Muros-Le Rouzic 2007). Endemic incidence is between one and three cases per 100,000 population and may be punctuated by clusters of disease or by hyperendemic periods of more frequent outbreaks. In developed countries, notifications from the early 20th century indicate incidence peaks at the time of the World Wars, for example there were more than 12,000 reported cases in England and Wales in 1940 (Jones 1988), but there have been few epidemics in the post-war era. Although a hyperendemic period occurred in England and Wales in the 1970s, this was not characterised by

multiple epidemics (Abbott *et al.* 1985, Kriz *et al.* 2011). An unusual epidemic, characterised by frequent and regular serogroup C cases, occurred in the Czech Republic during 1993; this is the focus of Chapter 3 of this thesis. In recent years, incidence in North America and Western Europe has been particularly low, with annual rates of two cases per 100,000 population in England and Wales in 2010/11 and 0.28/100,000 in the USA in 2009 (Ladhani *et al.* 2012a, Baccarini *et al.* 2013).

Global disease-associated lineages

The introduction of routine serogrouping in the 1970s and retrospective genotypic analysis of isolate collections pre-dating this time indicate that certain meningococcal groups are responsible for the majority of disease and that lineages associated with dynamic incidence rates have spread among geographic regions. For example, serogroup A meningococci were responsible for the military-associated outbreaks during both World Wars and one serogroup A clone caused two discrete global pandemics in the second half of the 20th century (Abbott *et al.* 1985, Achtman 1994). Although the serogroups responsible for the majority of disease (serogroups A, B, C, W, Y) have spread globally, their current distribution can be broadly categorized by region: for example, serogroups B and C are predominant in western Europe, and although serogroup A is more commonly identified in Africa and Asia than elsewhere, currently serogroup W predominates in the African Meningitis Belt (Oviedo-Orta *et al.* 2015).

The meningococci in possession of capsules corresponding to these major disease-causing serogroups are associated with a limited number of lineages. Characterisation of epidemiological collections, mostly from Europe, indicated that carried meningococci belong to a more diverse set of lineages than those causing invasive disease (Yazdankhah *et al.* 2004). At the time of writing there were 45 *N. meningitidis* clonal complexes (ccs) defined in the PubMLST database; of these, 12 (cc1, cc4, cc5, cc8, cc11, cc18, cc32, cc169, cc174, cc269, cc334, cc461) are recognized to be hyperinvasive, with relatively greater rates of isolation from disease than from carriage (Yazdankhah *et al.* 2004, Caugant 2008). In addition, cc41/44 and cc23 are responsible for a great deal of

meningococcal disease and have been included as hyperinvasive lineages in some studies (Bille *et al.* 2005, Caugant 2008, Watkins and Maiden 2012).

Knowledge of contemporary distributions of meningococcal serogroups and lineages is essential for vaccine design and deployment, as discussed in Chapter 5. Today, our knowledge of the geographical distribution of meningococcal lineages and their attack rates remains incomplete due to variability in the sensitivity and representiveness of surveillance systems; for example, many resource poor countries rely on clinical diagnosis and surveillance may be passive and from a subset of national hospitals (Harrison *et al.* 2009a). In addition, developing countries continue to have greater mortality and morbidity rates associated with meningococcal disease.

North and South America

Serogroup Y disease currently predominates in the USA, followed by serogroup C, then serogroup B, disease; the relative proportions vary by state and have altered significantly since the mid-1990s when incidence increased and ST-23cc (cc23) serogroup Y became more prevalent (Harrison *et al.* 2009a, Krauland *et al.* 2012, Baccarini *et al.* 2013) (Fig. 2). Serogroup W meningococci now cause little disease (Baccarini *et al.* 2013). Canada has seen broadly similar incidence trends, although outbreaks of cc11 disease from the early 2000s led to a predominance of serogroup C at that time; currently much serogroup B disease is caused by cc269 (Tyrrell *et al.* 2002, Gilca *et al.* 2012). Serogroup C conjugate vaccines, not licensed in the USA, are likely to be responsible at least in part for the different epidemiology between the USA and Canada (Harrison *et al.* 2009a).

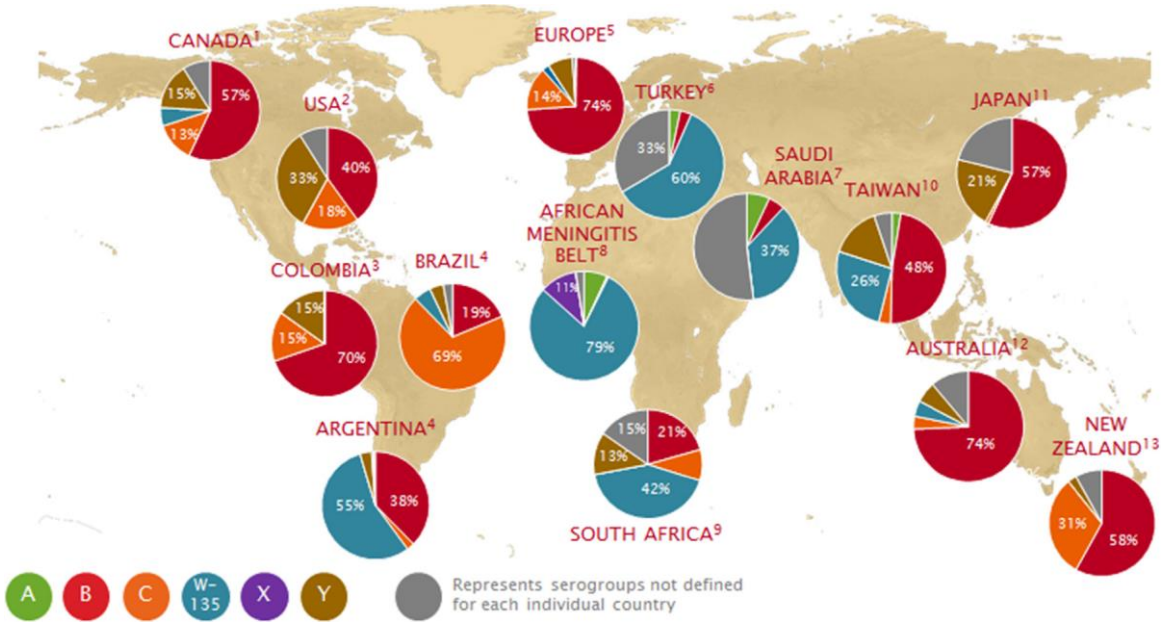


Figure 2. Global distribution of meningococcal serogroups in meningococcal disease. Source: Oviedo-Orta, E., S. Ahmed, R. Rappuoli and S. Black (2015)"Prevention and control of meningococcal outbreaks: The emerging role of serogroup B meningococcal vaccines" Vaccine. (Oviedo-Orta *et al.* 2015)

Surveillance quality is varied among Latin American countries (Safadi *et al.* 2013) (Fig. 2). In general, meningococcal disease is endemic and serogroup B is predominant (Harrison *et al.* 2009a). Serogroup A disease is rare and the 21st century has seen the emergence of cc11 serogroup W and cc23 serogroup Y disease; Brazil and Argentina have experienced little serogroup Y disease compared to serogroup W in recent years in contrast with other countries (Fig. 2) (Safadi *et al.* 2013).

Africa

The majority of meningococcal disease in the African Meningitis Belt has resulted from serogroup A meningococci since at least 1940. Successive epidemics have been caused by cc1, cc4, and cc5, and in the past 20 years there has been a replacement of disease caused by the central genotype of cc5, ST-5, by ST-7 meningococci of the same clonal complex (Caugant and Nicolas 2007). In some countries, a third cc5 ST, ST-2859, has replaced ST-7 as the primary cause of serogroup A disease. In

recent years disease caused by cc11 serogroup W and ST-181 serogroup X have been responsible for outbreaks and serogroup W now predominates in many countries; there have been few recent cases of serogroup A except in Chad (Halperin *et al.* 2012, Xie *et al.* 2013) (Fig. 2). Disease data are sparse for countries outside of the Belt, although the epidemiology in South Africa shares features of both the Belt and European countries with endemic disease (Harrison *et al.* 2009a). In North Africa and the Middle East there is little available information, although the epidemiology in these locations is likely to be similar (Halperin *et al.* 2012).

Asia

Available data from Asia indicate that endemic incidence rates are low, although anecdotal evidence suggests that the burden of disease, particularly in developing countries, may be under-reported (Harrison *et al.* 2009a, Vyse *et al.* 2011). Multiple serogroup A outbreaks have occurred in India, Pakistan, Mongolia, Nepal, and Indonesia, with a serogroup C epidemic in Vietnam, in the past 40 years, but relatively little is known about endemic disease (Vyse *et al.* 2011). Currently, the majority of disease in China is serogroup A following repeated serogroup A epidemics from 1950-1980, although there has been a recent emergence of cc4821 serogroup C disease. In contrast, the majority of meningococcal disease in Japan results from serogroup B, then serogroup Y, organisms (Fig. 2).

Australia and New Zealand

Australia and New Zealand have experienced rather different meningococcal disease epidemiology from one another in recent years: although current relative serogroup proportions are similar to those in other developed countries, the majority of serogroup B disease is caused by different strains in either country (Fig. 2). New Zealand experienced a prolonged, clonal, cc41/44 serogroup B epidemic from the 1990s that has diminished since introduction of the MeNZB vaccine in 2004 (Harrison *et al.* 2009a).

Europe

Surveillance data from Europe are relatively complete, having been collected by several surveillance programs (EU-IBIS, 1999-2006; EU-MenNet, 2001-2005; European Centre for Disease Prevention and Control (ECDC), 2007-present), although less information on the epidemiology of meningococcal disease is available from Eastern European countries. Recent incidence has been low, at 0.68/100,000 in 2012, following a decline from 1.9/100,000 in 1999 (Harrison *et al.* 2009a, European Centre for Disease Prevention and Control 2014). In part this is due to reduction of cc11 serogroup C disease following national immunization programs, however, serogroup B disease has also declined in countries such as England and Wales, The Netherlands, and Denmark in recent years for reasons incompletely known. Incidence is also variable among countries with little obvious geographical pattern. For example, in 2004 incidence was as high as 4.92/100,000 in the Republic of Ireland and 3.94/100,000 in Malta, whereas it was only 0.55/100,000 in Italy (Trotter *et al.* 2007). Serogroup Y disease has increased across Europe in recent years (Fig. 2). The prevalence of serogroup Y was significant in Sweden and the Czech Republic by the late 1990s and in 2010 Scandinavian countries had the highest relative proportion of serogroup Y (Broker *et al.* 2012b). There are few cases of serogroup A disease in Europe, although Russia, Greece, and Romania reported cases at the start of this century (Kriz *et al.* 2011).

Major hyperinvasive lineages

A limited number of lineages are associated with the prevalent disease-causing meningococcal serogroups. In the following description, the focus is on those currently of greatest concern in England and Wales and therefore relevant to the MRF Meningococcus Genome Library: in order of decreasing prevalence, these are cc41/44 (serogroup B), cc269 (serogroup B), cc23 (serogroup Y), and cc11 (serogroups W, B, C).

Serogroup A-associated lineages

Meningococci in possession of capsules corresponding to serogroup A are exclusively members of cc1, cc4, and cc5, although additional STs have been identified that are not designated to a known cc. The population structure of serogroup A meningococci was first described by Achtman and colleagues through application of MLEE to multiple isolate collections and cc1 and cc5 have been associated with epidemic disease or global pandemic spread (Olyhoek *et al.* 1987, Wang *et al.* 1992, Achtman 1994). cc1 (formerly subgroup I) meningococci were responsible for USA and UK outbreaks during the Second World War. Since then, these have caused epidemics in North Africa and the African Meningitis Belt (1960s), and outbreaks in China, Holland, Brazil, Australasia, USA (1970s), African countries south of the Belt, and South Africa (1990s) (Olyhoek *et al.* 1987, Wang *et al.* 1992, Achtman 1997, Caugant and Nicolas 2007). cc4 (formerly subgroup IV) meningococci replaced cc1 as the cause of epidemics in West Africa in the 1980s. Meningococci of cc5 (formerly subgroup III) were responsible for two global pandemics that each arose in China: the first, in the 1960s and 1970s, resulted in epidemics in China, Moscow, Norway, Finland, and Brazil; the second, in the 1970s and 1980s, resulted in epidemics in China and Nepal and among Hajj pilgrims who subsequently introduced cc5 to Africa resulting in major epidemics (Olyhoek *et al.* 1987, Moore *et al.* 1989, Wang *et al.* 1992). This lineage was responsible for the major African Meningitis Belt epidemic in 1996.

Serogroup B-associated lineages

Serogroup B meningococci are more genetically diverse than serogroup A meningococci and their epidemics tend to arise more gradually and last longer (Caugant *et al.* 1986b). However, several major lineages are predominantly responsible for serogroup B disease and cc41/44, cc269, and cc32 cause the majority of group B cases in England and Wales. cc32 (formerly ET-5) was first identified in the late 1960s prior to the beginning of an extended hyperendemic period caused by cc32 meningococci (strain type B:15:P1.7,16) in Norway from the mid-1970s to the late-1980s (Caugant *et al.* 1987a). This antigenic type caused increased disease in Oregon, USA, from 1993 to 2007 and an extended outbreak in Normandy, France, over the past ten years (Diermayer *et al.* 1999, Rouaud *et*

al. 2006). Other European countries experienced heightened serogroup B incidence during the 1980s; in England and Wales, this corresponded to increased prevalence of serotype 15 meningococci (Abbott *et al.* 1985). Antigenically distinct cc32 meningococci (B:4:P1.19,15) caused increased cases in Spain in the mid-1970s, and later an epidemic in Cuba in the 1980s and other parts of Latin America. Very recently, cc32 caused an outbreak at UC Santa Barbara (Whelan *et al.* 2015) and findings based on genome-wide variation suggest that distinct cc32 sub-lineages, associated with the different antigenic types, caused these localized outbreaks following expansion of the lineage (Harrison *et al.* 2015).

cc41/44 (formerly lineage III) was first identified in the Netherlands in the 1980s (Scholten *et al.* 1994b). It is the most diverse clonal complex with two central genotypes and over 1000 assigned STs, and is primarily associated with the serogroup B capsule. cc41/44 meningococci cause disease worldwide and have been responsible for multiple outbreaks, notably the extended New Zealand epidemic between 1991 and 2007 (B:4:P1.7-2,4). In 2013, cc41/44 caused an outbreak at Princeton University (Oviedo-Orta *et al.* 2015).

cc269 is thought to have diverged relatively recently. The earliest known isolates are from The Netherlands in 1970, and cc269 serogroup B meningococci emerged as an increased cause of meningococcal disease in Quebec, Canada, in the mid-2000s; it is postulated that selective pressure following introduction of A, C, W, Y tetravalent vaccines resulted in this increase (Law *et al.* 2006). cc269 is primarily associated with serogroup B, although cc269 meningococci that express capsules corresponding to serogroup C have been identified that may have undergone B to C capsule switching (Law *et al.* 2006). This lineage was responsible for a community outbreak in France in 2008/9, and a university outbreak in Ohio, USA, between 2008 and 2010 (Delisle *et al.* 2010, Mandal *et al.* 2013).

ST-11 clonal complex

cc11 (formerly ET-37 complex) is unusual in that it is associated with serogroups C, B, and W. This lineage was circulating as early as the First World War, and caused US army outbreaks in the 1960s, a serogroup C outbreak in Brazil in the early 1970s, a serogroup B epidemic in South Africa in the late 1970s, and serogroup B cases in China in 1974 (Caugant 1998). Serogroup C cc11 meningococci became increasingly prevalent in European countries in the 1980s and 1990s, and in the late 1980s a new variant of the ET-37 complex, ET-15, was identified to be responsible for increased serogroup C cases in Canada and subsequent global outbreaks. The increase in serogroup C cases in many countries in the mid-1990s resulted in implementation of meningococcal C conjugate (MCC) vaccines, and cc11 is now less prevalent in Europe. However, there have recently been clusters of serogroup C ET-15 disease among men who have sex with men (MSM) in Europe and the USA (Weiss and Varma 2013).

Outbreaks among Hajj pilgrims and their contacts in 2000 were the result of serogroup W cc11 meningococci (Taha *et al.* 2000) and the pilgrimage is likely to have accelerated the expansion of these organisms in Africa, resulting in a major epidemic in Burkina Faso in 2002 and outbreaks in many countries from 2002 onwards (Decosas and Koama 2002, Mayer *et al.* 2002, von Gottberg *et al.* 2008). Recent increases in England and Wales have resulted in addition of a quadrivalent vaccine to teenage vaccine schedules from August 2015.

ST-23 clonal complex

The predominant serogroup Y-associated lineage is cc23. Although now responsible for a third of disease in the USA following increased prevalence from the 1990s, cc23 was rare in Europe until the late 2000s when it became more prominent in central and western Europe and in Scandinavian countries in particular (Broker *et al.* 2012b, Halperin *et al.* 2012, Ladhani *et al.* 2012b). The proportions of serogroup Y disease remained high in Sweden in 2013 (Broker *et al.* 2015). Genome-wide genetic characterisation of isolates from the USA and Sweden indicate that increased cc23

disease has not been caused by a single virulent clone but has been due to fluctuations in the prevalence of antigenically distinct strains. In the case of Maryland, USA, it is postulated that the strain type initially responsible for increased disease was replaced with a second that escaped population immunity (Krauland *et al.* 2012), and in the case of Sweden, it was shown that cc23 sub-lineages defined at the whole-genome level were also circulating in England and Wales (Toros *et al.* 2015).

Meningococcal populations

Genetic diversity

A great deal of phenotypic and genetic diversity is observed among meningococci even in cross-sectional samples of one geographical area. For example, invasive disease isolates in Germany had a Simpson's Index of 0.963, where a value of 1 indicates complete diversity, even at the level of two antigens (Elias *et al.* 2006). Genetic variation in *N. meningitidis* is thought to be generated at a greater rate by recombination than by mutation: recombination rates generating new alleles among housekeeping loci have been estimated at 10 times those of mutation (Jolley *et al.* 2005) and at 3.6 times those of mutation (Feil *et al.* 1999).

Bacterial recombination occurs through conjugation, transduction of phage DNA, and, particularly in *N. meningitidis*, natural competence allows uptake of environmental DNA through transformation. The majority of HGT events in meningococci occur via homologous recombination between highly related cells, making variation generated by this process difficult to distinguish from mutational substitutions. Introduction of novel genome content, often by transduction, is a rarer occurrence.

Transformation among meningococci and other species is partly regulated by DNA uptake sequences (DUS), with a direct positive correlation between DUS sequence similarity and transformation rates (Frye *et al.* 2013). DUS are over-represented in essential genes of the core-genome relative to dispensable genes (Davidsen *et al.* 2004); it is proposed that this arrangement provides a mechanism

through which recombination acts as a regenerative, rather than diversifying, process (Treangen *et al.* 2008). The rates and selective advantages of recombination in bacteria have been subject to a great deal of debate. For example, it has been suggested that: (i) exogenous DNA is a nutritional resource; (ii) recombination allows faster generation of diversity and adaptation than does mutation, such as under frequency-dependent selection where recombination may allow pathogens to stay ahead of hosts (Red Queen hypothesis) through avoiding clonal interference; (iii) recombination prevents accumulation of deleterious mutations in clonal organisms (Muller's Ratchet); (iv) and acquired single-stranded DNA may function in DNA repair without the need for endonucleases. Recombination may also, however, break down linkage disequilibrium among beneficial alleles (Vos 2009).

It is clear that the balance between a parasexual lifestyle and clonality differs among bacterial species, with multilocus allele analysis demonstrating a range from completely clonal (e.g. *Salmonella*) to panmictic (e.g. *N. gonorrhoeae*) (Smith *et al.* 1993). Recombination in meningococcal populations was inferred from incongruent gene trees that indicated a lack of comparable vertical ancestry (Feil *et al.* 1996, Zhou *et al.* 1997) and as such meningococcal populations are likely to have a network structure (Holmes *et al.* 1999). In addition, homologous DNA shared among meningococci and other bacteria showed that interspecies recombination during co-colonization of the nasopharynx contributes to meningococcal evolution (Zhou and Spratt 1992, Zhou *et al.* 1997) and mobile genetic elements comprise a large proportion of the both the core- and dispensable-genomes (Hotopp *et al.* 2006). The gene pool from which diversity is generated, the pan-genome, is therefore large and dynamic: the species core-genome (DNA shared by all meningococci) has been estimated at 40% of the pan-genome (Schoen *et al.* 2009).

Population structure

Despite high rates of recombination, there is clear evidence of structure within meningococcal populations. Isolates can be grouped into clonal complexes that comprise particularly related

organisms; using neutrally evolving genes (MLEE/MLST loci), isolates are clustered based on a threshold of genetic distance from central genotypes (often the most prevalent and long-lived sequence type) with algorithms like eBURST (Feil *et al.* 2004). There is also evidence for structuring of antigenic diversity, with a limited number of the possible variant combinations (antigen types) frequently identified (Urwin *et al.* 2004, Russell *et al.* 2008). Although not absolutely concordant, antigen types are stably associated with meningococcal lineages in samples spanning multiple decades and geographic regions (Caugant *et al.* 1986b, Caugant *et al.* 1987a, Watkins and Maiden 2012); this is of great practical utility in the design of vaccines.

The existence of linkage among housekeeping loci could be inferred as rates of recombination too low to obscure patterns of clonal evolution. However, these patterns can arise alongside high rates of recombination with the existence of barriers to gene exchange, created, for example, by geographical isolation or mechanistic limitations to transformation. The strain-specific restriction-modification systems examined in Chapter 4 provide one such mechanism for limiting recombination: the incorporation of DNA is prevented by restriction endonucleases that recognize externally acquired DNA as foreign when it does not possess the modification signatures of their partner methyltransferase (Claus *et al.* 2000). Alternatively, transient linkage disequilibrium could arise from the local expansion of particular strain types that are subsequently broken up by recombination, and it has been suggested that *N. meningitidis* exhibits a microepidemic population structure despite having a sexual lifestyle (Smith *et al.* 1993). In addition, the neutral microepidemic model suggests that neutral genetic drift, combined with local transmission in heterogeneously distributed host populations, could create observed patterns of linkage disequilibrium and the temporary emergence of disease-associated lineages (Fraser *et al.* 2005).

However, meningococcal diversity is also structured by phenotype. Not only is meningococcal diversity greater in carriage than disease isolate collections, but lineages are also differentially distributed with hyperinvasive lineages isolated at greater rates from disease than from carriage

(Yazdankhah *et al.* 2004). The persistence of hyperinvasive meningococci presents a paradox, since invasive disease prevents onward transmission and eliminates reproductive fitness, yet these organisms are successful on global scales. Some lineages have distinct epidemiological behaviours, with outbreaks caused by serogroup C and cc11 organisms frequently responsible for short, sudden outbreaks, and epidemics caused by serogroup A meningococci in sub-Saharan Africa; these differences are likely to result from altered transmission characteristics, which forms the basis of explanations for observed strain structure and strain dynamics.

Strain theory

The majority of models invoke natural selection on meningococcal transmission fitness as the primary driver of the patterns observed in natural meningococcal populations. For example, the localized accumulation of variation during the transmission of meningococci may create 'genoclouds' of related meningococcal variants that do not completely overlap genetically with those elsewhere (Achtman *et al.* 2001, Zhu *et al.* 2001). Over time diversity may be purged by negative selection on unfit variants that are, for example, less transmissible due to variation in the capsule or due to rising herd immunity, or, by bottlenecks when subsets of variants are introduced to new locations. Fit variants may persist: for example, it has been suggested that the extended duration of B:P1.7-2, 4:F1-5:cc41/44 meningococci in New Zealand results from low immune responses to this serosubtype (Vogel *et al.* 2010).

Immune selection theory provides an explanation for the selective maintenance of a few lineages through co-evolution with host populations (Gupta *et al.* 1996). Antigen variants unique to lineages do not incur immune responses primed by previously carried lineages, and mathematical models demonstrate non-overlapping variants of antigens associated with dominant lineages (Gupta *et al.* 1996). Predictions from immune selection closely fit patterns of porin and Opa protein variation, and epistasis among particular antigen variants may lead to linkage disequilibrium of variants that are

stably associated with lineages (Callaghan *et al.* 2008, Buckee *et al.* 2010, Buckee *et al.* 2011, Watkins and Maiden 2012).

It is plausible that housekeeping loci hitchhike with positively selected antigen alleles, however, antigenic types are only markers for lineages and distinct lineages may share variants. A model that allowed sequence types to possess slight differences in transmission fitness showed that inter-strain competition in the nasopharynx could result in the existence of only a few dominant lineages; as immune selective pressure increased, antigen types went from a stable association with these lineages, to oscillating over time. This latter scenario was observed among meningococci isolated from the Czech Republic over almost three decades (Buckee *et al.* 2008).

Invasiveness

Global and temporal fluctuations in disease incidence can, in part, be explained by characteristics of meningococcal transmission. *N. meningitidis* has been labeled an 'accidental pathogen' due to the evolutionary dead-end of invasive disease (Moxon and Jansen 2005): transmission occurs via respiratory droplets among asymptomatic carriers of the bacterium. Indeed, most meningococcal disease patients have not had contact with other patients. Carriage studies are therefore invaluable in understanding the biology of *N. meningitidis*, the epidemiology of disease, and the population impact of vaccination.

Carriage studies

The majority of meningococcal isolates have been sampled from cases of disease, and the majority of isolates from asymptomatic individuals have been collected in defined projects over limited time periods as part of cross-sectional studies. The Czech carriage collection, composed of isolates spanning thirty years (Buckee *et al.* 2008), and the UK carriage collection, sampled at the same time as the serogroup C conjugate vaccine initiative (Maiden *et al.* 2008), are therefore invaluable.

The majority of carriage studies have focused on at-risk groups, such as adolescents (MacLennan *et al.* 2006), the military (Blackwell *et al.* 1992), and close contacts of cases, however, there have been some population-wide studies. These indicate that the point prevalence of meningococcal carriage is around 10% of the human population in non-epidemic situations (Caugant *et al.* 1994, Claus *et al.* 2005). The majority of individuals are likely to carry the bacterium at some point in their lives; longitudinal studies indicate that individual carriage periods may be transient or chronic, lasting from days to months (De Wals and Bouckaert 1985, Broome 1986, Gray *et al.* 2006, Bidmos *et al.* 2011). Although our understanding of within-host carriage dynamics has been limited by a lack of sensitive meningococcal characterisation in longitudinal carriage studies, it is likely that individuals harbor single strains during most carriage periods (Ala'Aldeen *et al.* 2000, Caugant *et al.* 2007, Glitza *et al.* 2008, Bidmos *et al.* 2011). Invasive disease is likely to occur within 10 days of acquisition of carriage, if at all (Edwards *et al.* 1977, Stephens 1999, Caugant and Maiden 2009).

Together with epidemiological surveillance, carriage studies have revealed multiple risk factors for meningococcal carriage. However, although carriage is a precursor of disease, the exact relationship between the two is uncertain; for example, fluctuations in disease incidence cannot necessarily be predicted by those of carriage (Caugant *et al.* 2007). Incidence fluctuations are, therefore, likely to result from a combination of population level immunity, various risk factors for exposure and acquisition of carriage, and the invasive potential of the meningococcal strain in question.

Risk factors

Climatic factors have consistently been implicated in meningococcal disease incidence: in North America and northern and western Europe, more cases occur during the first months of the year, and seasonal epidemics of disease occur in the Meningitis Belt concurrent with the dry and windy Harmattan. Carriage rates are largely unaffected by season in either case (Caugant *et al.* 1994, Trotter and Greenwood 2007). Meningococcal disease incidence may be affected by independent risk factors that display seasonal trends: for example, cases may increase following winter influenza

outbreaks (Cartwright *et al.* 1991). The host-meningococcal interactions underlying these associations are incompletely understood; however, increased invasion may occur during inflammation and concomitant temperature increases may aid immune evasion (Rowe *et al.* 2007, Loh *et al.* 2013). In addition, individuals with prior infection by other pathogens may have increased risk of carriage (Stephens 1999) and HIV-infected individuals may be at greater risk of meningococcal disease (Cohen *et al.* 2010).

Carriage rates are age-dependent, as are disease rates, but each exhibit different distributions. In endemic settings, carriage is lowest in infants and young children with peaks in adolescents and young adults (Claus *et al.* 2005), whilst disease incidence is greatest in infants with a secondary peak in adolescents. It is possible that *N. lactamica*, carried primarily by infants, provides protective immunity against meningococcal colonization, and it has been considered as a vaccine candidate (Bennett *et al.* 2005, Gorringe *et al.* 2009). Meningococcal carriage stimulates a systemic protective antibody response, and therefore active immunity against the meningococcus increases with age as additional carriage episodes occur (Kremastinou *et al.* 1999).

The relationship between age and meningococcal carriage may in part be due to behaviours characteristic of particular age-groups. Two of the most important risk factors for carriage are the number of contacts and their proximity, and adolescent social interactions may create contact networks that facilitate transmission (MacLennan *et al.* 2006). Infants, in contrast, are likely to be in close contact primarily with their parents (van Hoek *et al.* 2013). The extent of contact with carriers required for exposure is uncertain; carriage rates among household contacts of cases and among boarding school pupils have been variable (Olcen *et al.* 1981, Broome 1986, MacLennan *et al.* 2006), however, transmission and carriage rates in closed or semi-closed environments such as educational and military institutions are greater (Blackwell *et al.* 1992, Caugant *et al.* 1992, Neal *et al.* 2000). Meningococci may survive on glass and plastic surfaces for up to 72 hours (Swain and Martin 2007, Tzeng *et al.* 2013) and it is therefore possible that acquisition occurs in the sharing of kitchen

utensils. Kissing is a strong risk factor for carriage in teenagers, as is passive and active smoking (MacLennan *et al.* 2006). In addition, carriage prevalence has been found to be slightly higher in males and in those of low socioeconomic status in some studies (Kremastinou *et al.* 1994).

Meningococcal pathogenicity

Unlike some bacterial pathogens, there is a lack of definitive virulence factors possessed by disease-causing meningococci. The polysaccharide capsule has been identified as a virulence determinant, however, only certain capsule types are associated with invasive disease and capsule null meningococci have been isolated from patients. Epidemiological approaches have identified various meningococcal factors associated with disease or carriage, most notably the presence of genetically different meningococcal populations in carriage isolates versus disease isolates; hyperinvasive meningococci such as those belonging to cc11 and ST-41, are over-represented in disease (Yazdankhah *et al.* 2004). This over-representation results either from the increased virulence of these organisms, and/or due to a poorer colonization ability that results in decreased carriage durations balanced by high transmission rates (Jolley *et al.* 2000, Caugant *et al.* 2007).

There is some evidence for increased virulence among hyperinvasive lineages. cc11 in particular has been responsible for particularly severe symptoms and high mortality rates (Krizova and Musilek 1995, Whalen *et al.* 1995) and cc23 has been reported to be highly virulent in the Czech Republic (Kriz *et al.* 2011). Differences in the duration of carriage of particular lineages remain unresolved, however, some studies indicate that cc11 carriage is rapidly lost or progresses quickly to disease, whereas cc23 establishes stable carriage and is transmitted well (Edwards *et al.* 1977, Caugant *et al.* 2007). cc11 is known for causing sudden outbreaks that particularly affect adolescents and it has been suggested that the 'spreader' phenotype of lineages such as cc11 is mediated by an inability to establish stable biofilm-mediated colonization (Lappann *et al.* 2010).

Particular lineages have also been shown to be associated with disease in subgroups of the human population, a possible indication of niche differentiation. For example, in Europe and Canada,

disease associations such as cc11 with adults, cc23 with adolescents and the elderly, and cc41/44 with infants have been identified (Tyrrell *et al.* 2002, Brehony *et al.* 2014). Patients with disease caused by lineages associated with serogroups Y and W are more likely to experience pneumonia than meningitis or septicaemia (Ladhani *et al.* 2012b). cc11 serogroup W meningococci have also recently been associated with clusters of disease in men who have sex with men (MSM). The relative contributions of the behaviours characteristics of MSM and of adaptation of these organisms to transmission in the MSM niche are unclear; carriage is particularly high, particularly in urethral and rectal sites (Janda *et al.* 1980).

Epidemiologically-defined factors

The association of MLST-defined lineages with certain phenotypes suggests that the genetic basis of these phenotypes could be identified, whether this is sequence variation in the seven MLST loci themselves, allelic combinations among genomic loci linked to particular MLST profiles, or presence/absence of genome content associated with these MLST profiles. The extent of transformation in meningococcal populations may limit the ability of MLST to predict accessory-genome content, however, if rates of homologous recombination are high enough in the core-genome, clonal complexes are likely to be recently diverged lineages and therefore highly linked to the accessory-genome (Turner and Feil 2007).

Epidemiological studies employing data from whole-genome sequencing technologies have so far been unable to identify meningococcal factors that are consistently unique to either disease or carriage isolates, or to particular lineages. Indeed, certain genes involved in invasion of the mucosal barrier are present in commensal *Neisseria* species (Marri *et al.* 2010). However, certain factors have been identified that are associated with hyperinvasive lineages: presence of both of the iron acquisition systems HmbR and HpuA/B is associated with hyperinvasive lineages (Harrison *et al.* 2009b, Harrison *et al.* 2013a), as is a putative phage (the MDA phage).

The MDA phage is particularly associated with meningococci causing disease in adolescents, once the effect of hyperinvasive lineages is statistically accounted for (Bille *et al.* 2005, Bille *et al.* 2008). It has been suggested that the MDA phage may be responsible for increased probability of invasion through providing increased transmission efficiency among adolescents, which results in selective maintenance in lineages associated with disease in age-groups that encounter many new hosts (Moxon and Jansen 2005). Alternatively, it has been speculated that repeated gain and loss of the phage may result in oscillations in gene expression profiles that modulate invasion likelihood, appearing as short outbreaks of disease in adolescents (Schoen *et al.* 2008). Genome wide association studies in well defined isolate collections should provide further insights into differences in gene content between disease and carriage associated meningococci, however, animal models that mimic specific host sites will be necessary to profile phenotypic differences.

Models of pathogenicity

Given that invasive disease prevents onward transmission, meningococcal pathogenicity challenges longstanding views on both the evolution of virulence and on the relationship between transmission and virulence (Brown *et al.* 2012). For example, original theory postulated that evolution of pathogenic lifestyles was a tradeoff between increased transmission fitness through high within-host multiplication rates and the subsequent increases in host mortality (Lipsitch and Moxon 1997). Given that invasion offers zero reproductive fitness, only the selective forces acting on the carriage and transmission of meningococci are responsible for phenotypic differences among lineages. The aforementioned model by Buckee and colleagues (Buckee *et al.* 2008) indicated that slight differences in transmission ability provide hyperinvasive meningococcal lineages with the increased fitness necessary to compensate for their rates of invasion. However, this was true only when competition for hosts was high, a scenario plausible particularly with high acquisition rates created by adolescent behaviour. In addition, a stochastic mathematical model showed that differences in carriage duration and pathogenicity of hyperinvasive lineages determined the size and occurrence of outbreaks (Stollenwerk *et al.* 2004).

Both these models were based on lineages defined by MLST, which suggests that the seven housekeeping loci are selective-, rather than neutral-, markers. Six of these loci are metabolic enzymes, and it is becoming increasingly likely that transmission ability, and subsequent increases in invasion probability, are dependent on differences in metabolic efficiency among meningococcal lineages (Schoen *et al.* 2014). Metabolic phenotypes may be mediated by a variety of mechanisms, including variable genome content or within-host differences in the expression levels of metabolic genes (Joseph *et al.* 2010). The prevalence of contingency genes in meningococcal genomes provides one such mechanism for reversible alterations in gene expression, and Mod enzymes have been identified that mediate phase-variable expression of tens of genes (known as 'phasevarions') (Srikhanta *et al.* 2005). Indeed, it was recently suggested that hyperinvasive lineages differ not in their colonization ability, but in their ability to modulate the expression of genes, for example those involved in the oxidative stress response, that allow spread within the host (Schoen *et al.* 2014).

This idea is supported by mathematical models of 'short-sighted evolution'; these suggest that virulence is an inadvertent consequence of the contingency loci that are an adaptation for meningococcal survival in diverse human hosts (Levin and Bull 1994, Meyers *et al.* 2003). Within-host pressures may result in multiplication and dissemination of meningococci that have a local advantage from expression of particular genes, with no regard for onward transmission.

Vaccination

Given the unpredictability and speed with which invasion occurs, vaccination is crucial for controlling meningococcal disease. Despite much attention, prevention through immunization remains a challenge due to the lack of a universal vaccine; this partly results from the similarity of the serogroup B capsular antigen to human cell components, and from the diversity of group B organisms.

Polysaccharide vaccines

Vaccines based on the polysaccharide capsules associated with serogroups A, C, W, and Y have been available for many years. Monovalent vaccines were first developed in the 1970s and 1980s in response to outbreaks of serogroup C and A disease in the US military (Gotschlich *et al.* 1969a, Gotschlich *et al.* 1969b, Artenstein *et al.* 1970). Bivalent (A, C) and tetravalent (A, C, W, Y) polysaccharide vaccines are still successfully used to control outbreaks in African countries, to immunize pilgrims to the Hajj, and for at-risk individuals such as laboratory workers and travellers. However, whilst polysaccharide vaccines are safe and serum antibody against capsular polysaccharide activates complement mediated-bacteriolysis and/or opsonisation, in infants elicitation of antibodies is variable depending on the specific polysaccharide, and these vaccines fail to elicit immunological memory in individuals of any age (Makela *et al.* 1977, Reingold *et al.* 1985, Pace and Pollard 2007). Classically, polysaccharide vaccines were thought not to result in herd immunity through reduction of transmission, however, MenC polysaccharide has been shown to induce a mucosal response in adolescents (Pace and Pollard 2007).

Conjugate vaccines

Polysaccharide-conjugate vaccines, in which the immunogenicity of the polysaccharide is improved by chemical conjugation to a protein carrier such as tetanus toxoid, have largely superseded polysaccharide vaccines. Immunization results in a T-cell dependent immune response and memory B cell production. In addition, reduction of nasopharyngeal colonization (through stimulation of IgA and IgG) leads to protection of the unvaccinated population through herd immunity (Robbins *et al.* 1989, Pace and Pollard 2007).

Monovalent, bivalent, and tetravalent conjugate vaccines have been formulated against serogroup A, W, Y, and C polysaccharides. Particularly successful have been meningococcal serogroup C conjugate (MCC) vaccines. Their licensure was accelerated in the UK in the early 1990s in response to outbreaks of serogroup C cc11 disease, and in 1999 were incorporated into routine infant

immunization schedules; an additional catch-up campaign was instigated in children under 19 years of age. Subsequent surveillance and carriage studies showed that in addition to individual protection, serogroup C meningococcal carriage and transmission was significantly reduced in adolescents, particularly that of serogroup C cc11, resulting in population level herd immunity (Miller *et al.* 2001, Maiden *et al.* 2002, Ramsay *et al.* 2003, Maiden *et al.* 2008). Measurements of functional antibody and serum bactericidal activity (SBA) titres are used as correlates of protection (Borrow *et al.* 2001): these characterise bactericidal activity against meningococci using sera of vaccinated individuals and exogenous complement. It was observed that less than four years following immunization of children under three years of age with MCC vaccines in the UK, bactericidal antibodies waned with corresponding decreases in vaccine effectiveness (Pace and Pollard 2007), and therefore a 12-month booster was added to routine schedules in 2006. By 2009, MCC vaccines had been incorporated into routine schedules in 12 European countries, some as successfully as in the UK.

Tetravalent conjugate vaccines have been developed by various manufacturers: for example, Menactra® (Sanofi-Pasteur) was licensed in the US in 2005, Menveo® (Novartis) was licensed in the EU and in the US in 2010, and Nimenrix (GSK) was authorized in the EU in 2012. Based on estimates of reported vaccine failures and coverage, Menactra® is thought to have had 80% vaccine effectiveness in the few years following immunization (Macneil *et al.* 2011). Very recently, it has been announced that England and Scotland will be the first European countries to introduce a tetravalent conjugate vaccine on a national level: individuals aged 16-17 years of age will receive Menveo® or Nimenrix® from August 2015. An early study in students aged 18-24 years indicates that this vaccine may reduce carriage (Read *et al.* 2014).

Polysaccharide conjugate vaccines are costly, and a monovalent serogroup A conjugate vaccine, MenAfriVac™, was developed by the Meningitis Vaccine Project specifically for use in countries of the African meningitis belt (Marc LaForce *et al.* 2009). Mass vaccination began in 2010, and the few

cases of serogroup A disease in areas such as Burkina Faso and Chad subsequent to vaccine introduction suggest that it has been highly effective (Halperin *et al.* 2012, Kristiansen *et al.* 2013, Daugla *et al.* 2014). Although carriage studies indicated low levels of serogroup A carriage prior to vaccination, there is evidence for herd immunity in Chad (Kristiansen *et al.* 2014, MenAfriCar 2015).

Outer-membrane vesicle vaccines

Vaccines based on the serogroup B polysaccharide have been hindered by cross-reactivity with host polysialated glycoproteins, such as those on the neural cell adhesion molecule (NCAM), which results in immunological tolerance even when conjugated to a carrier, and possible induction of an autoimmune response (Finne *et al.* 1983). Attempts to formulate vaccines with chemically modified serogroup B polysaccharide have been hindered by safety concerns and lack of functional antibody activity in vivo (Jodar *et al.* 2002).

As a result of the immunogenicity of sub-capsular antigens, research has subsequently focused on formulation of serogroup B-substitute vaccines with outer-membrane proteins (OMPs); although often referred to as serogroup B vaccines, the protein variants they contain may be present in meningococci of other capsular groups. Outer-membrane vesicle (OMV) vaccines can be prepared using detergent extraction from meningococcal cells, and the resulting OMVs contain the major porins and lesser amounts of other proteins. Several have been developed in response to single strain serogroup B outbreaks. The first was VA-MENGOCOC-BC®, developed by the Finlay Institute in Cuba: this contained purified OMV from a B:4:P1.19,15 strain and serogroup C polysaccharide, and a 1987-1989 trial showed 83% efficacy in teenagers (Sierra *et al.* 1991). The second was MenBvac™, developed at the Norwegian Institute of Public Health (NIPH): this contained OMV from a B:15:P1.7,16 strain, and a school randomized, placebo controlled trial from 1988-1991 showed 57% efficacy in teenagers (Bjune *et al.* 1991). More recently, between 2004 and 2008, MeNZB® has been used to limit the serogroup B epidemic in New Zealand that started in 1991: this vaccine was based on a B:4:P1.7b,4 isolate, NZ 98/254, and vaccine effectiveness was estimated at 73% (O'Hallahan *et*

al. 2004, Arnold *et al.* 2011). In 2006, MenBvac™ was used in France to target an outbreak caused by a similar strain to that responsible for the Norwegian epidemic; because vaccine stocks were limited, the highest risk group, those aged 1-4 years, was targeted (Caron *et al.* 2011, Oviedo-Orta *et al.* 2015).

The utility of these vaccines is hampered by meningococcal diversity. Tailor-made vaccines to new outbreak strains may take years to develop (Jodar *et al.* 2002), and coverage of heterogeneous strains circulating in endemic settings is likely to be low since sequence variability of PorA VR2, the major immunogen, can significantly affect susceptibility of the strain to complement mediated lysis. Responses in infants in particular are restricted to the strain used in the vaccine. The nonavalent recombinant OMV PorA vaccine, NonaMen, was developed in the Netherlands to provide broader strain coverage: this includes OMVs from three genetically engineered strains each expressing three PorA sub-types, and has been shown to be immunogenic in mice and rabbits (van den Dobbelsteen *et al.* 2007, Kaaijk *et al.* 2013).

Outer-membrane protein vaccines

Two multicomponent vaccines, designed to provide broader protection against serogroup B disease, have recently been licensed. Both Bexsero® (GSK) and Trumenba® (Pfizer) are formulated with multiple recombinant outer-membrane proteins discovered through reverse vaccinology (Pizza *et al.* 2000, Giuliani *et al.* 2006, Nissen *et al.* 2013). This process was made possible through next generation sequencing (NGS) technology: the genome of a cc32 isolate, MC58, was bioinformatically mined for the identification of vaccine targets. Of 600 genes coding for surface-exposed or exported proteins, 25 were identified to induce bactericidal antibodies, and following assessment of their sequence diversity, a number were likely to induce immunity against multiple strains (Pizza *et al.* 2000). One such antigen was factor H binding protein (fHbp), also independently discovered using biochemical fractionation and known as LP2086 (Fletcher *et al.* 2004).

Two fHbp sub-variants, each from a different fHbp variant family, are included in the Trumenba® vaccine (also called rLP2086): fHbp-3.45 and fHbp-1.55 (Nissen *et al.* 2013). Trumenba® was FDA-approved in 2015 for immunization of those aged 10-25 years, but not for use in infants since it caused a fever in a large proportion of recipients (Martinon-Torres *et al.* 2014). This vaccine was immunogenic in children aged 8-14 years (≥4-fold rise in human serum bactericidal assay (hSBA) titre, or seroconversion, in up to 95.3% of recipients following three doses) (Nissen *et al.* 2013), and in adults aged 18-40 years (hSBA titre ≥1:4, or seroprotection, in 94.3%) (Marshall *et al.* 2013). Trumenba® was administered to students in response to USA university outbreaks (in Oregon and Rhode Island) of serogroup B disease in the winter of 2014/2015 (Oviedo-Orta *et al.* 2015).

The Bexsero® vaccine (also called 4CMenB) is formulated with recombinant fHbp-1.4 along with three additional components: NadA variant 3.1, recombinant NHBA variant 1.2, and the OMV used in MenNZB that includes PorA VR2 1.4; the anti-PorA immune response mainly targets the VR2 variant (Giuliani *et al.* 2006, Martin *et al.* 2006). Bexsero® was licensed for use in infants and adolescents in Europe, North America, and Australia in 2015, and was administered to 94% of undergraduates at Princeton University, and to students of the University of California, in response to serogroup B outbreaks in the final months of 2013 (Oviedo-Orta *et al.* 2015). Following some uncertainty concerning the cost-effectiveness of Bexsero® in national immunisation of infants and adolescents (Andrews and Pollard 2014), it is to be administered to those aged more than two months in the UK in a three dose schedule from September 2015.

Immunogenicity of Bexsero®

Estimating the degree of protection afforded by group B multicomponent vaccines at both the individual and population level is complex. This is partly because direct vaccine effectiveness studies cannot be carried out when incidence is very low (Seib *et al.* 2015b): for example, there were 424 group B cases in England and Wales in the 2013/14 epidemiological year, and incidence is declining year on year. In the absence of these data, the SBA assay was used to inform licensure of Bexsero, as

it was for MCC vaccines in the late 1990s. Whilst of great utility, there are some complexities with using SBA for this purpose (Andrews and Pollard 2014, Harrison 2015). For example, because the anti-fHbp response is human specific, human rather than rabbit serum is required as the antibody source (hSBA assay). Since many adults, who produce the greatest volume of suitable serum, naturally possess antibodies against the target strains, it can be difficult to identify suitable candidates, and a variety of different serum sources are needed to test the diverse serogroup B meningococcal strains circulating in the current endemic period. This reduces comparability among assays.

However, immugenicity to Bexsero® components investigated using this approach is high in infants, children, and adults (McIntosh *et al.* 2015, Seib *et al.* 2015b). For example, phase II and III studies show that 100% of infants had anti-fHbp hSBA seroprotection following three doses and that immune responses were high against the other antigens also (Snape *et al.* 2010, Gossger *et al.* 2012, Vesikari *et al.* 2013). Reactogenicity was greater when Bexsero® was administered alongside other routine infant vaccinations, but was not high enough to be considered unsafe. Persistence of the antibody response varied by antigen, and is lower in infants than in adolescents: children under five years show waning antibody responses and require a booster dose, and seroprotective titres were sustained in the majority of adolescent subjects up to two years following two doses (Santolaya *et al.* 2013, McQuaid and Snape 2014).

Bexsero® coverage

Quantification of the coverage of a given population by multicomponent vaccines such as Bexsero® is complex; the vaccine would need to be administered to an entire annual birth cohort to detect statistically significant effects in the UK, and post-implementation analysis and detailed surveillance in the coming years will be of great interest. Whether Bexsero® induces herd immunity or not will greatly affect efficacy; one study has indicated a reduction in carriage in students following Bexsero® immunisation, however, the magnitude of such an effect should the vaccine be routinely

administered to adolescents could not be estimated from these results (Read *et al.* 2014). Instead, current efficacy estimates are based on a combination of immugenicity data for strains homologous and heterologous to those used in vaccine design, combined with knowledge of the vaccine antigen profiles of temporally and geographically circulating strains.

In endemic periods in particular, there may be a diversity of disease causing serogroup B strains. These may not possess the gene for the vaccine antigen at all, they may possess diverse antigenic variants that may or may not stimulate cross-reactive antibody responses to variants included in vaccines, and expression of antigens may be variable. For example: *nadA* is absent from cc41/44 meningococci, the cause of a large proportion of disease in the UK, and isolates lacking *porA* have been identified (Comanducci *et al.* 2004, Lucidarme *et al.* 2010). Estimates of the prevalence of Bexsero® vaccine antigen variants in disease-causing meningococci range from 70% of endemic serogroup B strains in Norway (conservative estimate based on exact sequence matches to at least one component) (Holst *et al.* 2014), to 12% for fHbp and 27.7% for NHBA in serogroup B isolates from European countries (Vogel *et al.* 2013), and to 36% for PorA VR2 in the Netherlands (Bijlsma *et al.* 2014). There is evidence of cross-protection, such as among heterologous NHBA variants and among fHbp variant 1 sub-variants, however this may be less in infants (Giuliani *et al.* 2006, Brunelli *et al.* 2011, Gossger *et al.* 2012). In addition, anti-fHbp and anti-NHBA reactions may be synergistic (Vu *et al.* 2011). Expression of some NadA variants is phase-variable, and the abundance of fHbp expressed affects bactericidal activity (Seib *et al.* 2015b).

To circumvent some of these issues, the Meningococcal Antigen Typing System (MATS) was developed to allow estimation of vaccine coverage (Donnelly *et al.* 2010). This involves an antigen specific ELISA to estimate the extent of antibody cross-reactivity between the antigen variant in the strain of interest and the variant included in Bexsero, and a threshold value that predicts protection against a given strain is defined based on the level of antibody activity from the SBA assay. MATS predicts that 86% of adults and 77% of infants would be covered by Bexsero® globally (Donnelly *et*

al. 2010). Highly favorable strain coverage has also been predicted in Europe, with between 78% and 89% of serogroup B meningococci likely to be covered (Vogel *et al.* 2013, Tzanakaki *et al.* 2014), and it has been suggested that MATS may underestimate coverage (Frosi *et al.* 2013). However, MATS assays have been based on pooled infant sera since individuals produce low levels of antibody, and therefore these estimates may not be reliable indicators of individual protection (Andrews and Pollard 2014).

Population effects of vaccine implementation

In addition to post-implementation effectiveness monitoring, it will be extremely important to undertake detailed surveillance for detection of the effects of Bexsero® and the quadrivalent vaccine on meningococcal populations following national immunization. Vaccination may incur mucosal immunity, resulting in decreased rates of carriage acquisition among vaccinees and concomitant reduction in transmission rates to the unvaccinated population. This herd immunity effect, whilst of great importance in increasing vaccine efficacy, may provide the opportunity for expansion of non-vaccine strains to which the human population is not immune. In addition, selective pressure on vaccine strains may result in antigenic shift or expansion of altered, potentially virulent, variants (Harrison *et al.* 2006, Krauland *et al.* 2012).

Antigenic shift arising from population immunity is thought to contribute to fluctuations of meningococcal strains competing in natural populations (Buckee *et al.* 2008). Although capsule switching following MCC vaccine introduction was not recorded (Gray *et al.* 2006), expansion of serotype 19A of *S. pneumoniae* by recombination-mediated capsule switching was recently observed following routine use of a pneumococcal conjugate vaccine in infants in the USA (Golubchik *et al.* 2012). In addition, since carriage of meningococci and commensal *Neisseria* is immunizing, complete elimination of carriage-associated meningococcal lineages, or *N. Lactamica* with the *lactamica* vaccine (Gorringe *et al.* 2009), could have real consequences on natural immunity to the meningococcus. Bexsero® may affect carriage of commensal species that also possess NHBA

(Lucidarme *et al.* 2013b). It is therefore of great importance to not only monitor invasive disease lineages following vaccination, but to perform detailed pre- and post- implementation carriage characterisation that may capture these effects (Maiden and Spratt 1999).

Genomic epidemiology

Next-generation sequencing technology

The advent of cost-effective, whole-genome sequencing of multiple bacterial isolates is transforming molecular epidemiology. Since the first bacterial whole-genome sequence (WGS) was published in 1995 (Fleischmann *et al.* 1995), shotgun sequencing has given way to next-generation sequencing (NGS) technology: second-generation sequencing with Illumina technology is particularly popular, and third-generation technologies such Pacific Biosciences (PacBio) single-molecule real-time (SMRT) sequencing and Oxford Nanopore sequencing are becoming increasingly accessible (Quail *et al.* 2012).

Illumina sequencing

Currently, ultra-high throughput Illumina sequencing is carried out on a HiSeq machine and small-scale sequencing is rapidly performed on a bench top MiSeq. DNA libraries are created by enzymatic or mechanical shearing of the DNA sample into 200-500bp fragments, followed by ligation of adaptors for sample tagging, flow-cell attachment, PCR amplification, and sequencing. The library is fixed to a flow cell and each molecule is amplified over 1000 times to form clonal clusters through bridge-amplification. Nucleotides in the sequencing reagent are fluorescently labeled; each cycle of sequencing includes addition of a single base followed by termination with a 3'-OH terminator and imaging via laser excitation. Following removal of the terminator, the next cycle can begin (Bentley *et al.* 2008). The length of sequencing reads is dependent on the number of sequencing cycles carried out: the Sanger Centre now offers short-reads up to 100bp in length, and the number of

reads obtained per genome is dependent on the number of samples multiplexed per sequencing run.

PacBio sequencing

The PacBio RS sequencer was launched in 2011. A single-DNA molecule is sequenced in a tiny hole called a zero mode waveguide (ZMW), thousands of which are contained on SMRT cells. A single polymerase-DNA template complex is tethered to the bottom of each ZMW. As fluorescent dye-labeled nucleotides are held by the polymerase, a light pulse specific to the nucleotide is emitted that is terminated by cleavage of the flurophore upon incorporation to the growing DNA strand. This pulse is captured by cameras in real time (Eid *et al.* 2009). The sequencing reaction terminates when the DNA and the polymerase dissociate, and therefore read lengths have an approximately log-normal distribution. SMRT reads are substantially longer than Illumina reads, and are on average 3000bp and up to 20,000bp in length. In addition, since DNA is not amplified prior to sequencing, epigenetic modifications on the template are not removed; incorporation of nucleotides opposite methylated bases produces a particular pulse, which allows sequencing of epigenetically modified DNA sequence (Clark *et al.* 2012).

Genomes in public health

The choice of NGS technology and platform is dependent on financial considerations, the organism to be sequenced, and the biological question being addressed. For example, the long reads produced by SMRT sequencing are likely to encompass the short-sequence repeats that are numerous in the genome of *N. meningitidis*, and therefore facilitate complete genome assembly. Technologies which include an amplification step may suffer from amplification bias, particularly in GC rich regions (Quail *et al.* 2012). However, PacBio sequencing is currently costly and requires a statistical error reduction step to account for random errors. In general, as long as base calling is accurate, draft genomes are adequate for the majority of analyses (Guttman and Stavrinides 2010): meningococcal genomes sequenced with current Illumina chemistry assemble onto fewer than 300

contigs and this is more than sufficient for extraction of meningococcal typing information (Bratcher *et al.* 2014).

NGS technology delivers almost all the genetic information for multiple isolates in a single process. For molecular epidemiology, this means that information for routine diagnostics through to population structure resolution can be obtained from the same data source. When WGS data is obtained from bacterial populations, such as in routine national surveillance, this is of clear advantage for applications requiring an understanding of pathogen evolution. Resolution of fine-scale isolate relationships allows investigation of strain microevolution and transmission dynamics, which aids identification of unrecognized outbreaks and their sources, risk factors for acquisition, and intervention measures (Didelot *et al.* 2012). Genome sequencing of *Staphylococcus aureus* isolated from patients admitted to an intensive care unit over 14 months demonstrated that 19% of acquisitions occurred whilst on the ward, highlighting the importance of sanitation measures (Price *et al.* 2014).

At the macroevolutionary level, resolution of population structure and geo-temporal dynamics allows characterisation of emerging lineages that are of public health concern, and provides insight into the population processes shaping them. It is hoped that knowledge at this level will allow prediction of epidemiological patterns and advanced preparation of attendant control measures. One of the first demonstrations of population-level pathogen evolution with WGS data was performed on retrospectively sequenced *S. aureus* isolates; this demonstrated the role of nosocomial transmission in global dissemination of an antibiotic resistant lineage (Harris *et al.* 2010). WGS of well defined population samples are essential for designing tailor-made vaccines and post-implementation monitoring. In addition, genome-wide association studies, conceivable with large, unbiased, representative samples, will identify the genotypic basis of phenotypes such as antibiotic resistance or ecological niche specificity (Falush and Bowden 2006, Sheppard *et al.* 2013).

Population Genomics: methodological approaches

Whilst acquisition of great volumes of WGS data is now straightforward, methods for their analysis, both theoretical and computational, are still developing. This is particularly true for the aforementioned applications, which are dependent on reconstruction of intraspecies evolutionary history. Population genomics, described as the study of evolutionary processes that influence variation across populations using whole-genome data (Gulcher and Stefansson 1998), is a nascent field. Although classical population genetics theory is applied to WGS data, there are theoretical and methodological limitations since it was largely developed for analysis of single gene sequences; for example, forces on genomes and on individual genes are not equivalent, and population bottlenecks and neutral drift, heterogeneous genome-wide recombination, mutation, and selection rates, and hitchhiking effects, are incompletely understood. This rationale underlies the choice of unlinked, housekeeping genes for multilocus typing, since these portions of the genome are likely to contain neutral, synonymous variation. Conversion of nucleotide sequence variation at these loci into alleles reduces distortion of phylogenetic signal by recombination, and, although MLST was originally designed for epidemiological typing, temporally stable clonal groupings of meningococci are delineated. For isolates of the same ST, MLST distances are likely to represent the level of evolutionary divergence, however, for distantly related isolates, MLST loci are unlikely to provide the resolution necessary to reconstruct evolutionary history (Feil 2004). For example, it was shown that even with 20 loci, relationships of meningococcal clonal complexes could not be robustly reconstructed (Didelot *et al.* 2009).

The gene-by-gene approach

It is becoming apparent that clonal complexes are representative of consensus clonal patterns of descent: genealogies generated with meningococcal WGS data have demarked clusters analogous to clonal complexes (Budroni *et al.* 2011, Hao *et al.* 2011). Since housekeeping genes are a subset of the core-genome, and clonal complexes are temporally stable, this appears intuitive. In theory, therefore, seven-locus data can be expanded to include additional core-genome loci for high

resolution, sub-clonal complex isolate comparisons. For isolates closely related by short transmission chains, the majority of core-genome loci will be required, and the availability of multiple sequenced genomes permits careful locus selection. This approach has been termed ‘gene-by-gene’ analysis (Sheppard *et al.* 2012, Maiden *et al.* 2013) (Fig. 3). One scheme, rMLST, was introduced in 2012 and is based on 53 universally possessed ribosome-encoding genes: these loci are unlikely to be linked or under diversifying selection and are present in in all bacteria, and the level of variation present allows speciation down to strain typing (Jolley *et al.* 2012a).

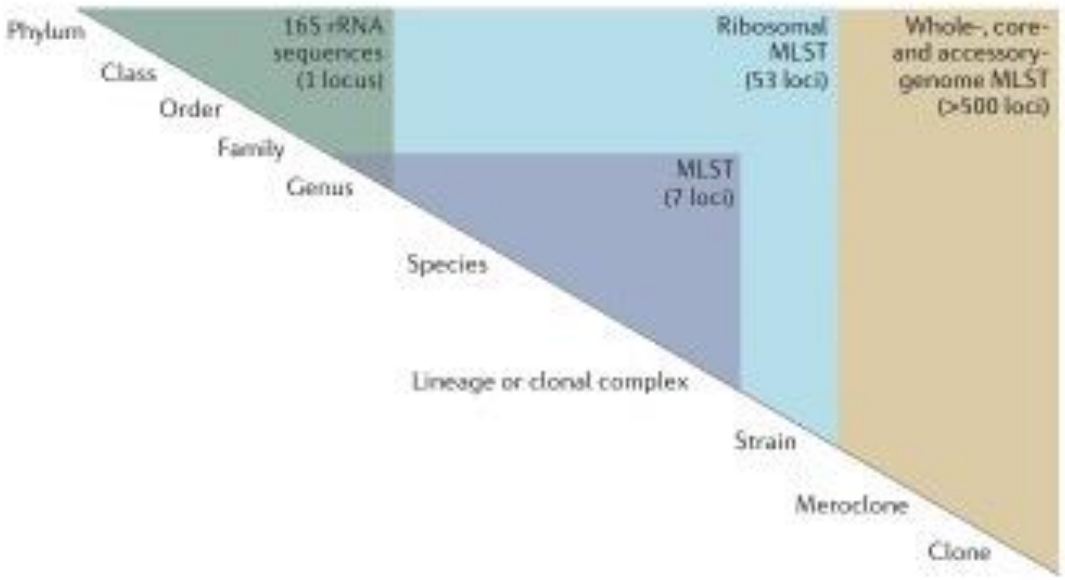


Figure 3: The gene-by-gene approach to genomic epidemiology. The number of loci required for discrimination of bacterial isolates is dependent on their relatedness. Source: Maiden, M. C., M. J. van Rensburg, J. E. Bray, S. G. Earle, S. A. Ford, K. A. Jolley and N. D. McCarthy (2013)"MLST revisited: the gene-by-gene approach to bacterial genomics” Nat Rev Microbiol 11(10): 728-736 (Maiden *et al.* 2013)

The methodologies used for analysis of seven-locus MLST data should therefore apply to genome level multilocus data. For example, eBURST is a clustering method that identifies clonal groupings in populations without inferring the relationships between them (Feil *et al.* 2004), and STRUCTURE identifies patterns of gene flow in order to cluster isolates into populations (Falush *et al.* 2003). Alternatively, distance based phylogenetic methods, such as Neighbour Joining, can be used on

multilocus data to generate a phylogenetic tree; where recombination rates are particularly low, or where isolates are highly similar, it may be appropriate to use nucleotide distances rather than allele data (Feil 2004). Statistical phylogenetic methods, such as Maximum-Likelihood and Bayesian analysis, are tree-searching methods that often rely on a heuristic strategy to identify the most probable tree from sequence data. However, all phylogenetic trees impose a bifurcating model of evolution that is likely to be inappropriate for a species known to be highly recombining (Hedge and Wilson 2014). ClonalFrame is a Bayesian algorithm designed specifically to account for recombination in multilocus sequence data, however, it assumes that alleles are imported from outside the population in HGT events (Didelot and Falush 2007). Where network population structures are likely, methods such as splits-decomposition and Neighbor-Net allow generation of distance networks (Huson and Bryant 2006): network analysis permits isolates that are extant ancestors of other isolates in the sample (likely in an outbreak sample) to be mapped as such, and identification of homoplasies.

The SNP approach

Much of the recent genomic epidemiology literature employs an alternative approach for the deduction of pathogen isolate evolutionary histories. This is based on the extraction of nucleotide sites in the genome that vary among isolates, referred to as single nucleotide polymorphism (SNP) (Croucher *et al.* 2013) or single nucleotide variation (SNV) data (Price *et al.* 2014). These data are often used for phylogenetic reconstruction and demographic inference using algorithms such as BEAST (Drummond *et al.* 2012). For monomorphic pathogens, such as *Mycobacterium tuberculosis* or *Yersinia pestis,* this approach obtains the most detailed genetic information, allowing generation of high resolution phylogenies reflective of evolutionary history that may be mapped against temporal and geographical information (Parkhill and Wren 2011, Ford *et al.* 2013). The suitability of these approaches for highly recombining organisms such as *N. meningitidis* is unclear: over long evolutionary timescales nucleotide substitutions are increasingly likely to have arisen through recombination, and among isolates sampled from an outbreak, may reflect recent, unfixed point

mutations that are phylogenetically uninformative. Studies in organisms such as *Streptococcus pneumoniae* have partially addressed this issue by the removal of clusters of SNPs likely to have been introduced by recombination (Croucher *et al.* 2011), and by the delineation of SNV cut-off values that permit identification of isolates from direct transmission events (e.g. <2 isolate pair-wise SNVs in *Clostridium difficile*) (Eyre *et al.* 2013b).

Genome assembly and alignment

These approaches encounter practical complications primarily created by the lag between the advance of NGS technologies and methodologies designed specifically for analysis of their data. Available computing power can be a severe hindrance on the analysis of large volumes of data using statistical methods, and even extraction of genetic variation information is not straightforward. To obtain SNP data, sequencing reads must be aligned against a reference genome for identification of homologous nucleotide sites. The reference genome must be sufficiently related that genetic information is not lost through absence of homologous regions, and genome alignment is a significantly complex theoretical and computational issue.

Alternatively, sequence reads can be assembled into a genome '*de-novo'*, or without reference to a complete genome. This is also a computationally difficult task: reads containing identical sequence such as homopolymeric tracts generated by insertion elements cannot always be distinguished from one another, which results in assembly break points at the ends of 'contigs'. The Velvet algorithm is particularly popular for de-novo assembly since it allows the user to adjust its parameters in order to balance specificity and sensitivity (Zerbino and Birney 2008, Salzberg *et al.* 2012). Mauve was developed for alignment of multiple assembled genomes (Darling *et al.* 2010). Whilst informative in studies of genome structure, this approach is not appropriate for alignment of many diverse genomes, e.g. for phylogenetic applications: only around 20 genomes may be aligned depending on their diversity.

The gene-by-gene approach escapes many of these issues. Genomes are assembled de-novo, and only variation at genes of interest needs extracting. For example, allele distances at each locus between each pair of genomes can be extracted to generate a distance matrix, which can be used to draw a network graph or Neighbour Joining tree. A prerequisite of this approach is identification and annotation of the loci required for analysis, however (Jolley and Maiden 2013). Previously characterised loci can be identified in new genomes using homology searching algorithms such as BLAST; these encounter problems in draft genomes where paralogues may not be distinguishable due to the lack of useful synteny information. Alternatively, bioinformatic prediction of novel open-reading frames can be achieved using programs such as Glimmer (Salzberg *et al.* 1998); prediction of encoded protein function may be possible by cross-referencing homologous sequence with large protein databases, however, when automated it can be inaccurate (Richardson and Watson 2013).

To date, a great deal of work has been undertaken to manually curate and annotate great numbers of meningococcal genomes in the PubMLST database in order to limit these issues. This process has been dubbed 'population annotation' (Bratcher *et al.* 2014) and the majority of analyses in Chapters 2 to 5 of this thesis make use of this process as a precursor to the gene-by-gene approach (Fig. 3) to genomic epidemiology.

Chapter 2. Genomic epidemiology of age-associated meningococcal lineages in national surveillance: an observational cohort study

Abstract

Background

Meningococcal disease is a potentially vaccine preventable global health problem; however, given the sporadic nature of meningococcal disease, and the high diversity of *Neisseria meningitidis*, epidemiological surveillance incorporating detailed isolate characterisation is crucial for effective control and for understanding its evolving epidemiology. The 'Meningitis Research Foundation Meningococcus Genome Library' (MRF-MGL) exploits whole-genome sequencing (WGS) for this purpose and presents, via the Internet, WGS data on a comprehensive and coherent meningococcal disease isolate collection from England and Wales.

Methods

WGS data were obtained from all meningococcal disease isolates available for England and Wales over the 2010/11 and 2011/12 epidemiological years. The data were annotated at >1,700 loci, analysed, and disseminated at http://pubmlst.org/neisseria using the gene-by-gene approach. This provided information on vaccine antigens and population structure, enabling identification of meningococcal disease-associated genotypes over time and by age of patient. A lineage nomenclature is suggested to facilitate compatibility between WGS-characterised lineages and traditional typing methods.

Findings

The methods efficiently characterised meningococcal disease isolates, providing plain language information that was compatible with previous approaches. At least 20 meningococcal lineages were identified, with three responsible for 59% of meningococcal disease isolates. Within lineages there

was high diversity and evidence of recombination. Certain lineages were associated with meningococcal disease in particular age-groups, increases in cc23 (lineage 23) and cc11 (lineage 11) disease were evident in 2011/12, and the increased incidence from 1984-2010 in England and Wales was due to successive and concurrent epidemics of different lineages.

Interpretation

The MRF-MGL represents an effective, broadly applicable model for the storage, analysis, and dissemination of WGS data that can facilitate real-time genomic WGS pathogen surveillance. For the meningococcus, these data reveal information crucial for effective vaccine deployment and assessment.

Introduction

Meningococcal disease, caused by *Neisseria meningitidis* and encompassing meningitis and severe sepsis, has been reported at varying rates for over two hundred years world-wide. The meningococcus is ordinarily a harmless commensal of the oropharynx and both disease incidence and asymptomatic carriage rates are age-specific (Caugant and Maiden 2009). Meningococcal disease is generally sporadic with endemic rates under three cases per 100,000 population, but elevated incidence also occurs as hyperendemic outbreaks, intense localised outbreaks, and epidemics (Caugant and Maiden 2009). These fluctuations are thought to be due to interactions between meningococcal population biology and human susceptibility, and possibly evolution and spread of particularly invasive strains (Stephens *et al.* 2007). Almost all meningococcal disease is caused by meningococci expressing the polysaccharide capsules that define serogroups A, B, C, W, X, and Y (Caugant and Maiden 2009), and multilocus enzyme electrophoresis (MLEE) (Caugant and Maiden 2009) and multilocus sequence typing (MLST) (Maiden *et al.* 1998) have demonstrated that most disease is caused by temporally stable hyperinvasive lineages, corresponding to certain 'clonal complexes' (ccs).

Successful control of meningococcal disease caused by several serogroups has been achieved by vaccination, such as the introduction of meningococcal serogroup C conjugate (MCC) vaccines to the UK. Here, herd immunity led to a dramatic reduction in serogroup C disease (Fig. 1A) (Maiden *et al.* 2002, Gray *et al.* 2006). However, due to poor immunogenicity and safety concerns, there are no equivalent vaccines against serogroup B polysaccharides. Serogroup B-substitute vaccines are based on proteins and often include outer-membrane vesicles (OMVs) (Giuliani *et al.* 2006, Serruto *et al.* 2012); vaccine coverage therefore depends on the serological cross-reactivity that they generate. As such vaccines are introduced, comprehensive surveillance is necessary to monitor antigenic variants, their expression, and their association with hyperinvasive lineages.

Meningococcal disease has been reportable in England and Wales since 1912, and, even accounting for alterations in reporting definitions, wide incidence fluctuations are evident (Fig. 1B) (Abbott *et al.* 1985). Currently, meningococcal disease reporting encompasses 'meningitis and septicaemia', the latter diagnosis referring to severe sepsis. Due to antibiotic therapy, around half of cases are culture-confirmed (CCCs) and the Public Health England (PHE) Meningococcal Reference Unit (MRU) receives the great majority of both these and specimens for non-culture case confirmation (NCCC) (Fig. 1A) (Gray *et al.* 2006). Linkage with epidemiological data demonstrates that the PHE-MRU collection of isolates and NCCC specimens represents 98% of clinically diagnosed cases in England and Wales (Heinsbroek *et al.* 2013). Such isolates are routinely characterised by serological methods and single gene sequencing (Gray *et al.* 2006).

The potential of whole-genome sequence (WGS) data for bacterial characterisation in clinical microbiology (Didelot *et al.* 2012, Jolley *et al.* 2012a), transmission analyses (Bryant *et al.* 2013), and post-vaccine implementation monitoring (Golubchik *et al.* 2012) has been demonstrated by a number of case studies. However, if this potential is to be realised in routine surveillance, sustainable, transparent, and compatible infrastructures are needed for data storage, analysis, and reporting. The Meningitis Research Foundation Meningococcus Genome Library (MRF-MGL) described here presents such a framework. WGS data from all 899 English and Welsh meningococcal disease isolates submitted to PHE-MRU in the 2010/11 and 2011/12 epidemiological years were obtained and used to investigate meningococcal disease epidemiology.

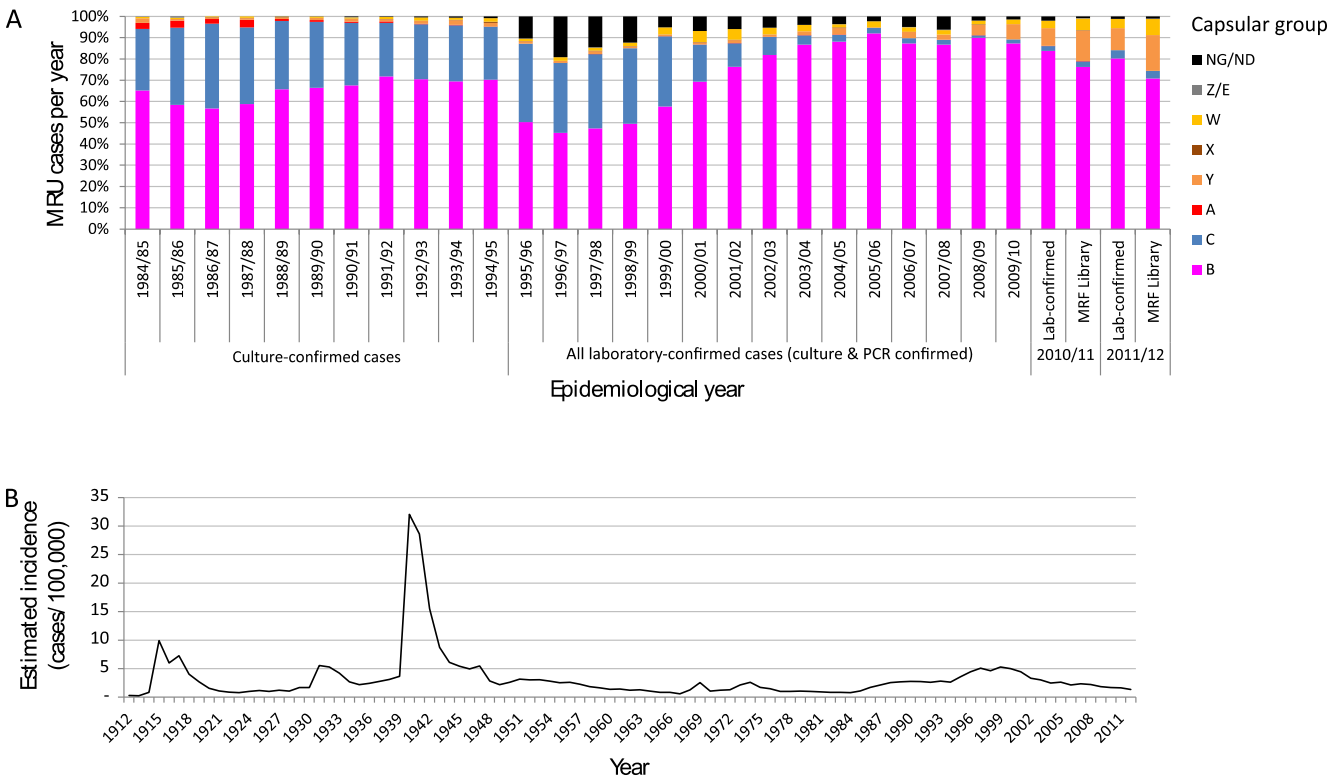


Figure 1. Capsular groups (1984-2012) and notifications of meningococcal disease in England and Wales (1912-2012). (A) Capsular groups determined from culture and non-culture specimens at the MRU (Gray *et al.* 2006), and from MRF-MGL genomes (culture-confirmed cases only). Not determined (ND): samples with insufficient meningococcal DNA to confirm capsular group or non-A, B, C, W or Y. Non-groupable (NG): capsule polysaccharide not expressed or non-A, B, C, W, X, Y, Z, 29E. (B) Estimated incidence of meningococcal disease calculated from: 1912-1997 notification data, supplied by Mary Ramsay, PHE, and 1998-2011 laboratory reports and 2012 Health Protection Report, Weekly report, Volume 7 Number 8, Published on 22 February 2013 (both www.HPA.org). PCR-diagnosis introduced in 1996 greatly increased MRU case ascertainment (Gray *et al.* 2006). Reporting criteria and methods have changed on multiple occasions.

Systematic literature review

PubMed was searched with the terms "*Neisseria meningitidis*", "genome sequence", and "epidemiology" for reports published in any language before April 20, 2015. Following the initiation of the MRF-MGL project in 2011, four studies reported the use of WGS data for retrospective resolution of single lineage *N. meningitidis* outbreak or hyperendemic scenarios (Jolley *et al.* 2012b,

Krauland *et al.* 2012, Harrison *et al.* 2015, Ladhani *et al.* 2015) and five case studies were identified in which clinical isolates were retrospectively whole-genome sequenced for strain characterisation (Rishishwar *et al.* 2012, Vogel *et al.* 2012, Koser *et al.* 2013, Reuter *et al.* 2013, Piet *et al.* 2014). Three studies undertook WGS of 16-108 meningococcal isolates to investigate meningococcal lineage structure using universally present genes (Budroni *et al.* 2011, Hao *et al.* 2011, Bratcher *et al.* 2014): both Budroni and colleagues (Budroni *et al.* 2011) and Hao and colleagues (Hao *et al.* 2011) presented robust phylogenetic relationships between isolates of different clonal complexes that contrasted with poor within-clonal complex resolution. In addition to PubMLST.org (Jolley and Maiden 2010, Jolley and Maiden 2013), used here, two publically available *N. meningitidis* comparative genomic databases were identified, both of which included only previously-published genomes and neither of which provided integrated clinical characterisation of strains or strain relationships (Rusniok *et al.* 2009, Katz *et al.* 2011). Nine reports had been published which make use of the MRF-MGL data in taxonomic, comparative genomic (e.g. core-genome delineation), and vaccine antigen characterisation studies (listed at http://www.meningitis.org/current-projects/genome) (Hubert *et al.* 2013, Hung *et al.* 2013, Lucidarme *et al.* 2013a, Oldfield *et al.* 2013, Bennett *et al.* 2014, Chatt *et al.* 2014, Clark *et al.* 2014, van Tonder *et al.* 2014, Wormann *et al.* 2014). No studies reporting either comprehensive *N. meningitidis* WGS data or open-access annotated WGS data from national surveillance of disease caused by other bacterial species were found.

Interpretation

This study adds to previous use of whole-genome sequencing in public health microbiology by providing the infrastructure needed to process, analyse and interpret bacterial WGS data in clinical settings. This framework can in principle be used in real-time analysis of WGS data collected on a case-by-case basis, from meningococcal disease or other bacterial diseases, for example in diagnosis or outbreak investigation. The approach was validated by demonstrating concordance between national reference laboratory typing of all *N. meningitidis* isolates collected over two years and

WGS-derived typing information, and between lineages defined by multi-locus typing methods and WGS-defined lineages. A nomenclature is suggested for the practical utilization of the latter. Use of MRF-MGL data is demonstrated by showing that current low meningococcal disease incidence results from a decline in cases caused by a small number of hyperinvasive lineages and that these lineages are associated with patients of different ages. The annotated WGS data are open-access, enabling community investigation of the genetic basis of these epidemiological traits, among other studies.

Methods

Specimen collection

All 899 meningococcal disease isolates submitted to PHE-MRU over two epidemiological years (July 1st – June 30th) were included: 501 isolates (47.57% of laboratory-confirmed cases) for 2010/11 and 398 isolates (51.96% of laboratory-confirmed cases) for 2011/12 (PHE 2014). The PHE Vaccine Preventable Invasive Bacterial Infections Forum agreed that data on year and UK region of isolation would be publicly available for each isolate, with other demographic data available from the PHE-MRU on request.

Procedures

Culture and DNA extraction at PHE-MRU

Isolates were cultured on Columbia Agar plus 5% (v/v) horse blood and stored on Microbank™ Bacterial & Fungal Preservation System vials (Prolab Diagnostics, Ontario, Canada) at -80°C. Prior to DNA extraction, isolates were streak plated onto Columbia Agar plus 5% (v/v) horse blood and incubated overnight at 37°C in an atmosphere containing 5% CO_2. DNA was extracted using the DNeasy Blood and Tissue kit (Qiagen) in which a sterile 1 µL inoculation loop was used to transfer and disperse (by rapid rotation between thumb and forefinger) heavy sweeps of non-confluent growth directly to a screw-capped microfuge tube containing 180 µL of buffer AL and 20 µL proteinase K. This was then incubated at 56°C with occasional vortexing for at least 2 hours. DNA extraction was then completed in accordance the manufacturer's DNeasy® Blood & Tissue Handbook (July 2006; Gram negative bacteria protocol) using two 75µl elution steps.

Whole-genome sequencing at the Sanger Institute

Standard Illumina libraries were generated using 1µg of genomic DNA sheared to 200-300bp using a Beckman FX robot. Except for twelve pilot samples which constituted a single pool and were sequenced to produce 75bp reads, 96 tagged DNA samples were pooled in an equimolar ratio for

sequencing in a flowcell lane on the Illumina Hi-Seq platform, generating 100bp paired-end reads. All sequence read data passed the Sanger Institute's in-house quality control assessment.

WGS assembly

FASTQ files were entered to an automated pipeline that integrates Velvet version 1.2.01 (Zerbino and Birney 2008), VelvetOptimiser version 2.2.0, and BIGSdb software (Jolley and Maiden 2010, Bratcher *et al.* 2014) for assembly of draft genomes. Contiguous sequences (contigs) were not scaffolded and were entered to the database only if >200bp. There was no manual improvement of assemblies at this stage of the project.

Assemblies were assessed for the presence of 1,605 full-length, core-genome (cg) loci belonging to the *N. meningitidis* cgMLST V1.0 scheme, and for incompletely assembled coding sequences (CDS) as defined in the PubMLST.org/neisseria sequence definitions database. Allele information obtained was more than sufficient for the results presented (typing and identification of lineages). The SPAdes assembler (Bankevich *et al.* 2012), introduced subsequent to the start of this project, produced outputs highly similar to those of Velvet: both assemblers resulted in identical allele results for up to 98% of the 1,605 loci. For consistency, all isolates were assembled using Velvet. Additional information on the methods chosen for processing genomic sequence data can be found in Bratcher *et al.* (2014) *BMC Genomics* (Bratcher *et al.* 2014).

WGS annotation

Assemblies were deposited in the PubMLST *Neisseria* database (http://pubmlst.org/neisseria/) (Jolley and Maiden 2010) where they are open-access and annotated with cgMLST loci and loci defining genogroup (Harrison *et al.* 2013b), typing antigens (Maiden *et al.* 2013), and vaccine antigens (Bambini *et al.* 2014). Features automatically reported by the database included: antigen variants; MLST sequence types (STs) and clonal complexes (cc); and rMLST sequence types (rSTs) based on 49 *Neisseria* ribosomal MLST (rMLST) loci (Jolley and Maiden 2010).

Where allelic variants were already stored in the database, or where BLAST hits were within 98% identity and alignment to stored alleles, annotation was automatically performed by the 'autotagger' and 'autodefiner' tools. Otherwise, web-based sequence tagging (Jolley and Maiden 2013) was used for manual curation: BLAST hits were extracted and aligned in MEGA6 (Tamura *et al.* 2013) for upload of new alleles, or, missing loci and loci interrupted by ends of contiguous sequences (contigs) were assigned alleles '0' and 's' respectively. Single MLST locus sequences were interrupted by the ends of contigs in eight genomes: these were individually sequenced at PHE-MRU to complete MLST profiles (Lucidarme *et al.* 2009).

Data analysis

Population genomic analyses were carried out with a hierarchical gene-by-gene approach (Maiden *et al.* 2013, Bratcher *et al.* 2014) enabling rapid, reproducible visualisation of genome similarity across the entire dataset. The BIGSdb Genome Comparator tool was used to generate allele-based distance matrices that were visualised as Neighbor-Net graphs in SplitsTree version 4.12.6 (Huson 1998). This approach accounted for the high rates of recombination in meningococcal populations, without discarding variant loci considered to be generated by recombination (Maiden *et al.* 2013). Loci missing in isolates, due to absence in the genome or incomplete assembly, were ignored in pair-wise comparisons. Congruence between clonal complexes and rMLST clusters was assessed by calculating adjusted Wallace Coefficients (AW) (Severiano *et al.* 2011) using the online tool at www.comparingpartitions.info. Maximum-likelihood trees were generated in MEGA6 from nucleotide sequence data extracted from the MRF-MGL: locus alignments were concatenated using the Genome Comparator MAFFT implementation and variable-sites were extracted. Trees were annotated using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Statistical analyses were undertaken in R (R Development Core Team 2012) with $\alpha = 0.05$. Chi square and Fisher's exact tests were carried out using standard R functions. Between-year relative risk ratios and confidence intervals (Wald estimates) (Rothman 2008) were calculated using the

'epitools' package (Aragon 2012) with the function 'epitab'. Human population data were extracted from the Office for National Statistics (ONS) (http://www.ons.gov.uk/) (1996-2012) and The National Archives websites (http://discovery.nationalarchives.gov.uk/SearchUI/Details?uri=C11521416) (1912-1995) in October 2013. Multinomial regression analysis of meningococcal disease-causing lineage on patient age was carried out with lineages modeled as categorical variables (with 'lineage 3 (cc41/44)' the baseline outcome) in the 'nnet' package with the function 'multinom' (Venables and Ripley 2002). Patient age was provided by the MRU and was grouped into four categories to reflect the peak incidence in young children and adolescents and low incidence in the elderly. For comparison of antigenic variation, *p*-distances (the proportion of amino acids at which two compared sequences differ) of deduced peptide sequences were calculated in MEGA6, with complete deletion of alignment gaps

Results

Genomic diversity of disease-causing meningococci

The 899 assemblies each contained an average of 209 contigs and 1,571 (97.86%) completely sequenced cgMLST loci, corresponding to ~79% of 2,000 anticipated coding sequences (Table 1) (Bentley *et al.* 2007). Additional assembly data is available in the MRF-MGL. There were two (NEIS2089 and NEIS0627, hypothetical proteins) to 313 (NEIS0829, putative membrane protein) alleles at each locus.

Table 1. Summary of automated Velvet/VelvetOptimiser genome assemblies.

Metric	Average (median)	Standard deviation
N50 contig number+	18.60	3.35
N50 contig length (L50) (bp)+	38,387	6,408
Assembly mean contig length (bp)	10,657	1,727
Largest contig length (bp)	121,962 (115,208)	37,572
Number of contigs	209	39
Number of contigs >1000 bp	114	17
Total assembled bases	2,166,197	44,680
Total assembled bases in contigs >1000 bp	2,128,374	40,019
Velvet final assembly k-mer size	85	5
Approx. number complete cgCDS†	1,571	10
%GC	51.60	0.15
Estimated sequencing coverage*	172.64	97.38
Estimated contig coverage‡	0.99	0.02

+N50 contig number: number of contigs collectively covering at least 50% of the assembly. N50 contig length (L50): ≥50% of the genome present on contigs of ≥Nbp. †cgCDS: 1,605 core-genome (cg) coding sequences (CDS) present in >95% of meningococcal genomes (Bratcher *et al.* 2014). *Estimated sequencing coverage: number of bases in sequence reads as a proportion of final assembly. ‡Estimated contig coverage: number of assembled bases as a proportion of FAM18 reference genome (2,194,961bp) (Bentley *et al.* 2007).

The genomes comprised 272 STs of which 219 were assigned to 20 MLST-defined ccs, and 498 rSTs, with 56 from isolates not designated to a known cc (Table 2). Meningococci belonging to cc41/44, cc269, and cc23 were responsible for 528 (58.73%) CCCs and there were nine 'rare' ccs (those with <10 members). Clonal complexes varied in their diversity, with the least diverse (cc1157) exhibiting an average of 3.47 rMLST locus differences between any two isolates, and the most diverse (cc41/44) exhibiting an average of 11.31 differences (Fig. 2A; Table 2).

Capsule groups were deduced from WGS-derived sequence data, providing a genogroup for each isolate. These were concordant with PHE-MRU serogroup designations and PCR-derived genogroups (Gray *et al.* 2006). WGS data provided additional discrimination: (i) of two isolates that were non-groupable (NG) by routine characterisation, one possessed a capsule null locus (*cnl*), and one had capsule group E loci; (ii) capsule loci of six serologically NG isolates were phase variable-off; (iii) four isolates were assigned to serogroup and genogroup W/Y (Tsang *et al.* 2008) since they possessed serine residues at codon 310 of the capsule polymerase, resulting in mixed galactose/glucose-containing sialic acid capsules and ambiguity in serological assays; and (iv) two serogroup/genogroup B isolates had six of the seven capsule group E synthesis genes, in addition to all capsule group B genes, but neither showed evidence of serogroup E expression.

Most isolates were genogroup B (668, 74.30%), followed by Y (141, 15.68%), W (56, 6.23%), and C (27, 3.00%), with an additional three genogroup E isolates, and single genogroup *cnl*, A, and X isolates. Serogroup characteristics of CCC isolates were similar to those of NCCC specimens, with a slight over-representation of group Y (6.23%, p-value = <0.0001, χ^2) and group W (2.44% p-value = <0.0001, χ^2) in the CCCs (Fig. 1A). Specimens from older patients were also present at slightly higher proportions among the CCCs than the NCCCs, and both of these differences were consistent with known differences in clinical presentation and diagnosis among demographic groups (Ladhani *et al.* 2012a). The serogroup differences, and the proportion of laboratory-confirmed cases confirmed by culture, did not alter significantly between the two epidemiological years (Table 3).

Table 2: Frequency and diversity of meningococcal disease-causing lineages, MRF-MGL 2010/11 and 2011/12.

Clonal complex	# (%) 2010/11	# (%) 2011/12	# (%) Total	Genogroup (% if <100%)	# unique STs (CC)	# unique rSTs (CC)	# incongruent rSTs (lineage)	Lineage	# STs (lineage)	# rSTs (lineage)	Mean pairwise rST distance (SD)
cc41/44	134 (26.75)	103 (25.88)	237 (26.36)	B (99.2), C (0.8)	81	156	0	3	84	158	11.31 (4.93)
cc269	98 (19.6)	73 (18.34)	171 (19.02)	B (98.8), C (1.2)	29	89	19 (2)*	2	50	110	11.09 (6.56)
cc23	60 (11.98)	60 (15.08)	120 (13.35)	Y	14	21	0	23	14	21	4.76 (3.29)
cc213	38 (7.58)	37 (9.30)	75 (8.34)	B	21	53	0	13	21	53	7.48 (2.88)
cc11	21 (4.19)	38 (9.55)	59 (6.56)	W (59.3), C (25.4), B (15.3)	5	16	0	11	5	16	7.27 (3.93)
cc32	28 (5.59)	14 (3.52)	42 (4.67)	B	14	23	0	5	16	24	4.62 (2.49)
cc22	13 (2.59)	11 (2.76)	24 (2.67)	W (83.0), Y (17.0)	8	15	0	22	9	16	4.95 (2.95)
cc60	15 (2.99)	5 (1.26)	20 (2.22)	B (85.0), E (15.0)	11	20	0	6	13	22	7.91 (3.43)
cc162	6 (1.20)	9 (2.26)	15 (1.67)	B	1	4	0	25	1	4	3.66 (1.75)
cc461	10 (2.00)	4 (1.01)	14 (1.56)	B	3	9	0	39	3	9	4.02 (1.99)
cc174	8 (1.60)	2 (0.50)	10 (1.11)	Y (80.0), W (10.0)	2	5	0	14	2	5	5.19 (4.80)
cc18	5 (1.00)	4 (1.01)	9 (1.00)	B	6	8	0	18	7	9	9.67 (5.68)
cc35	7 (1.40)	2 (0.50)	9 (1.00)	B	5	7	0	35	6	8	8.54 (4.35)
cc103	5 (1.00)	3 (0.75)	8 (0.89)	C (62.5), B (25.0), Y (12.5)	4	5	1 (26)	20	4	4	5.67 (1.85)
cc1157	3 (0.60)	4 (1.01)	7 (0.78)	B	5	7	0	15	5	7	3.47 (1.99)

cc167	4 (0.80)	2 (0.50)	6 (0.67)	Y	4	4	0	26	5	5	10.09 (4.54)
cc282	5 (1.00)	1 (0.25)	6 (0.67)	B	3	3	0	32	3	3	7.51 (3.44)
cc226	1 (0.20)	0	1 (0.11)	cnl	1	1	0	29	1	1	na
cc5	1 (0.20)	0	1 (0.11)	A	1	1	0	10	1	1	na
cc865	0	1 (0.25)	1 (0.11)	B	1	1	0	41	1	1	na
ccND	39 (7.78)	25 (6.28)	64 (7.12)	B (92.2), Y (3.1), C (3.1), X (1.6)	53	56	0	NA	21	21	na
Total	**501 (100.00)**	**398 (100.00)**	**899 (100.00)**	**B (74.21); Y (15.71); W (6.07); C (3.14); E (0.33); cnl (0.22); A, W/Y, X (0.11)**	**272**	**498**	**20**		**272**	**498**	**26.82 (8.45)**

Unique rSTs: six rSTs are found among both ccND isolates and isolates designated to a clonal complex and are counted once in the column total. Incongruent rSTs: rSTs that do not cluster into a lineage with other rSTs from isolates of the same clonal complex. *cgMLST clustered isolates possessing these rSTs into a lineage with other isolates of cc269. Mean pairwise rST distance: the number of rMLST loci at which there is an allelic difference in a comparison between an rST pair (single occurrences of rSTs used to reduce skew). na: not applicable.

Table 3. Characteristics of the 899 meningococcal isolates in the MRF-MGL for the epidemiological years 2010/11 and 2011/12. The changes in total number of isolates, lineages, age-groups, and serogroups with most variation between years, are shown.

Characteristic	Number of isolates 2010/11 (%)	Number of isolates 2011/12 (%)	Ratio 2011/12:2010/2011	Relative risk ratio (CI)	Chi square p-value
Total MRF-MGL*	501 (100)	398 (100)	0.79	1.09 (1.00, 1.20) †	0.723
Lineage 5 isolates (cc32)	28 (5.59)	14 (3.52)	0.50	0.63 (0.34, 1.18) ‡	0.1927
Lineage 6 isolates (cc60)	15 (2.99)	5 (1.26)	0.33	0.42 (0.15,1.14) ‡	0.1267
Lineage 23 isolates (cc23)	60 (11.98)	60 (15.08)	1.00	1.26 (0.90, 1.76) ‡	0.2082
Lineage 11 isolates (cc11)	21 (4.19)	38 (9.55)	1.81	2.28 (1.36, 3.82) ‡	0.002031
Isolates from patients under 5 yrs of age	257 (51.30)	171 (42.96)	0.67	0.84 (0.73, 0.97) ‡	0.01563
Isolates from patients 16-18 yrs of age	18 (3.59)	27 (6.78)	1.50	1.89 (1.06, 3.38) ‡	0.04282
Genogroup Y isolates	74 (14.77)	67 (16.83)	0.91	1.14 (0.84, 1.54) ‡	0.4515
Genogroup C isolates	13 (2.59)	14 (3.52)	1.08	1.36 (0.64, 2.85) ‡	0.5429
Genogroup W isolates	26 (5.19)	30 (7.54)	1.15	1.45 (0.87, 2.41) ‡	0.1909

*The total number of laboratory-confirmed cases was 1,053 in 2010/2011 and 766 in 2011/12: this corresponds to >98% all reported cases, and comprised 899 culture confirmed cases which were included in the MRF-MGL, and 920 non-culture confirmed cases (PHE 2014). The relative disease incidence ratio 2011/12:2010/11 was 0.72 (CI 0.66, 0.79; p-value<0.0001, population sizes from ONS). Relative risk ratio: ratio of 2011/12 to 2010/11 isolate counts as a proportion of: all laboratory-confirmed cases (†), culture-confirmed cases/MRF-MGL (‡).

Genealogical relationships among isolates

rMLST analyses provided high resolution comparisons among all isolates and enabled visualisation of lineages and structure within them (Fig. 2A). To take advantage of this resolution and facilitate comparison of WGS surveillance data with MLST designations, a unified lineage nomenclature was implemented (Table 2). Where possible, lineages were assigned a number previously associated with the meningococci constituting them: for example, the lineage containing cc41/44 isolates, previously identified as MLEE 'lineage 3', was assigned lineage 3 in this scheme.

rMLST clusters were congruent with isolate MLST clonal complexes, with the exception of the rST from isolate ID 21505 (cc103) and certain cc269-associated rSTs (Table 2; Fig. 2A): although the three lineage 2 (cc269) rMLST clusters (Fig. 2A) in general contained distinct STs, their relationships were inconsistent with MLST clusters (Table 4; Fig. 3). It appeared that rMLST cluster 3 had undergone recombination in ten (20.41%) ribosomal loci (Fig. 4) with the result that rMLST clusters 1 and 2 were more similar to lineage 5 (cc32) that they were to cluster 3 (Fig. 2A). However, analysis of cc269 and cc32 isolates at the 1,605 locus cgMLST level (Fig. 2B) confirmed that lineage 2 was indeed a separate lineage, consistent with MLST (Fig. 3). Two predominant sub-lineages were visible: 2.1, composed of rMLST clusters 1 and 3; and 2.2, composed of rMLST cluster 2 plus 29 isolates (3.23% MRF-MGL) not designated to a known clonal complex (Table 4). A putative third sub-lineage, 2.3, was rare in the MRF-MGL but was prevalent in PubMLST. Only 261 (16.26%) cgMLST loci distinguished sub-lineages 2.1 and 2.2, compared to 451 (28.10%) distinguishing lineages 2 and 5.

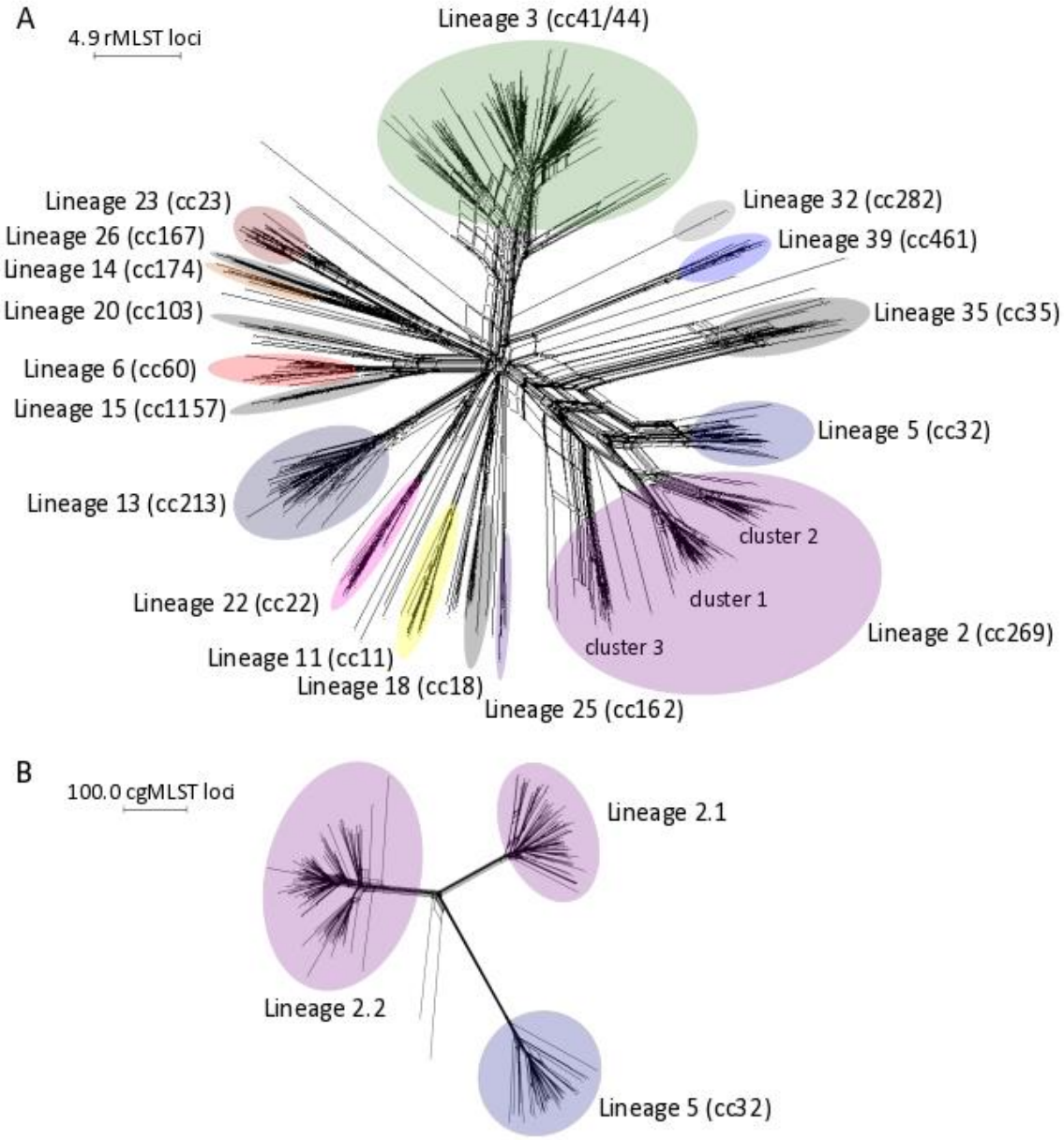


Figure 2. Structure of disease-causing meningococcal population in England and Wales, 2010/11-2011/12. (A) Neighbor-Net diagram generated from unique ribosomal multilocus sequence typing (rMLST) sequence types (rST) in MRF-MGL (n=498). Seven-locus MLST clonal complexes (cc) causing <10 cases annotated in grey; ccs represented by single rSTs and STs not designated to a cc are not labeled. (B) Neighbor-Net diagram generated from 1,605 core-genome MLST (cgMLST) loci in lineage 2 (cc269) (n=171) and lineage 5 (cc32) isolates (n=42).

Table 4. Association of cc269 MLST sequence types (STs) with lineage 2 sub-lineages.

Lineage 2 sub-lineage	MLST ST	rMLST cluster	MLST ST	rMLST cluster
2.1	269	1 (+ single rST in 3)	9836	1
	1049	1	9840	1
	1195	1	9843	1
	1092	1	467	3
	1942	1	479	3
	2873	1	283	3
	7226	1	1774	3
	7939	1	10264	3
	9823	1	10291	3
2.2	1161	2	5335*	2
	275	2	6781*	2
	1163	2	7143*	2
	4713*	2	9839*	2
	1159*	2	9880*	2
	6604	2	9837*	2
	5849*	2	3934*	2
	1831*	2	2307*	2
	4401	2	9827*	2
	6428	2	9887*	2
	7789	2	9845*	2
	7833	2	10265*	2
	9004	2	10290*	2
	9826	2	10263*	2
	9829	2	10297*	2
	10288	2	10283*	3
Putative 2.3	13	NA		
	9311*	2		

* indicates sequence types not designated to MLST-defined clonal complexes, but which would be part of cc269/275 if the second central genotype were incorporated. NA: not applicable.

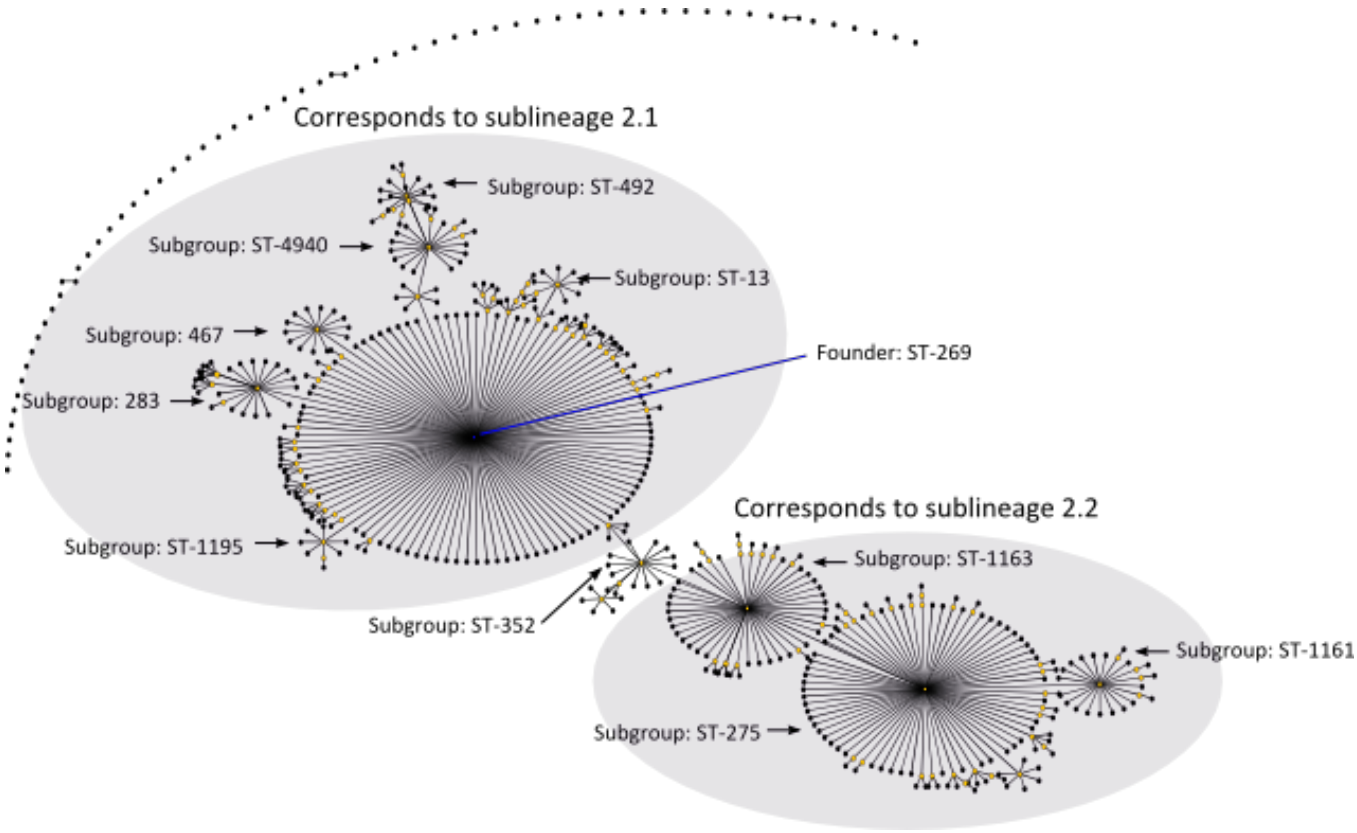


Figure 3. eBURST of MLST sequence types (ST) associated with lineage 2. All cc269-designated STs in PubMLST, and STs not designated to MLST-defined clonal complexes that would be part of cc269/275, were included in the eBurst diagram (Feil *et al.* 2004). STs are represented as coloured circles: the founding ST is blue, subgroup founding STs are yellow, and other STs are black. Each ST was included once, and sub-groups were labeled according to their founders if the founder was prevalent in http://pubmlst.org/neisseria. The approximate clustering of lineage 2.1 and 2.2 isolates as identified in Fig. 2B is indicated with purple bubbles.

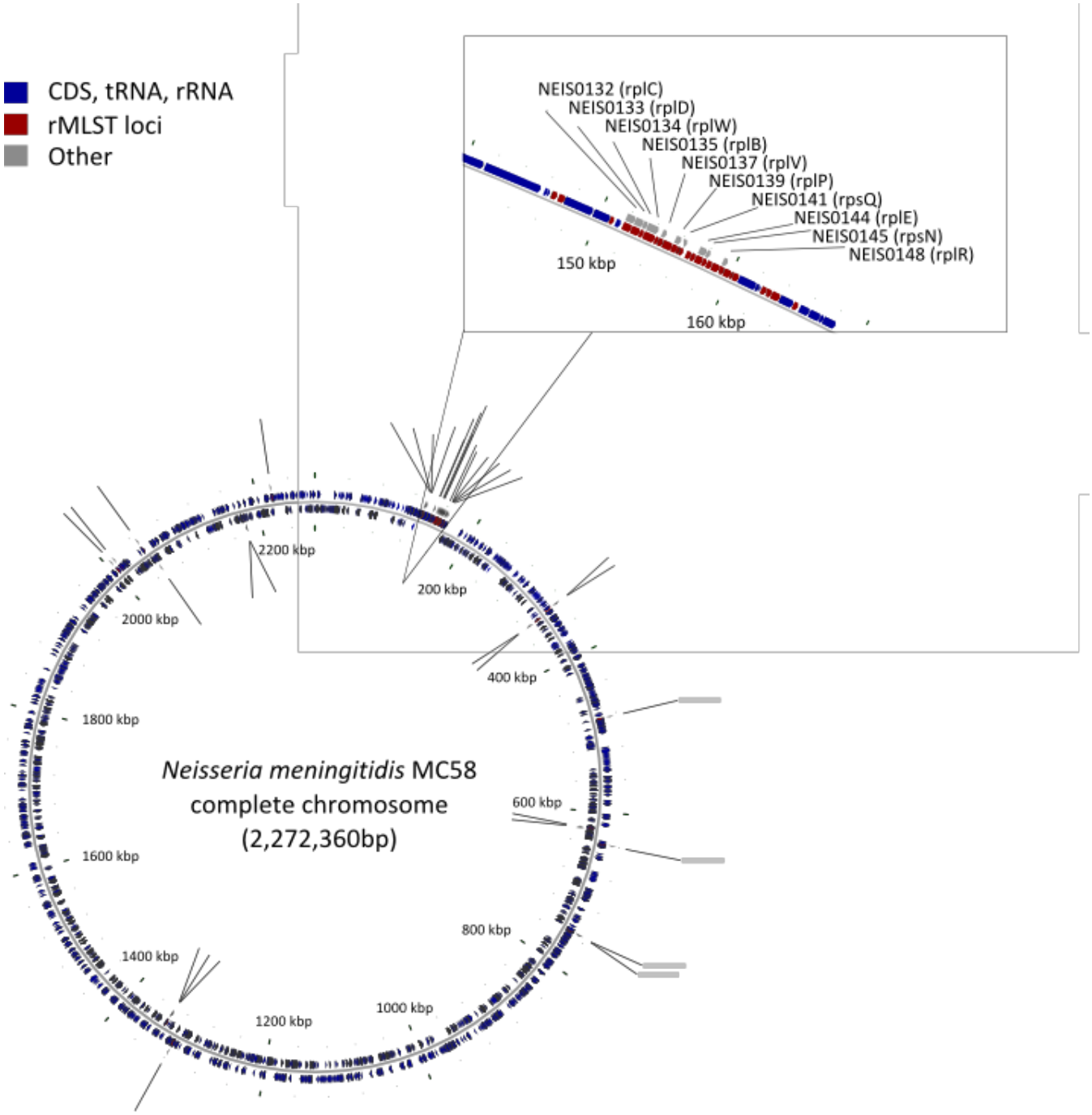
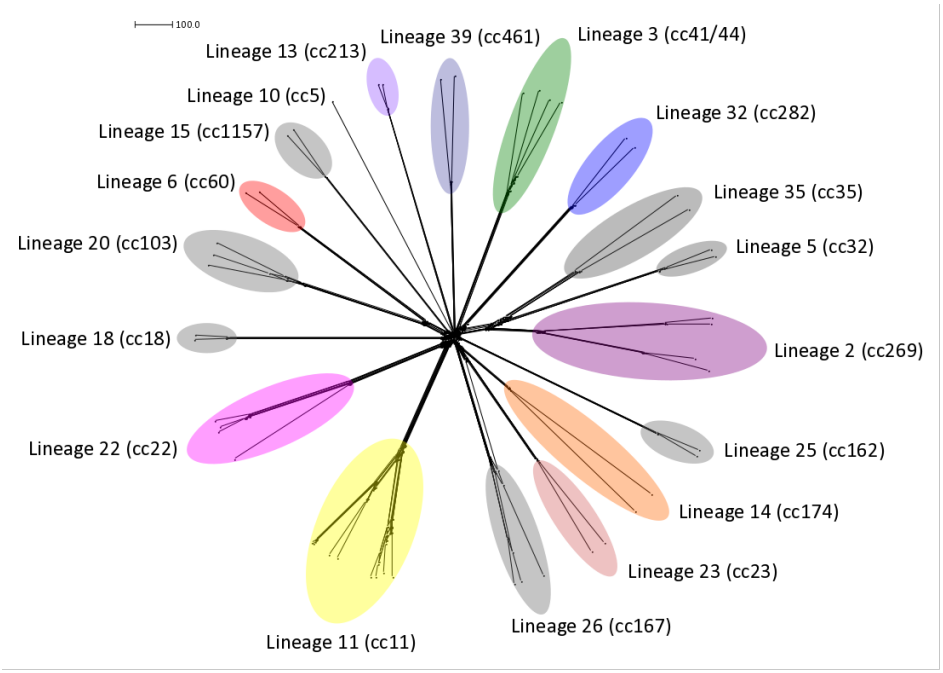


Figure 4. Ribosomal MLST loci responsible for lineage 2 rMLST versus MLST/cgMLST incongruence. *Neisseria meningitidis* genes are indicated by triangles on the MC58 chromosome, with gene product indicated by colour. *N. meningitidis* rMLST loci are additionally indicated by grey tick marks. Zoom panel: rMLST loci that distinguish rMLST sub-cluster 3 from rMLST sub-clusters 1 and 2 are annotated by name on grey tick marks. This figure was generated with CGView (Stothard and Wishart 2005).

Networks and maximum-likelihood trees generated from cgMLST alleles and variable-sites identified lineages and sub-lineages analogous to those identified using the gene-by-gene approach (Figs. 5-6). Clusters in an rMLST variable-site maximum-likelihood tree (Fig. 7) were analogous to those in the rMLST network (Fig. 2A), except lineage 3 (cc41/44) was composed of additional clusters, indicating conflicting phylogenetic signal in lineage 3 ribosomal loci.

rSTs were manually linked to lineage and sub-lineage designations in PubMLST to ensure consistent future utilisation of the nomenclature: where an rST clustered within rSTs associated with a clonal complex, it was assigned to the lineage corresponding to that clonal complex; rSTs from isolates not designated to known clonal complexes and not clustering within a clonal complex group were not assigned to lineages. Across the whole dataset, isolates from the same rMLST cluster had a 97% chance of being placed in the same clonal complex by MLST, whereas isolates of the same clonal complex had an 81% chance of belonging to the same rMLST cluster, reflecting the additional sub-division of cc269 by rMLST ($AW_{rMLST->MLST}$ = 0.97 [0.91-0.99], $AW_{MLST->rMLST}$ = 0.81 [0.80-0.82]). Isolates of the same clonal complex had a 96.7% (95% CI: 96.4-97.0) chance of being placed in the same lineage once lineage 2 rSTs were assigned to the consensus cgMLST sub-lineages. To date, sub-lineages have been defined by visual inspection and comparison of phenotypes, such as the 'ET-15' and 'non-ET-15' sub-clusters of cc11 (lineage 11), named lineage 11.2 and lineage 11.1 respectively (Jolley *et al.* 2012b, Harrison *et al.* 2015, Lucidarme *et al.* 2015, Toros *et al.* 2015); ultimately, WGS data may enable formal definition of the high resolution population structure of the meningococcus. The current lineage nomenclature is searchable within the PubMLST.org/neisseria database and the MRF-MGL where it can be compared with cc designations.

(A)



(B)

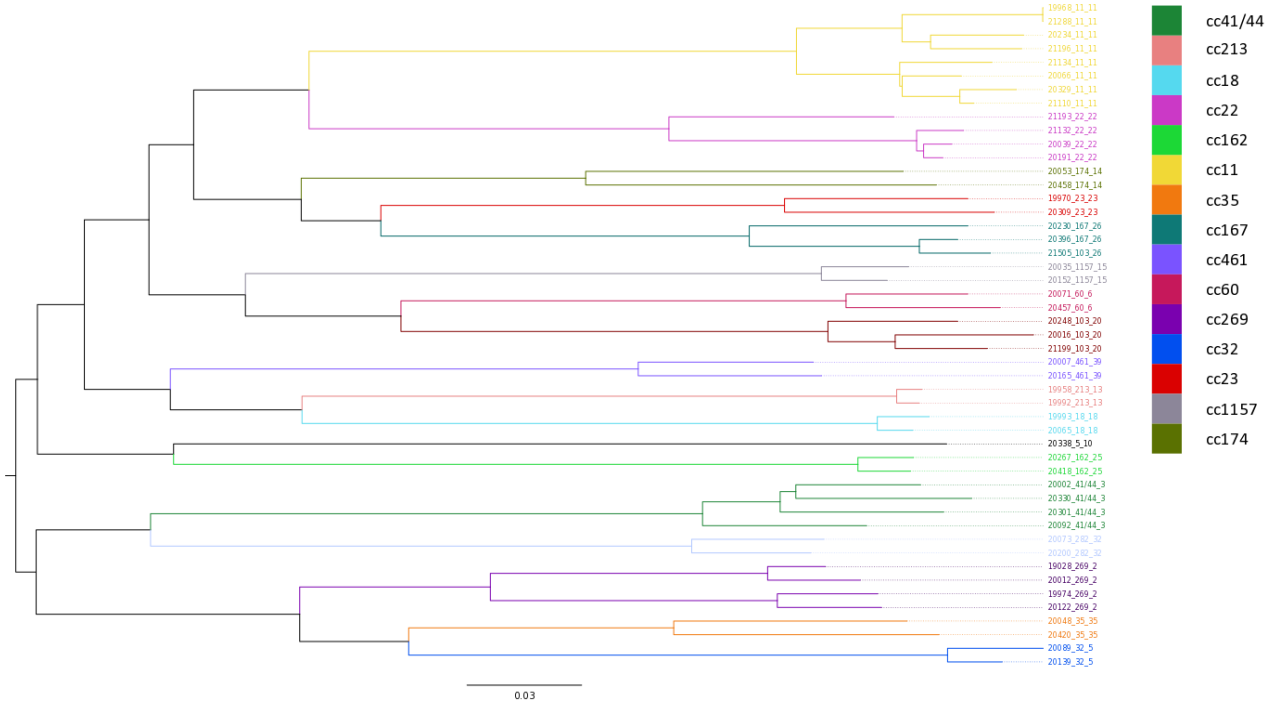


Figure 5. MRF-MGL lineages identified using core-genome loci. A subset of isolates (n=50) representing MRF-MGL lineages was extracted to demonstrate that core-genome lineages, inferred using allele or nucleotide data, are analogous to those in Fig. 2. (A) Neighbor-Net graph generated from allelic distances among core-genome loci. (B) Maximum-likelihood tree generated from variable-sites (n=104,993) in the core-genome MLST locus nucleotide alignment.

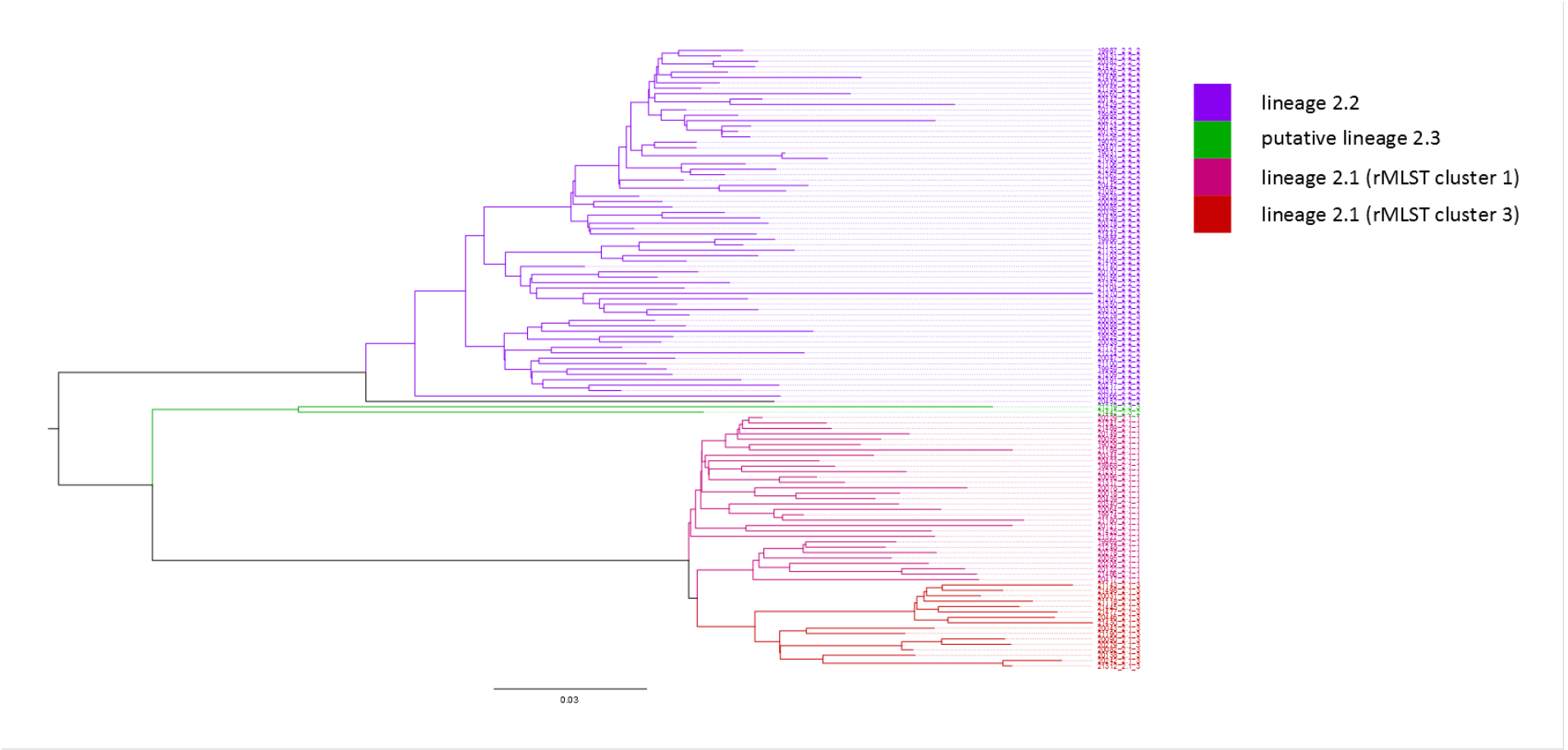


Figure 6. Lineage 2 core-genome maximum-likelihood tree. A subset of lineage 2 (cc269) isolates (n=116) representing each lineage 2 sub-lineage in the MRF-MGL demonstrates that sub-lineages identified using the gene-by-gene approach (Fig. 2B) are congruent to those in a maximum-likelihood tree. Variable-sites (n=79,316) were extracted from the concatenated alignment of core-genome loci. As in Fig. 2B, lineage 2.1 is composed of rMLST clusters 1 and 3 (Fig. 2A, Table 4) due to recombination in the rMLST loci (Fig. 4).

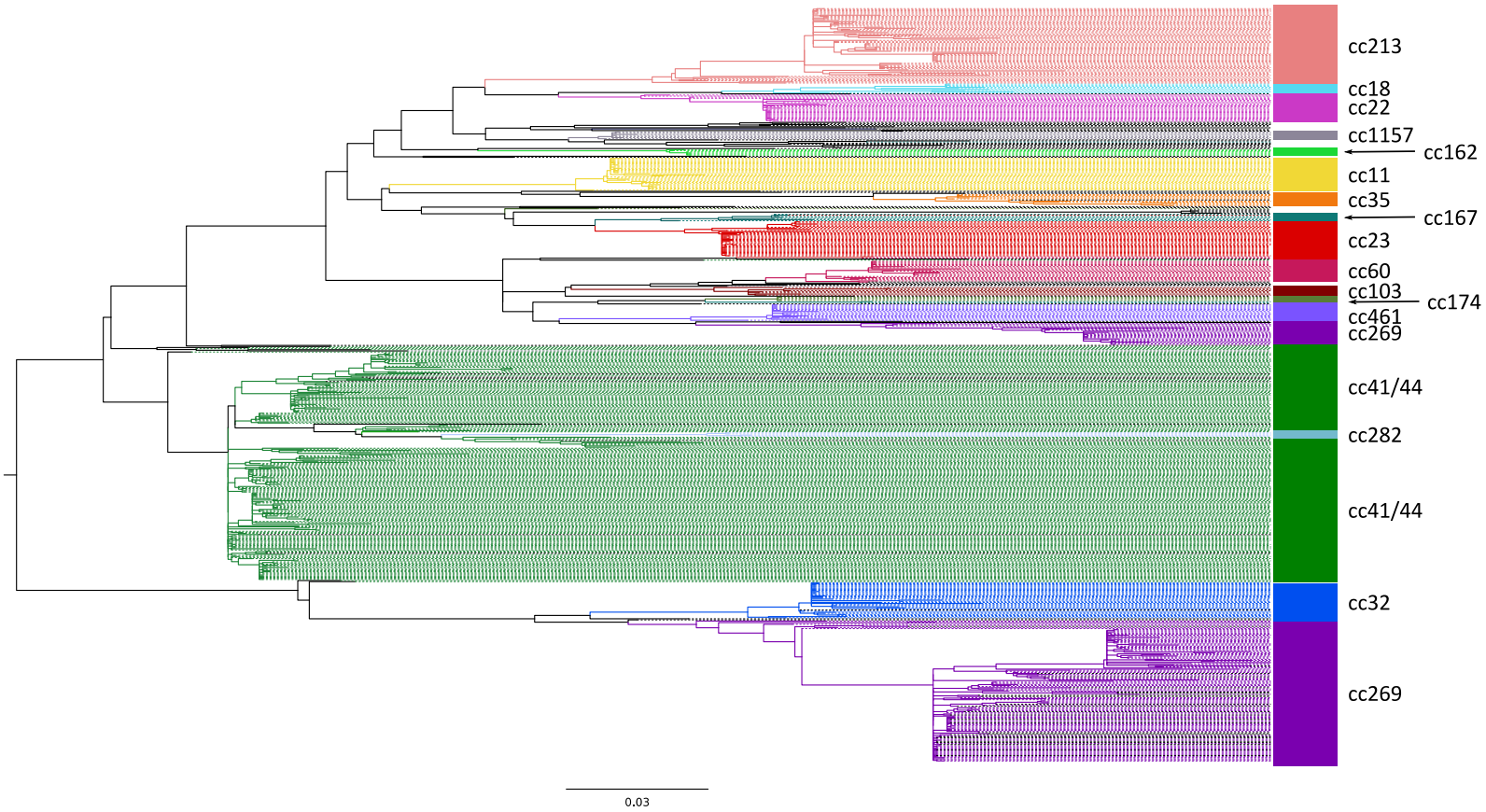


Figure 7. rMLST variable-site maximum-likelihood tree. Nucleotide alignments of concatenated non-paralagous rMLST loci from all unique rSTs (n=498) in the MRF-MGL were extracted using Genome Comparator. A maximum-likelihood tree was drawn from all variable-sites (n=1,373). Tips and branches were coloured by the corresponding clonal complex.

Annual fluctuations in meningococcal disease characteristics

Culture-confirmed cases in England and Wales declined by 20.56% in 2011/12, reflecting a decline to the lowest incidence of recent years (Fig. 1B; Table 3; Fig. 8A). Genogroup was more diverse among 2010/11 than 2011/12 (groups B, C, E, Y, and W only) isolates. No individual lineage was significantly associated with the decline in cases (Table 2): the largest prevalence decreases were observed in lineage 5 (cc32) and lineage 6 (cc60), which dropped by half and two-thirds respectively (Table 3). The prevalence of lineage 23 (cc23) and lineage 11 (cc11) increased in 2011/12, however, with five of six lineage 11-associated STs causing additional cases and ST-11 from lineages 11.1 and 11.2 responsible for the biggest increase (Table 3). There was a concomitant increase in genogroup Y, C, and W cases (Table 3).

Nearly one quarter (22.58%) of culture-confirmed cases were from individuals aged less than one year, with a maximum in patients aged 4-6 months (7.71% of yearly cases). After one year of age, cases declined in number before a marked increase in those aged 16-to-24 years (11.07% of cases); there was a small increase in meningococcal disease patients aged over 39 years (28.73% of isolates) with a minor peak in those 82-84 years old. The age characteristics of MRF-MGL isolates reflected data from NCCCs, except, isolates from individuals aged over 30 years comprised a slightly greater proportion of CCCs (an additional 0.01-1.07% per two-year age-group) (PHE 2014). The disease age distribution altered in 2011/12 (Table 3) with a marked reduction in cases from those aged under five years but a significant increase in cases from those aged 16-18 years (including four additional cases each of lineages 23 and 3, and two additional cases each of lineages 11, 5, and 13).

Association of lineage with patient age

There was a strong association between patient age and the lineage of the disease-causing meningococcus (Fig. 8B). The proportion of disease caused by lineages 2 and 3 declined with age (Fig. 8B). Patients aged four years and under were less likely to have been infected with lineage 13 relative to lineage 3 disease (relative risk ratio (RRR) 0.32 [0.23, 0.45]), and after the age of four,

patients were two orders of magnitude more likely to have had lineage 23 disease, with the RRR of lineage 23 to lineage 3 at 27.73 [12.91, 59.56] in the over 25s (Table 5). After 14 years of age disease caused by lineage 11 increased significantly, with the over 25s most likely to have been infected with lineage 11 (RRR = 7.17 [3.53, 14.57]). Disease caused by other meningococcal lineages was greatest in those aged under five years and over thirty years, although the former were still most likely to have had lineage 3 disease (Table 5). Additional analyses with alternative cut-offs/bands for patient age, which retained the relative peaks in incidence (e.g. <2, 2-12, 13-28, >28 (Bille *et al.* 2008)), produced the same general trends, with some apparently non-linear associations.

Patients at the extremes of age were affected by a greater range of meningococcal genotypes, although four of the least diverse lineages (lineages 39, 25, 22, and 23) were common (Table 2) and were frequently isolated from patients aged less than five or more than 50 years (Fig. 8). The association of genogroup with patient age was consistent with the association of genogroup with lineage, for example, genogroup C meningococci were more frequently identified in two year age-groups between 25 and 49 years (age-group average 2.2 isolates, 11.09%) than from any other (age-group average 1.5 isolates, 5.43%) (p-value <0.0001, Fisher's exact test).

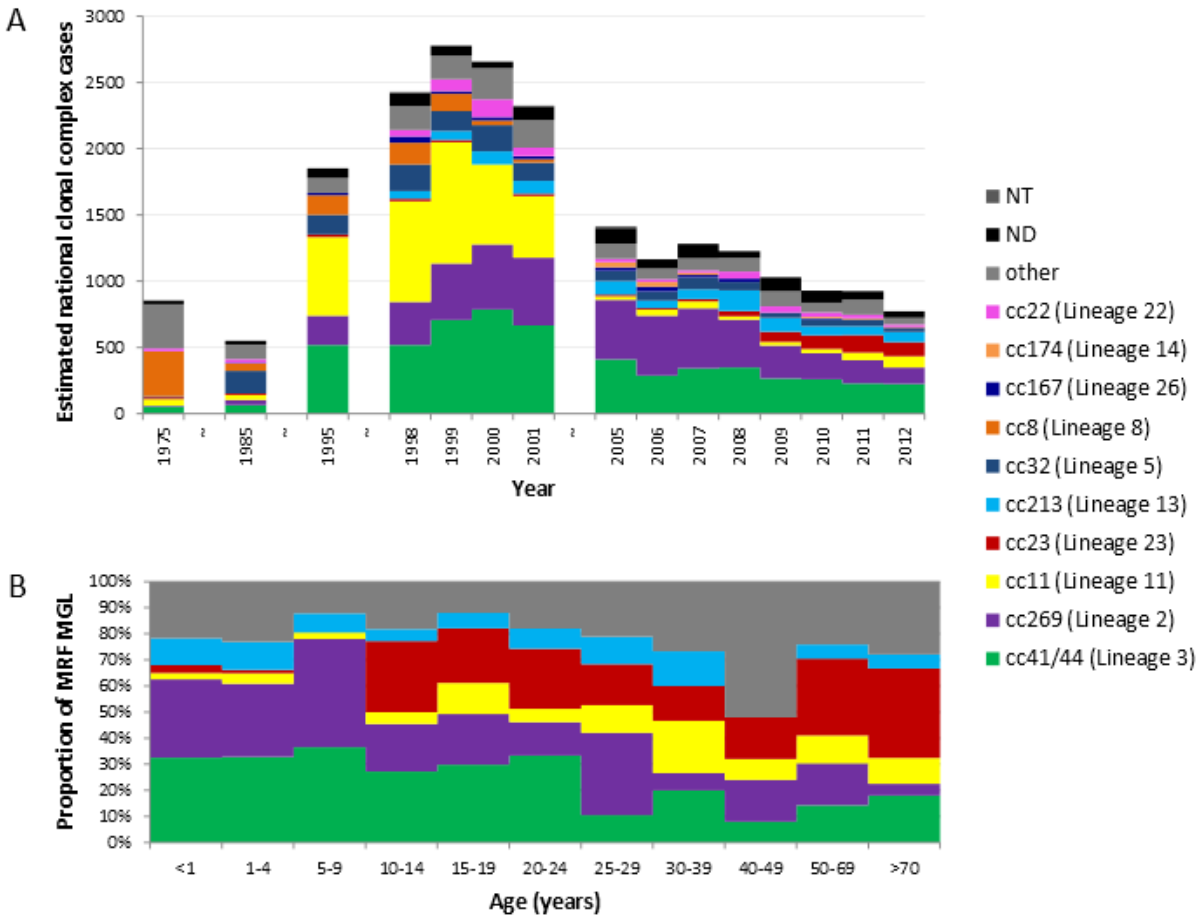


Figure 8. Prevalence of clonal complexes and age-association of lineages in England and Wales. (A) Yearly frequencies of clonal complexes estimated from proportions in available MRU disease isolate MLST data (Russell *et al.* 2008) multiplied by disease notification data (pre-1998 supplied by Mary Ramsay; 1998-2012 from http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1317136087786). ND: ST not designated to a clonal complex. NT: MLST not performed. (B) Proportion of MRF-MGL cases in each age-group by lineage.

Table 5. Multinomial regression of meningococcal lineage against patient age-group, with a baseline outcome of lineage 3 (cc41/44).

Lineage	Age (years)	# isolates	Relative risk ratio (RRR)	95% confidence intervals
2	≤4	124	0.89	0.70, 1.13
2	5-14	21	1.13	0.59, 2.17
2	15-24	18	0.62	0.33, 1.15
2	≥25	35	0.86	0.52, 1.42
23	**≤4**	**9**	**0.06**	**0.03, 0.13**
23	**5-14**	**6**	**4.44**	**1.44, 13.76**
23	**15-24**	**23**	**10.84**	**4.59, 25.59**
23	**≥25**	**82**	**27.73**	**12.91, 59.56**
13	**≤4**	**45**	**0.32**	**0.23, 0.45**
13	5-14	4	0.59	0.19, 1.82
13	15-24	7	0.66	0.27, 1.59
13	≥25	18	1.22	0.64, 2.31
11	**≤4**	**14**	**0.10**	**0.06, 0.17**
11	5-14	2	0.95	0.20, 4.49
11	**15-24**	**10**	**3.03**	**1.24, 7.42**
11	**≥25**	**33**	**7.17**	**3.53, 14.57**
other	**≤4**	**96**	**0.69**	**0.53, 0.89**
other	5-14	9	0.62	0.27, 1.42
other	15-24	15	0.66	0.34, 1.29
other	**≥25**	**83**	**2.63**	**1.69, 4.10**

Significant relative risk ratios in bold.

Vaccine antigens

Bexsero® vaccine antigen peptide sequences (fHBP, NHBA, PorA, and NadA) (Giuliani *et al.* 2006, Serruto *et al.* 2012, Bambini *et al.* 2014) were diverse within the MRF-MGL. Despite possessing the greatest numbers of rare peptide sub-variants (fewer than 10 occurances), fHBP and NHBA were relatively conserved at the peptide sequence level, with average *p*-distances of 0.20 (80% amino-acid identity) and 0.13 (87% identity) respectively (Table 6; Fig. 9). Consistent with previous studies, PorA VR2 was highly variable (38% amino acid identity) with 55 different peptide sub-variants identified, and NadA was absent in 797 isolates (88.65% of MRF-MGL) (Table 6; Fig. 9). Vaccine antigen variants were associated with particular lineages; for example, Bexsero® fHBP peptide 1 (variant 1.1) was found mainly in lineage 5 (cc32) isolates (Fig. 9) (Serruto *et al.* 2012). Exact amino-acid sequence matches to any Bexsero® vaccine antigen variant occurred in 259 (28.81%) isolates from eight lineages. Among these, 247 of the 668 genogroup B isolates (36.98%) contained at least one exact match to a Bexsero® antigen variant. Lineage 3 was most likely to contain an exact match to any Bexsero® antigen variant (203 isolates), and only one lineage 2 isolate contained an exact match. Almost one third of isolates from patients younger than one year (62 of 203 isolates) contained exact peptide matches to any variant, consistent with the association of lineage 3 with this age-group (Fig. 8B).

Table 6. Prevalence and diversity of vaccine antigen peptide variants in England and Wales, 2010/11-2011/12.

Antigen	# unique peptide sub-variants		Average amino-acid p-distance (SD)	Most prevalent variant MRF-MGL			Bexsero® sub-variant		
	prevalent*	rare*		sub-variant	frequency MRF-MGL	frequency genogroup B	sub-variant	frequency MRF-MGL	frequency genogroup B
fHbp	12	87	0.20 (0.13)	4	159 (17.69%)	159 (23.80%)	1.1	34 (3.78 %)	32 (4.79%)
NadA	3	6	0.21 (0.20)	No value†	797 (88.65%)	530 (79.34%)	8	12 (1.33%)	2 (0.30%)
NHBA	14	70	0.13 (0.04)	2	172 (19.13%)	172 (25.75%)	2	172 (19.13%)	172 (25.75%)
PorA VR2	15	40	0.62 (0.29)	1.9	137 (15.24%)	133 (19.91%)	1.4	136 (15.13)	136 (20.4%)

*Prevalent and rare variants: present in greater than ten, or fewer than ten, isolates, respectively. †No value: isolates missing gene; gene contains frame-shift mutation; gene interrupted by the end of a contig.

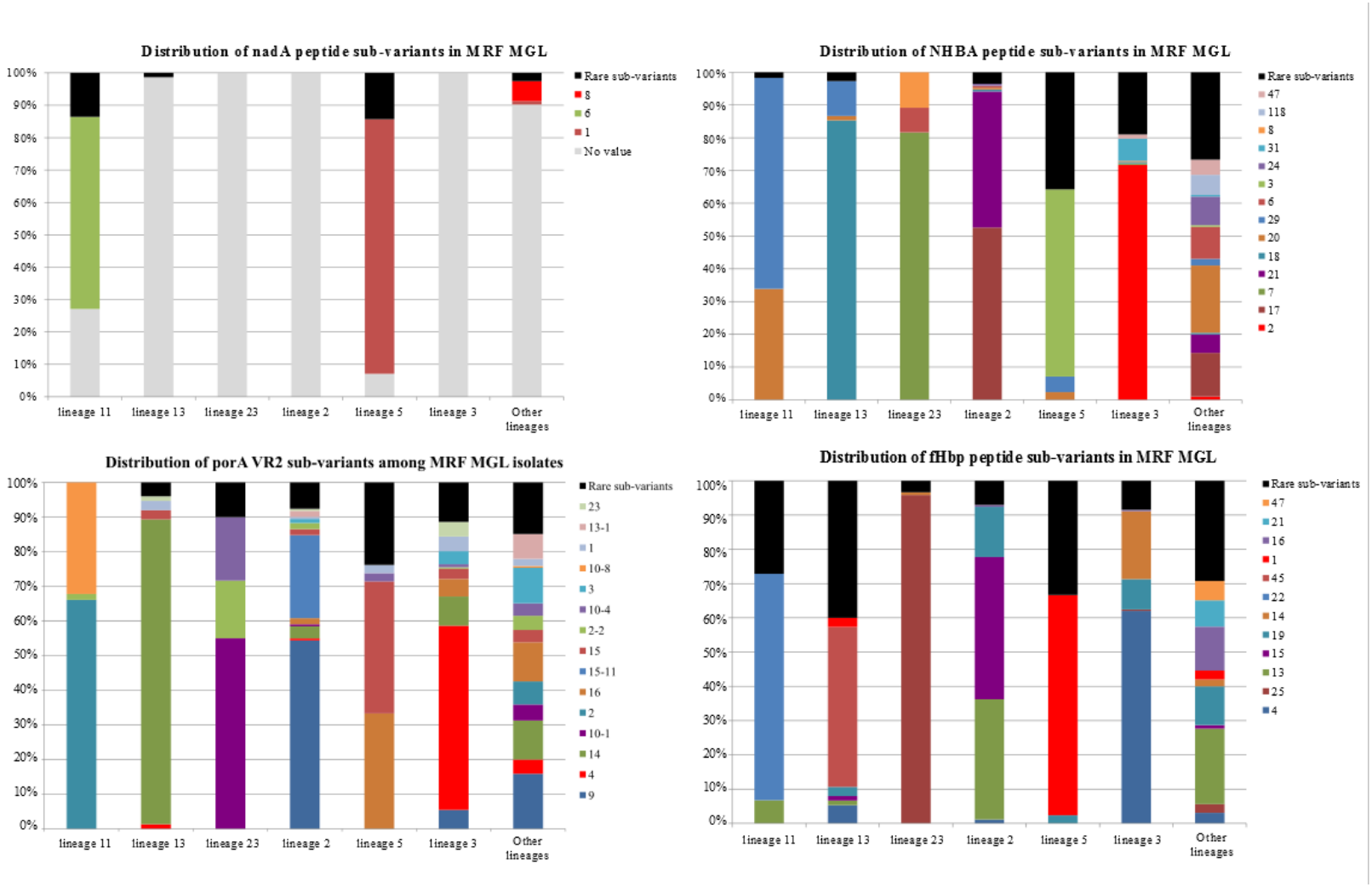


Figure 9. Prevalence of vaccine antigen peptide variants in meningococcal disease-causing lineages in England and Wales, 2010/11-2011/12. The variant of fHBP, NHBA, NadA, and PorA VR2 included in Bexsero® (GSK) is coloured red. 'Rare' sub-variants are those with <10 occurrences. 'No value' refers to absence of the peptide (the result of missing, frame-shifted, or incompletely assembled antigen genes).

Discussion

NGS technology permits genetic characterisation of bacterial isolates at >95% of their genome, enabling unification of clinical, epidemiological, and microbiological studies (Bratcher *et al.* 2014). To impact patient care, however, these data have to be provided in an accessible format that is readily interpreted by medical and public health practitioners. The MRF-MGL achieves this for a coherent national collection of meningococcal disease isolates by presenting WGS data within an open-access web-interface, with inbuilt reporting, analysis, and export options. It employs the PubMLST.org infrastructure to consolidate and extend the comprehensive case ascertainment data and specimen collections available in England and Wales, where virtually all meningococcal disease isolates are referred to the PHE-MRU (Jolley and Maiden 2010, Maiden *et al.* 2013). These data are of particular value with imminent changes in the UK meningococcal vaccine schedule.

Consistent with previous studies, meningococcal disease isolates in the MRF-MGL were genetically and antigenically diverse, with evidence for the role of recombination in generating this diversity (Caugant and Maiden 2009, Hao *et al.* 2011) (Fig. 2; Fig. 4). Gene-by-gene analyses (Maiden *et al.* 2013) rapidly and effectively organised this diversity into coherent groups that were consistent with meningococcal clonal complexes (Maiden *et al.* 1998). As in previous studies, there was good agreement among different analysis approaches with at least 20 distinct meningococcal lineages present (Fig. 2; Figs. 5-7) (Budroni *et al.* 2011, Kohl *et al.* 2014).

The PubMLST.org/neissseria database enables the comparison of different nomenclatures. In common with other geographic regions, meningococcal disease incidence in England and Wales has fluctuated (Gray *et al.* 2006, Ladhani *et al.* 2012a), and, using the inbuilt analysis tools, the MRF-MGL data were readily placed in the context of previous studies. The largest meningococcal disease epidemics observed in England and Wales over the past 100 years (Fig. 1A) can be attributed to the disruption of the World Wars and were probably caused by single lineages. By contrast, the elevated disease that occurred in 1985-2005 (Fig. 1B) was due to concurrent and successive epidemics caused

by lineage 5, 11, 2, and 3 meningococci (Fig 8A) (Table 2). These observations support the view that recent epidemics are a consequence of interactions of meningococcal diversity, human social behaviour, and host immunity. Continued surveillance employing WGS data will further define meningococci responsible for incidence fluctuations and provide the genetic resolution necessary to detect outbreaks and future trends in disease incidence. This will provide the basis for the prediction of disease trends and development of targeted interventions.

The pronounced associations of certain lineages with disease in particular age-groups (Fig. 8B) were consistent with those reported for pan-European MLST data from the early 2000s (Brehony *et al.* 2014), with greater diversity in isolates from older and younger individuals perhaps due to lower immunity in these cohorts. Two major differences were evident, however, with fewer lineage 11 isolates and more lineage 23 isolates in the MRF-MGL. The difference in lineage 11 serogroup C prevalence is attributable to MCC vaccines, but the reasons for changes in the prevalence of lineage 23 serogroup Y meningococci principally affecting older individuals are unclear (Broker *et al.* 2012a, Ladhani *et al.* 2012b). Identification of such associations alongside WGS data presents the prospect of investigating the genomic basis of meningococcal phenotypes, exemplified by the overrepresentation of the meningococcal disease-associated phage in lineages with a propensity to cause disease in adolescents (Bille *et al.* 2005). It will be possible to undertake further studies of this type by comparing WGS data from invasive and carried meningococci.

An important advantage of WGS is the resolution of isolate relationships within lineages, such as the distinction of sub-lineages 11.1 and 11.2, indistinguishable by MLST. Such sub-lineages frequently have distinct epidemiological behaviour, with outbreaks and epidemics attributable to particular sub-lineages (Jolley *et al.* 2012b). As more WGS data become available, the role of sub-lineages will become increasingly apparent. For example, lineage 11.1 isolates were predominant among global meningococcal disease isolates in the mid-20th century, but from the mid-1990s to early 2000s serogroup C lineage 11.2 was the principal cause of serogroup C disease in many countries including the UK (Caugant 1998). This outbreak was successfully contained by herd immunity generated with

MCC polysaccharide vaccines (Maiden *et al.* 2002, Gray *et al.* 2006), but 10-14 years after implementation some lineage 11.2 serogroup C disease remains in the MRF-MGL, perhaps as a consequence of waning immunity in adolescents (Table 2, Fig. 8) (Pollard *et al.* 2013). Serogroup C lineage 11.2 meningococci have recently been associated with outbreaks in 'men who have sex with men' communities in Europe (Weiss and Varma 2013). Finally, serogroup W lineage 11.1 meningococci were introduced in England and Wales in the early 2000s as part of Hajj-associated outbreaks (Gray *et al.* 2006) but the recent increase in lineage 11.1 serogroup W meningococci (Fig. 8A, Table 3) is due to a closely related but distinct variant distinguished with WGS data (Lucidarme *et al.* 2015).

With no serogroup B conjugate polysaccharide vaccines, protection against serogroup B depends on protein-based vaccines, such as Bexsero®, which will be introduced in the UK immunisation program in 2015. This vaccine contains an outer-membrane vesicle from a lineage 3 meningococcus with three other antigens from organisms of other lineages (Giuliani *et al.* 2006). Peptide sequences of vaccine antigens are readily deduced within the MRF-MGL, enabling the potential coverage of actual and proposed vaccine formulations to be assessed. Exact peptide matches to Bexsero® components were rare in the MRF-MGL (Table 6; Fig. 9), especially in isolates from individuals aged less than one year, indicating the importance of immunological cross-reactivity among peptide variants for coverage by this vaccine. Cross-protection among fHbp family 1 sub-variants and among NadA-1, -2, and -3 sub-variants has been described, and were identified in 497 (55.28%) and 97 (10.79%) of MRF-MGL isolates respectively. Therefore, 567 (63.07%) of the MRF-MGL, 498 (74.55%) of genogroup B isolates, and all 35 of the lineage 11.1 genogroup W isolates, exhibited at least one peptide sequence match to one of the vaccine antigens. This coverage is similar to results predicted with the MATS assay (Vogel *et al.* 2013), but remains subject to a number of assumptions. Over time fluctuations in disease-associated meningococcal lineages and their antigen variants (Fig. 9) may warrant the development of alternative vaccine formulations, with WGS in surveillance providing the necessary data (Table 6).

Genomic analyses of meningococci from disease cases are increasingly feasible in real-time as technology develops: combined with the four-day turnaround for WGS determination, achievable at the time of writing, the gene-by-gene approach allows genome-level characterisation of multiple isolates and comparisons with reference databases within hours (Jolley *et al.* 2012b). Further, WGS involves the processing of a single specimen from which much information can be extracted, including: high resolution genetic relatedness of isolates; variation of vaccine and typing antigens; and likely vaccine and antimicrobial susceptibility (Jolley and Maiden 2013). The delivery of these data through easily accessible interfaces will enhance our ability to intervene in all bacterial infections from the level of the individual patient to entire populations. In the case of the meningococcus this will be especially important during the introduction of novel vaccines.

Chapter 3. Dissemination of lineage 11.2 (ET-15) meningococci during the 1993 Czech Republic epidemic

Abstract

Pathogen whole-genome sequence (WGS) data facilitate high resolution insight into the source and spread of epidemic strains. Meningococci belonging to clonal complex (cc) 11 have been responsible for worldwide outbreaks for at least six decades, and the ET-15 clone of cc11, which caused a severe epidemic across the Czech Republic in 1993, is associated with high mortality rates. In recent years, clusters of ET-15 meningococcal disease have arisen among 'men who have sex with men' in multiple countries, and cc11 disease incidence, which includes cases caused by ET-15 and serogroup W meningococci, continues to rise in England and Wales. These increases have prompted UK-wide adolescent immunization with a quadrivalent ACWY vaccine from August 2015 and surveillance will be instrumental in understanding its effects. ET-15 meningococci are, however, indistinguishable from other cc11 meningococci using seven-locus MLST, and the relationships among ET-15 cases have remained incompletely understood.

A comprehensive set of isolates from the well-characterised Czech Republic epidemic provided an opportunity to understand the population biology of ET-15 meningococci and to retrospectively investigate dissemination of the ET-15 clone across the Czech Republic using WGS data. ET-15 meningococci were members of a discrete cc11 sub-lineage, designated lineage 11.2, that is likely to have diverged some time before the clone was discovered. It was likely that the epidemic strain was introduced only once to the Czech Republic and that human demographics had a large impact on epidemic characteristics: transmission of meningococcal genoclouds was localised with increased incidence in highly populated districts. Although surveillance with WGS data will be essential for lasting control of cc11 disease, comprehensive epidemiological data remain a crucial component of outbreak investigation.

Introduction

Clonal complex (cc) 11 meningococci have caused high profile outbreaks of meningococcal disease for 60 years. These organisms are particularly invasive, being infrequently isolated from asymptomatic carriage relative to their incidence in disease, and are associated with high mortality and morbidity rates (Ashton *et al.* 1991, Maiden *et al.* 1998, Yazdankhah *et al.* 2004, von Gottberg *et al.* 2008). cc11 was circulating in the USA at least as early as the First World War and cc11 outbreaks were reported in the USA army in the 1960s and 1970s (Wang *et al.* 1993, Caugant 1998, Brundage *et al.* 2002); by the 1970s and 1980s, serogroup B and serogroup C cc11 meningococci were globally disseminated, causing disease in South America, Africa, China, USA, and Europe (Caugant *et al.* 1987b, Wang *et al.* 1993, Guibourdenche *et al.* 1996).

In 1988, multilocus enzyme electrophoresis (MLEE) of isolates from a serogroup C cc11 outbreak in Ontario, Canada, identified the ET-15 clone. This new group of cc11 meningococci subsequently caused most meningococcal disease in Canada (Ashton *et al.* 1991, Whalen *et al.* 1995). Since the earliest ET-15 isolate is from Canada in 1986, it was suggested that the clone emerged from the local Canadian meningococcal population shortly prior to its epidemic spread (Ashton *et al.* 1991, Caugant 1998). Retrospective characterization of isolate collections indicates rapid global dissemination throughout the 1990s: ET-15 meningococci were present in Europe at least as early as 1989, caused disease clusters in the USA from the early 1990s, and were observed in Australia in 1994 (Schaaf *et al.* 1989, Jackson *et al.* 1995, Jelfs 1997, Jelfs *et al.* 2000). Increased disease incidence in Iceland, Israel, and England throughout 1990s was the responsibility of ET-15 meningococci and a nationwide ET-15 epidemic occurred in the Czech Republic in 1993 (Krizova and Musilek 1995, Kaczmarski 1997, Caugant 1998, Feavers *et al.* 1999).

National epidemics and restricted clusters of ET-15 meningococcal disease from 1988-2001 were characterized by particularly severe symptoms and high case fatality rates in young adults, with ET-

15 described as a 'virulent clone' (Ashton *et al.* 1991). Cases from contained outbreaks in the UK (Wales and Southampton University), Australia, and the USA were linked to shared accommodation or bar visits (Ashton *et al.* 1991, Edmond *et al.* 1995, Krizova and Musilek 1995, Whalen *et al.* 1995, Erickson and De Wals 1998, Tsang *et al.* 2003). However, outbreaks were epidemiologically variable: carriage of the outbreak strain ranged from 0% to 12% of meningococci sampled during large epidemics, and single strain outbreaks lasted from four days to more than 11 months (Le Saux *et al.* 1992, Edmond *et al.* 1995, Imrey *et al.* 1996, Jelfs *et al.* 1998, Feavers *et al.* 1999, Gilmore *et al.* 1999, Round *et al.* 2001, Jolley *et al.* 2012b).

An extended period of high cc11 disease incidence, particularly in adolescents, occurred in the UK from the mid-1990s to the mid-2000s (Ch. 2 Fig. 1) (Ladhani *et al.* 2012a). Whilst the subsequent decline in incidence can be attributed to implementation of serogroup C conjugate vaccines in 1999 (Balmer *et al.* 2002, Maiden *et al.* 2008, Borrow *et al.* 2013), a resurgence of serogroup C cc11 disease in England and Wales in the past five years is thought to result from declining immunity in individuals immunized as adolescents, combined with expansion of a new ET-15 strain associated with 'men who have sex with men' (Tsang *et al.* 2003, Kupferschmidt 2013, Pollard *et al.* 2013, Weiss and Varma 2013). A further cc11 strain, characterised by a serogroup W capsule, is also becoming increasingly prevalent in England and Wales (Ladhani *et al.* 2015); the very high mortality rates reported in cases of serogroup W lineage 11 disease in South Africa (von Gottberg *et al.* 2008) and South America (Sorhouet-Pereira *et al.* 2013, Abad *et al.* 2014) are of great concern. Together, these prompted the decision to include a quadrivalent ACWY conjugate vaccine in UK immunization schedules from August 2015 (Ladhani *et al.* 2015).

Comprehensive molecular epidemiological studies of isolates from meningococcal disease are essential for public health control measures, whether during outbreaks or in surveillance following vaccine implementation. However, cc11 meningococci are highly conserved at both the antigenic and genotypic levels (Wang *et al.* 1993, Kriz and Musilek 1996, Buckee *et al.* 2010). Seven-locus

MLST identifies the majority of cc11 meningococci as the central genotype: although the FumC enzyme permits MLEE discrimination of ET-15 from cc11 meningococci, the *fumC* fragment sequenced in MLST does not include the characteristic ET-15 polymorphism (Vogel *et al.* 2000). In addition, although MLEE successfully discriminates ET-15 and non-ET-15 members of cc11, it offers insufficient resolution among the ET-15 strains responsible for outbreaks (Krizova and Musilek 1995, Jelfs *et al.* 2000, Tyler and Tsang 2004). IS1301 has been suggested as an alternative ET-15 marker, however, this is a mobile element that has been repeatedly observed in meningococci of other lineages (Hilse *et al.* 2000, Elias and Vogel 2007). Assessment of the population processes during outbreaks of ET-15 meningococcal disease has, therefore, been limited. For example, it is unclear whether ET-15 arises repeatedly from within endemic meningococcal populations through recombination and mutation in the core- and/or accessory-genomes, whether endemic populations of the clone temporarily expand through selective and/or neutral processes, or whether chance introductions to host populations leads to rapid spread (Wang *et al.* 1993, Achtman 1994, Achtman 1995, Morelli *et al.* 1997, Fraser *et al.* 2005).

Population genomic analysis of WGS data obtained from the isolates collected in national surveillance should achieve high resolution of meningococcal population structure and aid analysis of transmission patterns. The MRF-MGL provides one such opportunity for investigating the temporal distribution and relatedness of lineage 11 (cc11) disease cases in the UK (Lucidarme *et al.* 2015). However, it is unclear whether the inherent lack of cc11 diversity will permit detailed exploration of lineage 11 transmission dynamics at the small geotemporal scales characteristic of outbreaks. Isolates from national surveillance during 1993 Czech Republic epidemic provide the opportunity to perform a retrospective case study of the utility of WGS data in resolving the diversity and dispersal of cc11 meningococci during outbreaks. The established Czech Republic surveillance system rapidly recognised an increase in cases caused by C:2a:P1.2(1.5), ET-15, a strain not observed in the Czech Republic prior to 1993 (Krizova-Kuzemenska *et al.* 1993, Krizova *et al.* 1994, Krizova and Musilek 1995). Incidence of serogroup C disease almost doubled from 0.4 to 0.7 cases per 100,000

population following the first case in January and was associated with increased likelihood of Waterhouse-Friderichsen syndrome and a 20% case fatality rate (Roznovsky *et al.* 1994, Krizova and Musilek 1995, Krizova and Musilek 1997). In total, 21 disease and 32 carriage isolates of the epidemic strain were collected from across the Czech Republic in 1993 (Fig. 1). Here, WGS data were obtained from these isolates to: investigate the genomic diversity of ET-15 meningococci; understand further the characteristics of ET-15 outbreaks; and explore the utility of WGS data in routine national surveillance of cc11 meningococcal disease.

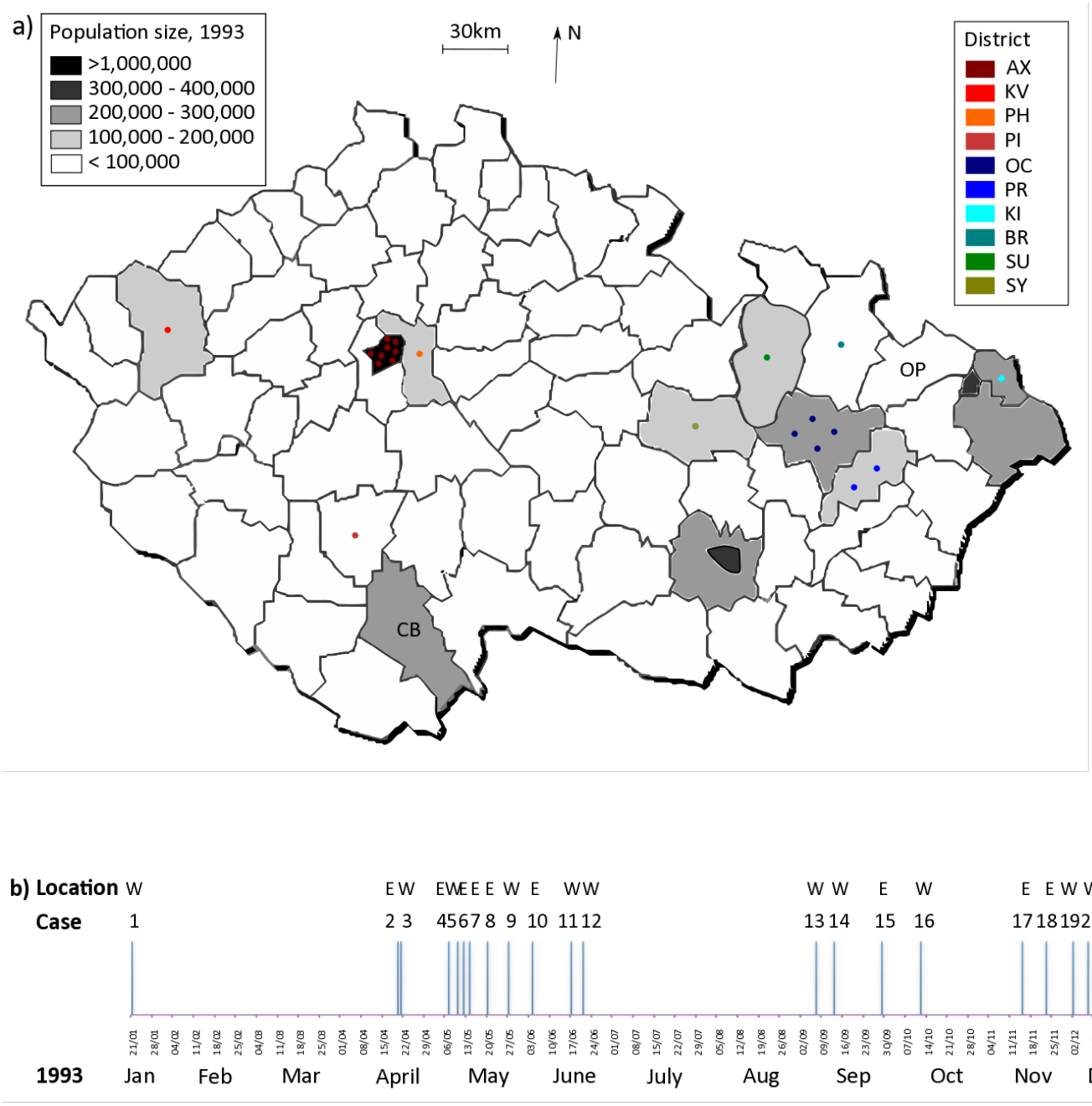


Figure 1. ET-15 meningococcal disease in the Czech Republic in 1993. a) Geographical distribution of ET-15 meningococcal isolates from the epidemic in 1993. Districts with populations greater than 100,000 individuals in 1993 are shaded. Coloured dots within districts represent single cases: cases are coloured according to the district from which isolates were obtained; the position of cases within districts is not significant. Carriage isolates were available from OP and CB, which had no disease cases, and from OC and AX. **b)** Timeline of ET-15 epidemic cases. W: the West of the Czech Republic (districts in red shades). E: the East of the Czech Republic (districts in blue/green shades).

Materials and Methods

Isolates

The Czech Republic National Reference Laboratory for Meningococcal Infections provided all available ET-15 isolates from the Czech Republic epidemic in 1993. These included 21 disease isolates (Fig. 1) and 33 carriage isolates collected from individuals aged 16-21 years, mainly from five classrooms in four districts at seven sampling times (Kriz 2004). The molecular epidemiology of these isolates has been described using traditional serological and genetic typing techniques (Krizova *et al.* 1994, Krizova and Musilek 1997, Jelfs *et al.* 2000, Jolley *et al.* 2000, Jolley *et al.* 2005).

In addition, 22 cc11 isolates with whole-genome sequence (WGS) data were available in the PubMLST database at the time of writing: 10 ET-15 isolates from an outbreak at Southampton University in 1997 including four identified to be the outbreak strain (Feavers *et al.* 1999, Jolley *et al.* 2012b); 10 isolates from the global EMGM reference collection (Maiden *et al.* 1998); and the cc11 reference genome, FAM18 (Bentley *et al.* 2007). These were used to contextualise the genetic diversity of the epidemic isolates. The cc8 reference genome G2136 was also included in cc11 population level analyses (Bennett *et al.* 2007).

Whole-genome sequencing

Isolates were cultured using standard methods (Harrison *et al.* 2015). Genomic DNA was prepared using the Wizard Genomic DNA Purification Kit (Promega) with the first and the final steps of the manufacturer's instructions altered for non-broth cultures as previously described (Jolley *et al.* 2012b). Standard Illumina libraries were generated using 1µg of genomic DNA sheared to 200-300bp using a Covaris E210 acoustic shearing device. Twelve libraries were pooled in an equimolar ratio for sequencing in a flowcell lane on the Illumina Genome Analyser IIX platform, generating 76bp paired-end reads. All sequence read data passed the Sanger Institute's in-house quality control assessment. Contiguous lengths of sequence (contigs) were assembled using Velvet (v1.2.01) and the

VelvetOptimiser script (http://bioinformatics.net.au/software.velvetoptimiser.shtml) that automatically scans the parameter space to produce optimised assemblies. A range of likely optimal *k*-mer lengths for input to VelvetOptimiser were initially estimated using Velveth and Velvetg (Zerbino and Birney 2008). Contigs longer than 100bp were made publically available in the PubMLST *Neisseria* database where all assembly metrics are available (Jolley and Maiden 2010).

Genome annotation

As in Chapter 2, WGS data were annotated, automatically or manually as necessary, using tools available in the PubMLST database. Seven-locus MLST sequence types (STs) and clonal complexes (cc), ribosomal STs (rSTs), and antigenic strain-types were automatically reported by the database. Original ST and strain-type designations (Kriz *et al.* 1999b, Jolley *et al.* 2000, Jolley *et al.* 2005) were cross-referenced with typing data extracted from WGS data: discrepancies were checked by alignment of Illumina reads to reference loci using Bowtie and Tablet (Langmead *et al.* 2009, Milne *et al.* 2010) and where inconclusive, individual loci were re-sequenced using standard methods (Jolley *et al.* 2000, Russell *et al.* 2004).

Population genomic analyses

Reference-based genome comparisons

At the time of analysis (April 2012) there were approximately 1,200 loci defined in the PubMLST *Neisseria* database, the majority of which were those universally present in the *Neisseria* genus. To increase the data necessary for highest resolution gene-by-gene comparisons (Ch. 2) (Jolley and Maiden 2010, Maiden *et al.* 2013), Genome Comparator was used to generate genome-wide allele profiles from among 1,975 annotated loci in the FAM18 reference genome (Bentley *et al.* 2007). At the time of analysis, distance matrices automatically reported by Genome Comparator included 'missing loci' (loci not present in isolate genomes or loci interrupted by ends of contigs) as a distance measure. To avoid artificial inflation of genome distances, missing loci were manually filtered from genome-wide allele profiles before calculation of distance matrices using the SplitsTree tool at

http://pubmlst.org/analysis. Neighbor-Net graphs (Huson and Bryant 2006) were generated from these distance matrices. In addition, concatenated alignments of loci present in all isolates under comparison were obtained using Genome Comparator. At the time of analysis, missing portions of any loci interrupted by the ends of contigs were automatically included in alignments as gaps; therefore, alignment columns containing gaps were manually filtered using Gblocks version 0.91b with default parameters (Castresana 2000, Talavera and Castresana 2007). Neighbor-Net graphs and maximum-likelihood trees with 1,000 bootstrap reconstructions were constructed from the variable-sites of resultant alignments, as in Chapter 2.

Reference-free genome comparisons

Phylogenies were also reconstructed from the nucleotide sequence data of homologous regions identified *de-novo* in un-annotated genomes. Contigs of the Czech Republic genomes were ordered against FAM18 using Mauve Contig Mover (Rissman *et al.* 2009) and multiple-genome alignments were generated using progressiveMauve version 2.3.1 with default parameters (Darling *et al.* 2010). This algorithm searches for homology using measures of genome similarity and synteny, generating a set of homologous Locally Collinear Blocks (LCBs). LCBs shared by all taxa were concatenated using the StripSubsetLCBs script (available at http://darlinglab.org) to create alignments of the sample core-genome. These were used to generate a 75% consensus ClonalFrame tree (three iterations of ClonalFrame were run with the parameters: –x 1000, -y 1000 and z -1) (Didelot and Falush 2007), variable-site Neighbor-Net graphs, and maximum-likelihood trees as above (Tamura *et al.* 2013). A binary distance matrix was created from the presence/absence of LCBs missing from at least one genome for generation of an accessory-genome Neighbor-Net graph.

Prokka (Seemann 2014) was used to bioinformatically predict novel coding sequences (CDS) in the cc11 FAM18 reference genome and in four ET-15 isolates that had been assembled to completion or on to a few contigs from PacBio WGS data (Ch. 4) (Eid *et al.* 2009). CDS were clustered using the BLASTclust script and accessory-genome CDS were identified by searching for CDS longer than 100aa

that were not present in FAM18. BLAST searches of these CDS in the five genomes were conducted for confirmation. Products and functions of loci were predicted using NCBI's conserved domain database (Marchler-Bauer *et al.* 2015) and NCBI's Clusters of Orthologous Groups (COG) functional classifications (Tatusov *et al.* 1997).

Statistical analysis

The latitude and longitude coordinates of the midpoints of Czech Republic districts and of the cities Prague and Olomouc were obtained from Google Earth. Pairwise geographical distances were calculated using The Geographic Distance Matrix Generator (Ersts). Statistical significance of the correlation between spatial and allelic distance matrices was determined in R using the Mantel test from the Ape 3.0-1 package (Paradis *et al.* 2004, R Development Core Team 2012). Geographical subdivision (F_{ST}) of ET-15 variation was assessed by analysis of molecular variance (AMOVA) in Arlequin version 3.5.1.2 (Excoffier *et al.* 1992, Excoffier *et al.* 2005). Recombination was detected using the pairwise homoplasy index (PHI) in PhiPack (Bruen *et al.* 2006): the default window size was used (100nt) and significance of the test statistic Φ_w was calculated under a normal approximation.

Results

Typing information derived from WGS data

On average, Czech Republic genomes were: 2,251,622bp in length; composed of 413 contigs; and had an N50 contig number of 26.72. There were discrepancies between typing information extracted from these data and original typing designations in only three isolates: (1) isolate 0075/93, originally designated ST-2510 with *abcZ* allele 9, was ST-11 with *abcZ* allele 2 in read-mapping and re-sequencing; (2) isolate 0104/93, originally ST-672 with *aroE* allele 117, was ST-11 with allele 4 in read-mapping; (3) isolate 0263/93, originally serogroup B, possessed genogroup C capsule genes (Kriz *et al.* 1999b, Jolley *et al.* 2000, Jolley *et al.* 2005). All Czech Republic ET-15 isolates possessed *fumC* allele 50, which contains the characteristic ET-15 adenine at position 640 (Vogel *et al.* 2000). All were strain-type C:P1.5,2:F3-6:ST-11(cc11) except for a single ST-733 (cc11) isolate. Isolate 0104/93 had the same PorA variable regions as the rest of the isolates but a different whole PorA coding sequence.

Lineage 11 population structure

Ribosomal MLST (rMLST) has been used for strain level characterisation of bacterial pathogens (Jolley *et al.* 2012a). A Neighbor-Net graph generated from concatenated nucleotide sequences of 49 non-paralagous meningococcal rMLST loci confirmed that ET-15 is a discrete lineage 11 sub-lineage (Fig. 2). This was designated lineage 11.2 and the sub-lineage of non-ET-15 isolates was designated sub-lineage 11.1, in accordance with previous nomenclature (Ch. 2) (Harrison *et al.* 2015, Lucidarme *et al.* 2015, Toros *et al.* 2015). The alleles at eight (16.33%) rMLST loci were sub-lineage specific. There was little ribosomal diversity within lineage 11.2: 49 (93%) Czech Republic isolates, the four Southampton outbreak strain isolates (Jolley *et al.* 2012b), and L93/4286 from England were indistinguishable (Fig. 2). The remainder of Czech Republic isolates varied by single nucleotide

polymorphisms in different rMLST loci except for 0023/93 and 0093/93 which possessed the same mutation in the *rpsA* locus (Fig. 2).

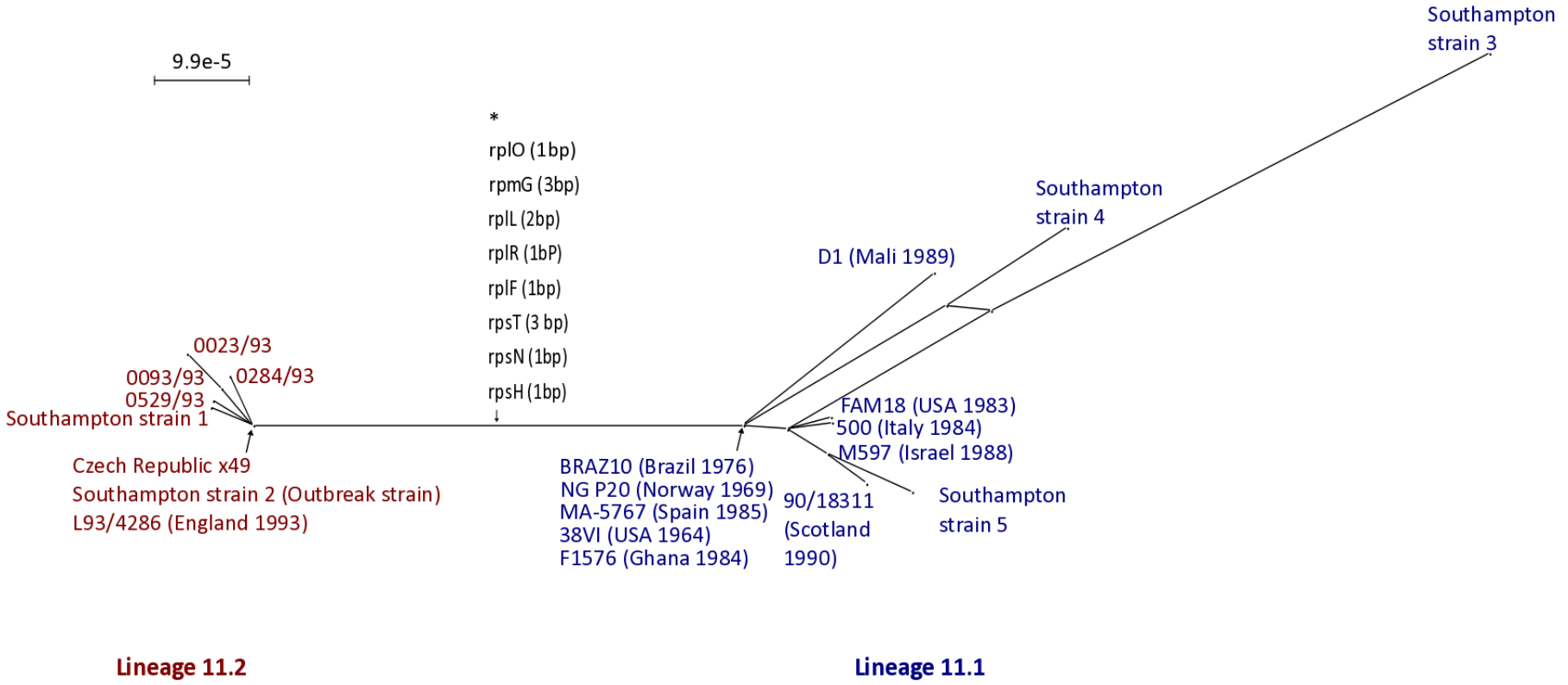


Figure 2. ET-15 meningococci form a discrete sub-lineage, lineage 11.2, within lineage 11 (cc11). Neighbor-Net graph generated from concatenated nucleotide sequence alignments of the 49 non-paralagous meningococcal rMLST loci. Isolates in red: MLEE designated members of the ET-15 clone. Isolates in blue: cc11 isolates not belonging to ET-15. *rMLST loci that distinguish sub-lineages 11.1 and 11.2.

The lineage 11 core-genome was composed of 1,514 loci. Although 510 (33.7%) of these loci were identical in all isolates, the lineage 11 core-genome was completely discriminatory of ET-15 isolates: each lineage 11.2 isolate was situated on a unique node in a Neighbor-Net graph (Fig. 3). Lineage 11.2 was distant from 11.1 and 308 (20.4%) loci had sub-lineage specific alleles; however, the lineage 8 (cc8) outgroup isolate was more distant, differing from lineage 11 at 780 loci (Fig. 3). A large proportion of sub-lineage discriminatory loci had uncharacterised functions, although seven had products implicated in host-pathogen interactions such as iron acquisition and utilization (e.g. *hpuB* and *lbpB*); eight were involved in adhesion to human cells (e.g. pilus assembly and glycosylation); and 69 were cell wall and capsule associated. The alleles at a second full-length MLST locus, *abcZ*, were also characteristic of sub-lineages.

Thirteen additional loci, excluded from distance matrix calculations, also had fixed sub-lineage specific alleles. These included: loci coding for the vaccine antigens FHBP (NEIS0349), NHBA (NEIS2109), FetA (NEIS1963), and NadA (NEIS1969); *tspA* and *iga2,* whose products interact with the human immune system; and *tbpA* and *tbpB,* which encode iron acquisition proteins (Boulton *et al.* 1999a, Boulton *et al.* 1999b, Frasch *et al.* 2001, Harrison *et al.* 2008, Brehony *et al.* 2009, Serruto *et al.* 2010). Lineage 11.1 isolates possessed the smaller isotype I *tbpB* in contrast to the lineage 8 and 11.2 isolates, which possessed isotype II (Harrison *et al.* 2008).

IS1301 has been described as an ET-15 marker (Elias and Vogel 2007). *nadA* was disrupted at position 471 by this insertion sequence in all lineage 11.2 (ET-15) isolates except in those belonging to Southampton University strain 1. IS1301 was not found in the genomes of lineage 11.1 isolates. IS1301 insertions could not be accurately quantified in lineage 11.2 genomes due to consistent assembly failure at the homopolymeric thymine stretch at position 303. Unlike in a previous study, sulphonamide sensitivity was not absolutely associated with ET-15 isolates (Radstrom *et al.* 1992, Jolley *et al.* 2012b): all lineage 11.2 isolates and the oldest lineage 11.1 isolate (from the USA, 1964) possessed the same allele, predicted to confer sulphonamide susceptibility, at the *folP* (*dhpS*) locus (Millar *et al.* 1963, Alexander *et al.* 1968, Qvarnstrom and Swedberg 2000, Fiebelkorn *et al.* 2005).

The *penA* locus, responsible for penicillin susceptibility, sub-lineage specific alleles, and all lineage 11 isolates were predicted to be penicillin susceptible in agreement with the phenotypic antibiotic susceptibility data available for these isolates (Antignac *et al.* 2001, Taha *et al.* 2007). These observations are consistent with a single divergence of lineage 11.2 (ET-15) from the rest of lineage 11 sometime before 1964.

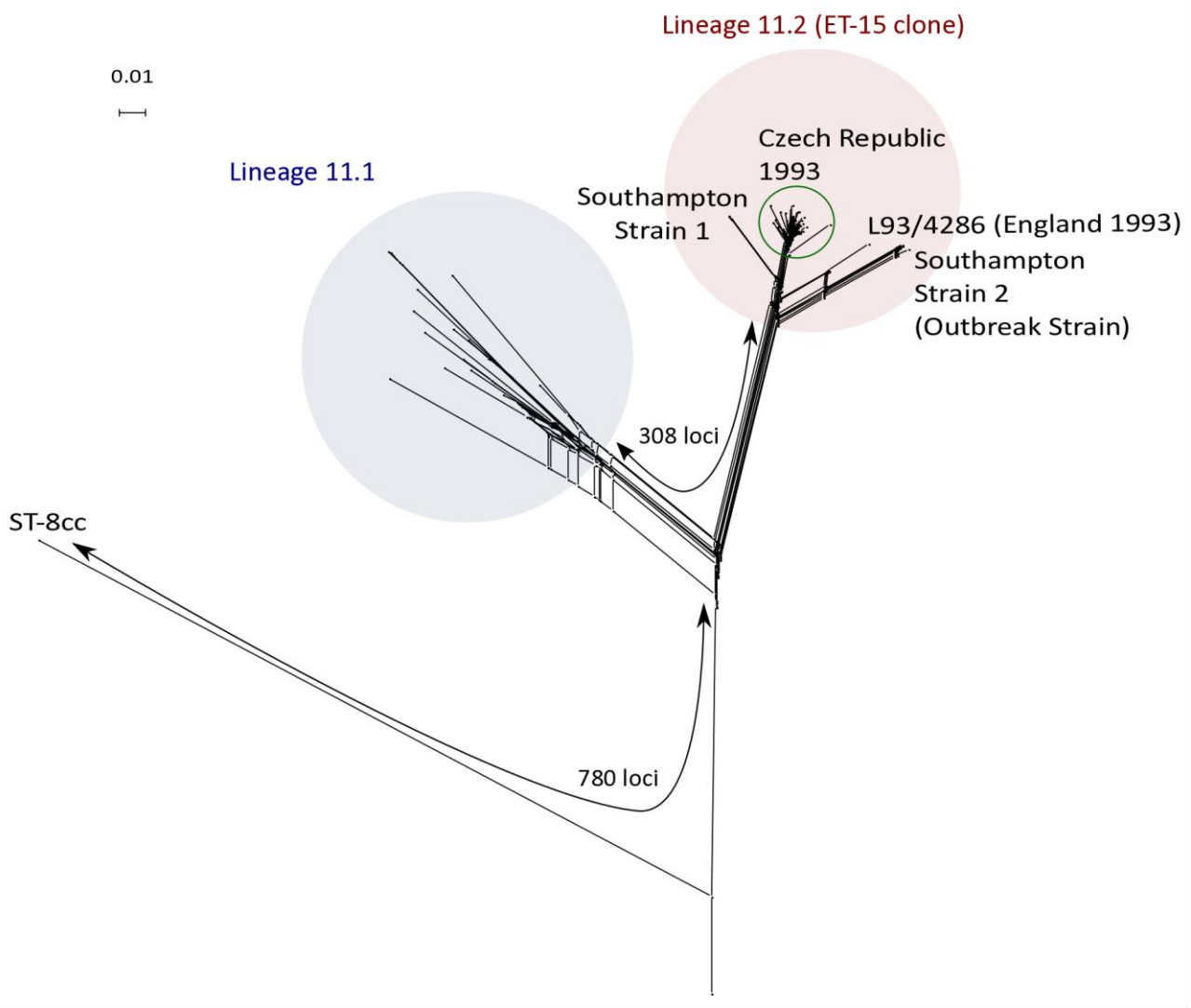


Figure 3. Lineage 11 (cc11) population structure. A Neighbor-Net graph generated from allelic distances among lineage 11 core-genome loci (n=1,514) showed that lineage 11 was composed of two sub-lineages, one of which (lineage 11.2) corresponded to the ET-15 clone identified with MLEE. The lineage 8 (cc8) isolate G2136 was included as an outgroup. Southampton strain nomenclature as in Jolley et al. 2012.

Genomic diversity of lineage 11.2 (ET-15 clone)

Among the 1,514 lineage 11 (cc11) core-genome loci, the average pairwise distance (pwd) among all lineage 11.2 isolates was 45.8 locus differences (Table 1; Fig. 3). Within this diversity, Czech Republic isolates from 1993 formed an outbreak cluster with a pwd of 21.9 loci and were more similar to each other than to English ET-15 isolates (Table 1; Fig. 3). In contrast, isolates obtained during the six weeks of the 1997 Southampton University outbreak (Feavers *et al.* 1999) had an average pwd of 91.4 loci: these were split into strain 1 and the outbreak strain (Jolley *et al.* 2012b) by the 1993 English isolate. The average pwd of the outbreak strain was 10.6 loci and comparable to the diversity of the Czech Republic epidemic (Table 1, Fig. 3).

Table 1. Gene-by-gene lineage 11.2 pairwise diversity

Isolates	Isolates (n)	Sample core-genome (# loci)	Av. pwd [SD]*
Lineage 11	75	Lineage 11 (1,514)	174.9 [190.9]
Sub-lineage 11.2	61	Lineage 11 (1,514)	45.8 [48.1]
England 1993-1997	11	Lineage 11 (1,514)	95.8 [67.0]
Southampton outbreak strain	4	Lineage 11 (1,514)	10.6 [6.13]
Czech Republic 1993	54	Lineage 11 (1,514)	21.9 [10.6]
'West' epidemic cluster	11	Czech Republic (1,610)	29.8 [17.0]
'East' epidemic cluster	43	Czech Republic (1,610)	24.9 [7.8]
Classrooms (n=5)	30	Czech Republic (1,610)	11.3 [na] to 22.2 [5.0]

*Av. pwd: average pairwise diversity (mean number of loci differing between two isolates). SD: standard deviation. na: not applicable.

The alleles of 31 loci were specific to lineage 11.2 strains circulating in England or the Czech Republic: these included six capsule loci and NEIS1699, which encodes an enzyme related to the *E. coli* mutator MutT (Appendix T1). It was possible that there were accessory-genome loci unique to the Czech Republic epidemic strain or to the Southampton outbreak strain. Using de-novo gene finding (Seemann 2014), a single coding sequence (CDS) was identified in two PacBio WGS Southampton isolates that was not present in two PacBio WGS Czech Republic isolates: this was a conserved hypothetical protein predicted to possess a conserved protein domain with matches to a

ski2-like helicase, found in other *N. meningitidis* genomes in GenBank such as the serogroup A reference genome Z2491 (Parkhill *et al.* 2000).

Genomic variation of the Czech Republic epidemic strain

The 54 epidemic isolates obtained across the Czech Republic between January and December 1993 were highly conserved. On average, 1,887,187bp (89%) of each genome (progressiveMauve alignment) were homologous among isolates (Appendix T2). Within this, the Czech Republic sample core-genome contained 1,610 loci, of which 1,290 (80.1%) were identical.

The remaining 320 loci each had between two and 16 alleles. The majority had two (80.9%) or three (13.8%) alleles and were hypothetical, conserved hypothetical, or putative protein-coding loci. The most variable categories were: capsule-associated loci (average 3.17 alleles); posttranslational modification, protein turnover, and chaperones (average 3.00 alleles); and pilus-associated loci (average 3.00 alleles). Following the elimination of pseudogenes and transposases, there were 17 loci with more than three alleles in the sample, including: four encoding cell wall proteins (*tspA*, NEIS1943, NEIS1623, NEIS0418 and NESI0417); three involved in amino acid transport and metabolism (*purC*, *argD*, and *gltS*); two pilus-associated loci (*pilQ* and *pilU*); and loci predicted to function in capsule polysaccharide modification (*lipB*), cell motility (*lepA*), protein glycosylation (*pglE*), and nucleotide excision repair (*uvrA*). All were characterised by multiple rare alleles. The most variable locus was a hypothetical protein (NEIS1702) with 16 alleles which mostly occurred once; all variation at this locus was generated by an alpha-alpha Correia element however. Loci with more than four alleles were dispersed around the genome except for *lipB* and NEIS0068, and NEIS0417 and NEIS0418, which were adjacent on the Fam18 chromosome (Bentley *et al.* 2007).

Geotemporal structuring of lineage 11.2 variation during the epidemic

The high ratio of external to internal branches and low diversity of the Czech Republic cluster suggested rapid expansion of the epidemic strain in the Czech Republic (Fig. 3) (Achtman 1997). Neighbor-Net graphs generated from the genome-wide diversity of the epidemic isolates revealed

two clusters associated with their geographic origin (Fig. 4). Isolates from the west (districts AX, KV, PH and PI) and the east (districts OC, BR, KI, PR, SU and SY) of the country clustered separately, except for several isolates from the west (case 5 from AX and CB classroom carriage isolates) that were situated in the 'East cluster' (Fig. 1, Fig. 4). There was strong statistical support for geographical subdivision during the epidemic (p-value = 0.001, Mantel test) and 21% of genetic variation in epidemic isolates could be explained by their geographical source (F_{ST} = 0.21450, p-value <0.0001, AMOVA).

The extent to which diversity was generated within the Czech Republic following strain introduction was unclear. Three loci had alleles specific to East and West clusters: a septum formation inhibitor (NEIS0160), a putative lipoprotein (NEIS0522), and a putative transmembrane transport protein (NEIS1208). There was some positive association between dates isolates were collected and their genetic distance from Fam18 (from 1983): case 1 was the most similar, differing from Fam18 at 513 lineage 11 core-genome loci and 14,308 single nucleotide polymorphisms, and the final two cases were among the three most divergent isolates with 541 and 530 loci differing from Fam18 respectively. There was some structure within East and West clusters: the West was more diverse than the East despite containing fewer isolates and there was a bottleneck within the East cluster composed of the anomalous west isolates (CB classroom isolates and case 5), cases 6, 7 and 15, and OC classroom isolates (Fig. 1; Table 1; Fig. 4b). No association between the date of the first case in a district and its distance from major local cities (Olomouc and Prague) was found (Fig. 1).

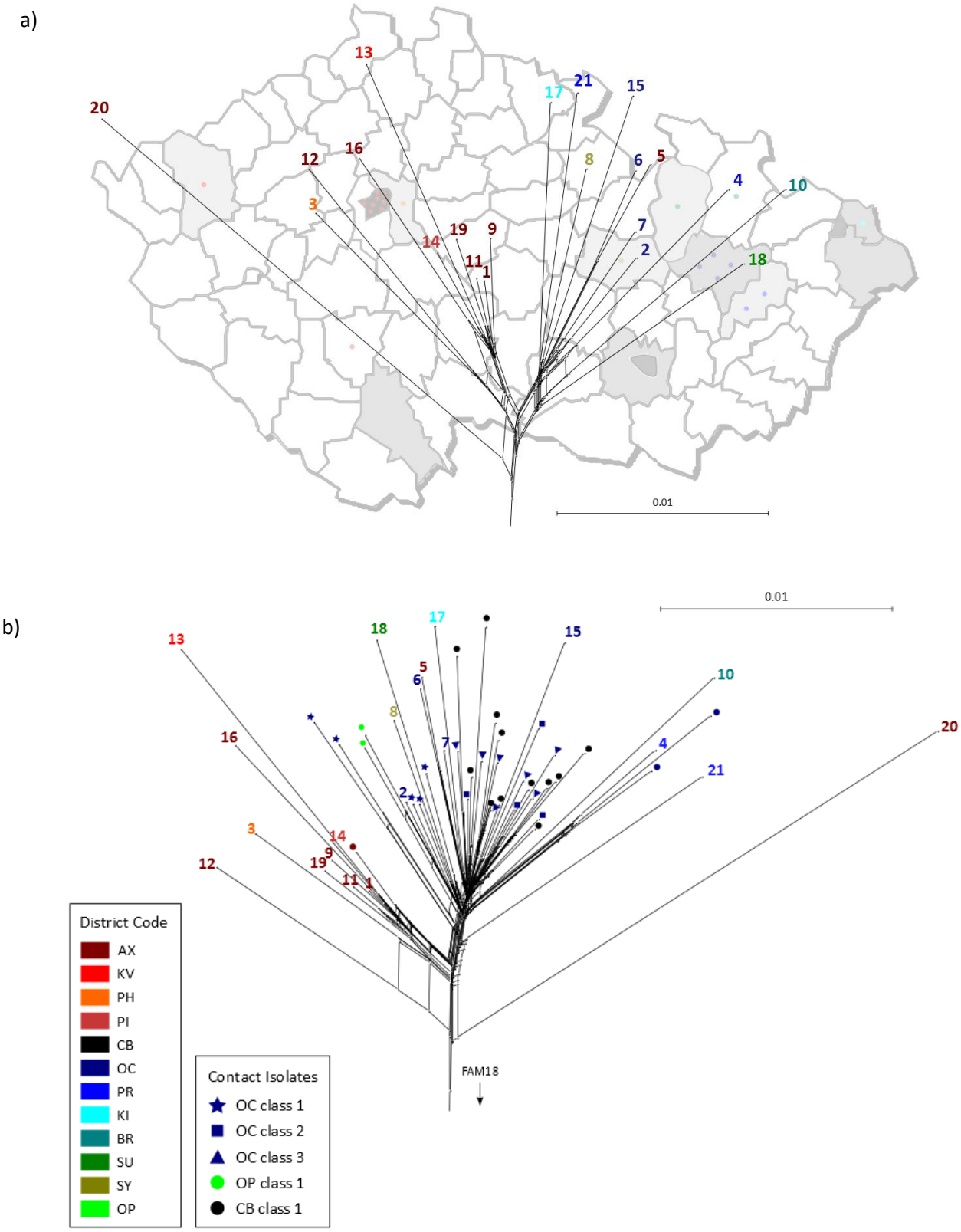


Figure 4. Geographical subdivision of lineage 11.2 (ET-15) during the Czech Republic epidemic. a) Neighbor-Net of disease ET-15 isolates generated from allelic distances among 1,772 shared loci. **b)** Neighbor-Net of 1993 Czech Republic ET-15 disease and carriage isolates generated from allelic distances at 1,610 shared loci. Cases 1 to 21 labelled by order of occurrence; classrooms from which carriage isolates collected denoted by shapes; districts denoted by colours, with western districts in red and orange and eastern districts in green and blue.

Detection of localised transmission

Within districts

There was little genetic clustering of isolates from the same district; even cases 11 and 12, and 19 and 20, collected within a week in the AX district did not cluster (Fig. 4). Case 4 from PR shared a 'recent' internal node with two carriage isolates sampled five days later from the neighbouring OC district (Fig. 1; Fig 4b).; it is possible that these carriage isolates were sampled during epidemiological investigation of case 4, although there was no information available to investigate this.

Within classrooms

There was, however, some clustering of carried meningococci isolated from adolescent school classmates. Isolates from a classroom had an average pwd of 17 loci: the most similar and most distant pairs had pwds of 9.7 and 35.4 respectively (Table 1). Isolates collected from the same classroom on a single day clustered loosely, for example: OP class 1 isolates shared a reticulated internal node at the exclusion of other isolates; OC class 1 isolates collected on the same day were highly similar; and all CB class 1, OC class 2, and OC class 3 isolates clustered within the bottleneck of the East cluster (Fig. 4b). However, there was no particular similarity among isolates collected from different classes in the same district,

Short transmission chains

These data were indicative of localised transmission of multiple ET-15 variants, or genoclouds (Zhu *et al.* 2001), and a great deal of unsampled diversity. To explore whether isolates part of short chains of transmission had been captured during investigation of the epidemic, any epidemiological links between genetically similar isolates were investigated. Isolates with the lowest pwds were separated by large geotemporal distances. For example, the most similar isolates differed at only 3 loci, were adjacent in the Neighbor-Net graph, and were isolated within two days of each other; however, they

originated from the distant OC and CB districts (Fig. 1a). Cases 9 and 11, both AX district, differed at only six loci but were temporally remote (six months separated disease onset) (Fig. 1; Fig. 4).

Cases 5 (male aged 45 years) and 6 (female aged three years) had the lowest combined geotemporal and genetic distances. Unlike other pairs these isolates shared an unreticulated, 'recent', internal node, and differed at 10 loci, comparable to the diversity of classroom carriage isolates (Fig. 4; Table 1). Disease onset occurred within two days (Fig. 1b). Although case 5 occurred in AX in the West and case 6 occurred in PR in the East, case 5 was the only case that did not cluster according to its geographical source and it was possible that invasive disease occurred subsequent to travel (Fig. 1).

Isolates from an epidemiologically supported short chain of transmission during the 1997 Southampton University outbreak were investigated for comparison. Contact tracing had identified carrier 1 to have been a direct personal contact of case 1 (Feavers *et al.* 1999, Gilmore *et al.* 1999) and these isolates had a diversity comparable to that of cases 5 and 6, with eight locus differences. Half of the loci were adjacent on the chromosome, however, and they possessed almost double the nucleotide polymorphisms (163 SNPs) than variable loci of cases 5 and 6 (85 SNPs) (Fig. 5). A putative rotamase-coding locus (NMC0276) containing a 9bp simple sequence repeat was the only locus to differ within both transmission pairs. Two case 5/6 polymorphic loci have been implicated in host-pathogen interactions: *tspA* (T-cell stimulating protein A, NEIS1829), which possesses a repeat tract and differed by a single indel (Oldfield *et al.* 2007), and *hpuA,* which was phase-variable off in case 5 (Fig. 5). HpuA expression has been shown to vary within individual hosts during blood infection (Omer *et al.* 2011). The remaining variable loci in the putative Czech Republic transmission event were a putative transmembrane potassium transporter *trkH*, a hypothetical protein coding locus *lepA,* which functions in cell motility, and *dnaK*, a putative chaperone: the alleles of these loci were not seen in the other 52 Czech Republic isolates. These data indicated that variation in the Southampton pair was more likely to have been horizontally generated, and therefore generated over a shorter evolutionary timescale, than that in Czech Republic cases 5 and 6.

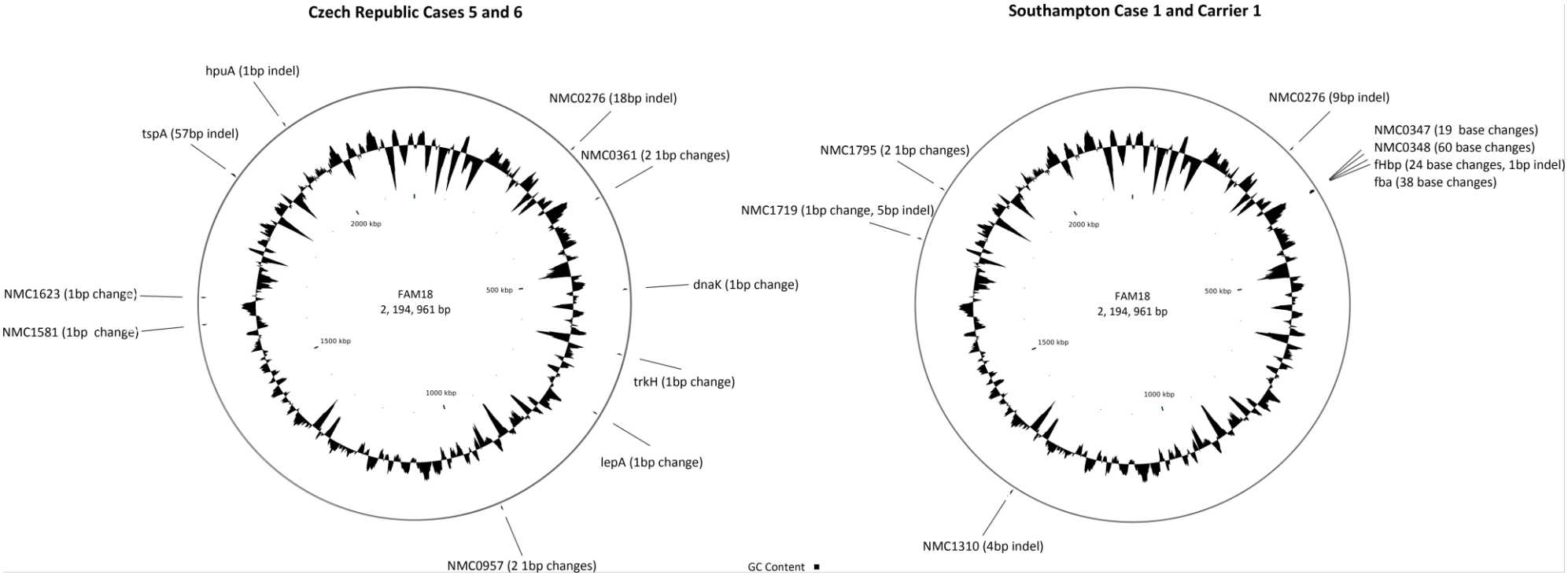


Figure 5. Genomic variation in ET-15 isolate pairs from short chains of transmission during outbreaks. Genomic differences between case 5 and case 6 of the 1993 Czech Republic epidemic and between case 1 of 1997 Southampton University outbreak and known close contact, carrier 1: four loci are clustered on the chromosome and likely to have undergone recombination. Figures generated with CGView (Stothard and Wishart 2005).

Alternative population genomic approaches

Transmission in bacterial outbreaks has been investigated using phylogenetic trees reconstructed from genome-wide single nucleotide polymorphism (SNP) data since these permit full resolution of the genetic changes among very similar isolates (Koser *et al.* 2014, Price *et al.* 2014, Stucki *et al.* 2015). Relationships of the Czech Republic epidemic isolates were therefore investigated using phylogenies constructed from the variable-sites in core-genome nucleotide alignments (Fig. 6a-d).

Polymorphic sites extracted from gene-by-gene data

An alignment was generated from concatenated nucleotide sequences of the 1,610 Czech Republic core-genome loci used in Fig. 4b. The alignment was 1,626,123bp and had 5,681 variable-sites of which 3,998 were phylogenetically informative. First, a Neighbor-Net graph (Fig. 6a) was generated from these variable-sites for comparison with Fig. 4. Deep structure was lost, however, six neighbouring isolates were arranged as in Fig. 4 (indicated with red brackets): these included the aforementioned case 4 and OC carriage isolate; the aforementioned cases 5 and 6; cases 8 and 18; and some classroom isolates. Since the phylogenetic method was consistent these differences can largely be attributed to the use of nucleotide data: the loss of deep structure suggested that much of the genetic distance among ET-15 epidemic isolates was homoplastic and generated through recombination. A PHI test for recombination was highly significant (p-value=0). Isolate neighbours common to both phylogenies were more likely to differ by mutation rather than recombination events, supporting the possibility that cases 5 and 6 were part of a short transmission chain.

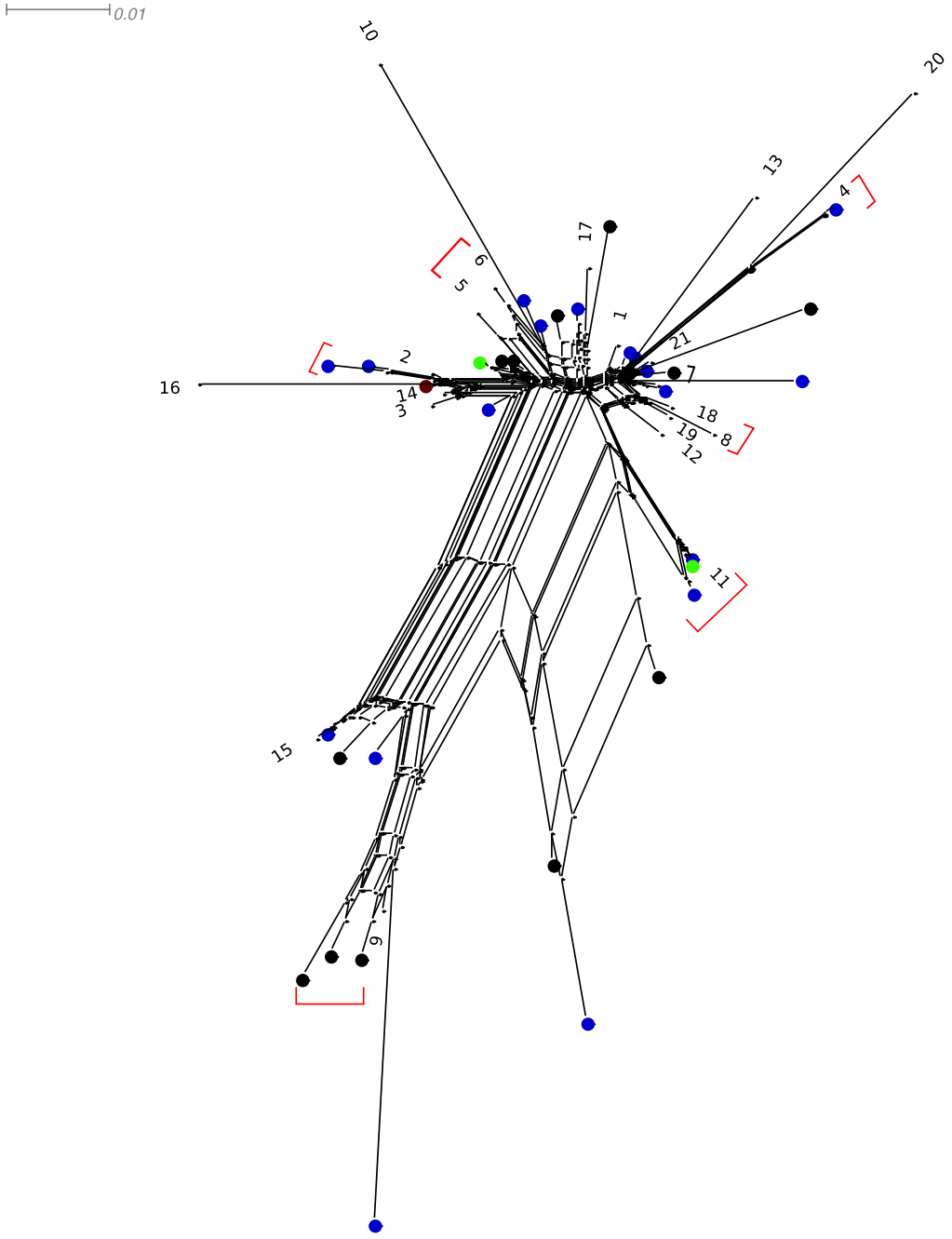
Second, a Maximum-Likelihood tree was generated from these nucleotide data (Fig. 6b). As expected, this was similar to the nucleotide Neighbor-Net (Fig. 6a): deep structure was lost and the same six pairs of isolates were neighbours (indicated with red brackets). Nodes leading to the consistently identified neighbours had 100% bootstrap support and nodes nearer the root had lower bootstrap values: it was likely that the deep structure of this tree was unreliable given the evidence for recombination.

Genome-wide polymorphic sites

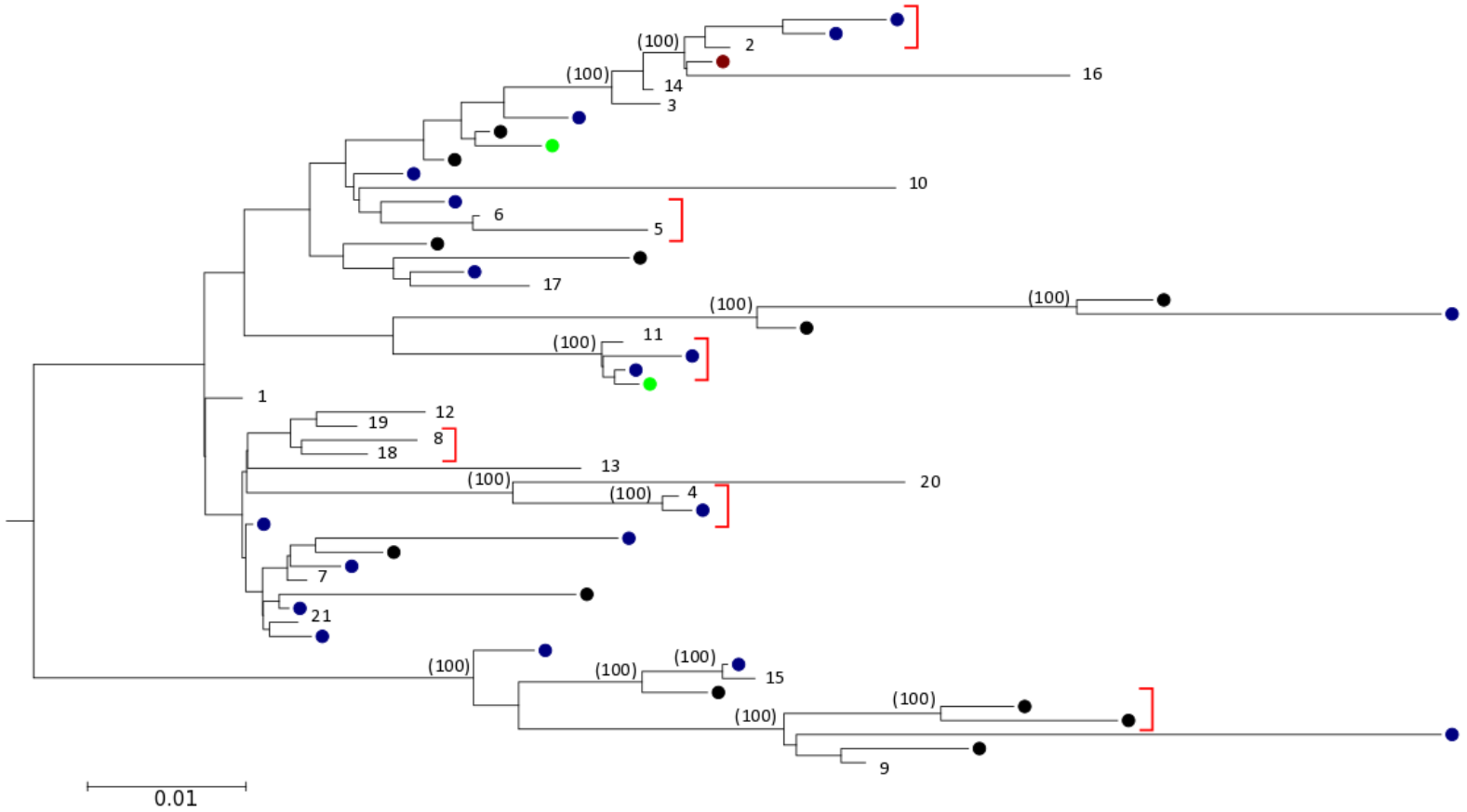
SNPs have also been extracted directly from WGS reads mapped to reference genomes or from reference-free multiple-genome alignments for generation of genomic genealogies (Harris *et al.* 2010, Darch *et al.* 2015, Montano *et al.* 2015, Struve *et al.* 2015). A progressiveMauve multiple-genome alignment of 2,777,363bp was generated from all Czech Republic ET-15 isolates. The sample core-genome totalled 1,887,665bp, 261,542bp longer than the Genome Comparator core-genome alignment, and included 5,681 variable-sites. In theory, this provided additional evolutionary information to the Genome Comparator approach that included only coding sequences, however, a Maximum-Likelihood tree produced from these data (Fig. 6c) provided no additional insight to the previous phylogenies (Fig. 4a, Fig. 6a-b): only five of the neighbouring isolate pairs were retained (indicated with red brackets) and the majority of nodes had bootstrap values less than 75%.

In all phylogenies, cases 16, 10, and 20 were positioned at the ends of long branches, which implied that the meningococci causing disease in these cases had undergone extensive diversification by recombination or mutation; the amplification of this effect in maximum-likelihood trees (Fig. 6b-c) was particularly interesting. In general, long external branches relative to internal branches may be evidence of expansion of epidemic strain in new environments. A PHI test for recombination in the progressiveMauve core-genome alignment was highly significant (p-value=3.36e-16). A ClonalFrame tree created from the same progressiveMauve variable-site data had negligible structure and retained only the relationship of the putative transmission pair case 5 and case 6 (indicated with red brackets) (Fig. 6d). Since this method accounts for homologous recombination events that have disrupted clonal signal among isolates (Didelot and Falush 2007), the lack of phylogenetic information in this tree was also indicative of extensive recombination.

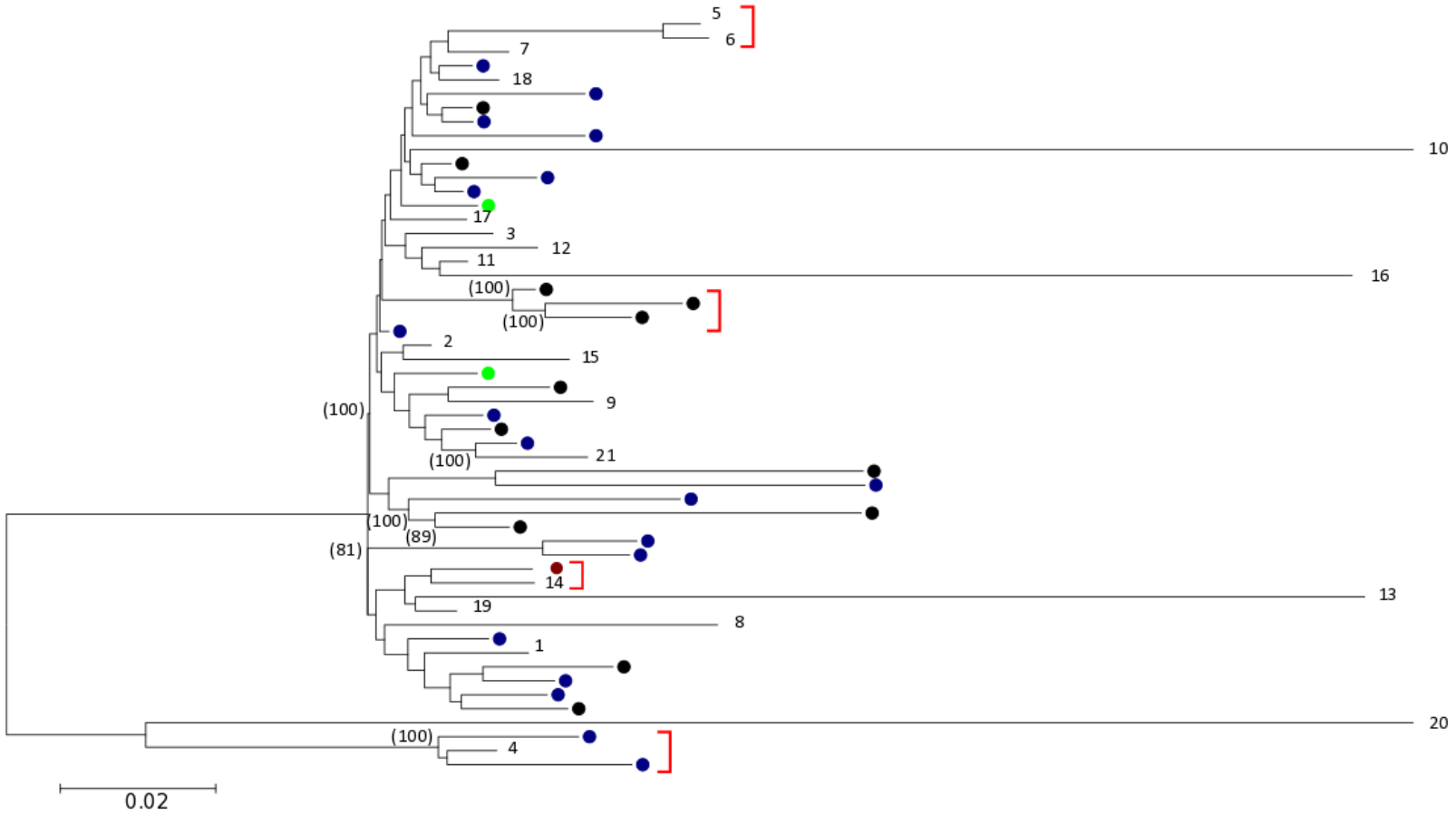
a)



b)


(100)
2
(100)
16
14
3
10
6
5
17
(100)
(100)
(100)
11
1
19
12
18
8
13
(100)
20
(100)
4
7
21
(100)
(100)
(100)
15
(100)
(100)
9
0.01

c)


5
6
7
18
10
17
3
12
11
16
(100)
(100)
2
15
9
(100)
(100)
21
(100)
(89)
(81)
14
13
19
8
1
20
(100)
4
0.02

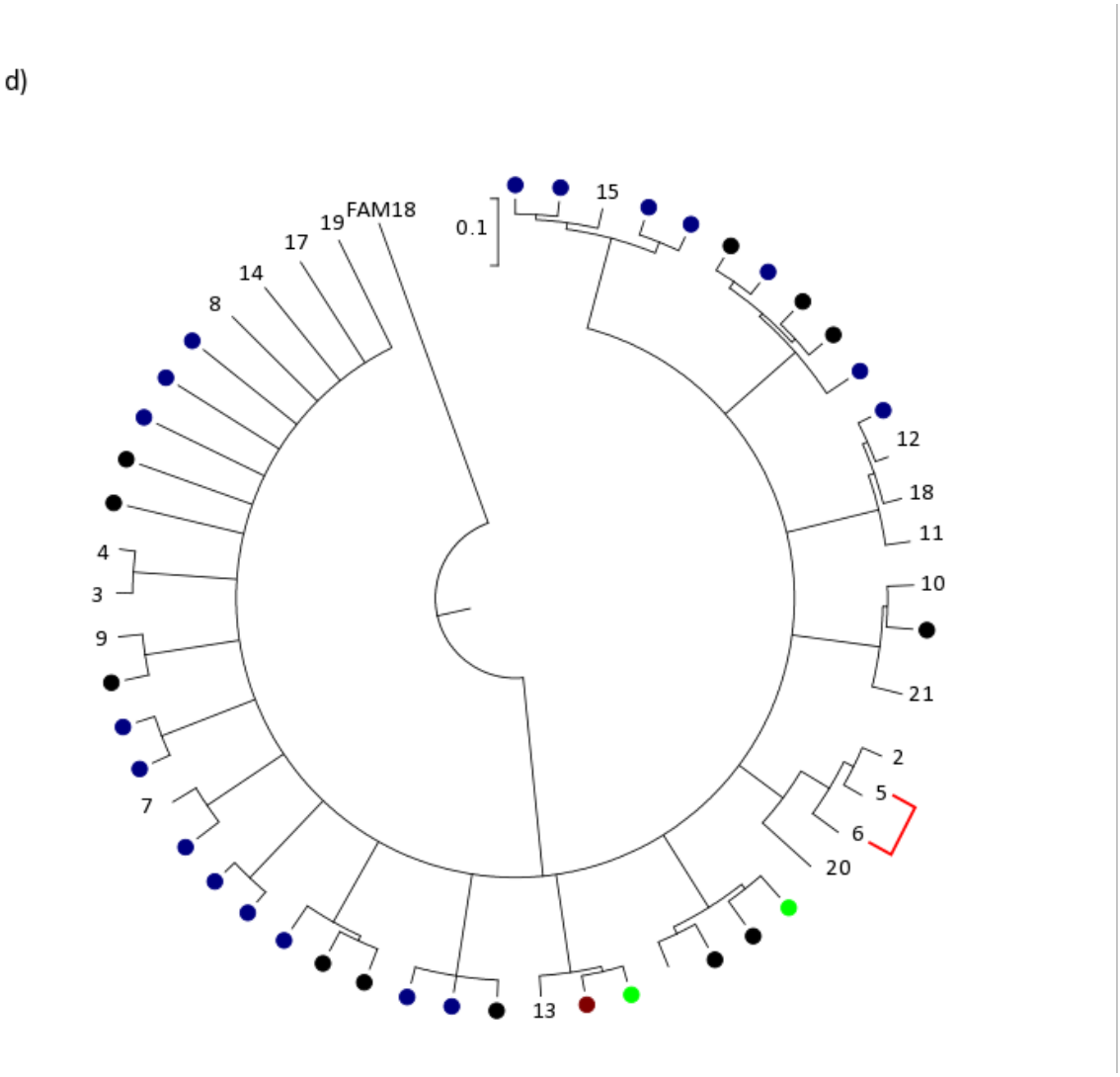


Figure 6. Nucleotide phylogenies of Czech Republic ET-15 isolates. The variable-sites (5,681bp) in the concatenated alignment (1,626,123bp) of the Czech Republic ET-15 core-genome loci (n= 1,610) were used to construct **(a)** a Neighbor-Net graph and **(b)** a Maximum-Likelihood tree. Multiple-genome alignments were generated using progressiveMauve for construction of **(c)** a core-genome variable-site (4,292bp) Maximum-Likelihood tree and **(d)** a core-genome ClonalFrame tree (1,887,665bp). Neighbouring taxa in the core-genome allele Neighbor-Net generated using the gene-by-gene approach (Fig. 4) are indicated with red brackets. Bootstrap values of 75% or greater are indicated in brackets.

Genomic differences among ET-15 disease and carriage isolates

It was possible that ET-15 meningococci from disease cases possessed genomic factors conferring increased likelihood of invasion that were absent from carriage isolates; some disease isolates were notably diverse (Fig. 4; Fig. 6). Variation in the core- and accessory-genomes was compared between disease and carriage isolates, however, no genes of the core-genome had alleles exclusive to isolates from disease or carriage, and genomic material comprising the accessory-genome (7% of each genome, on average; Appendix T2) also showed no association with disease or carriage (Fig. 7). Although there was some clustering of classroom isolates and of the isolate neighbours from core-genome analysis (edges indicated in red), Neighbor-Net graphs generated from the core- and accessory-genomes were largely incongruent (Fig. 4; Fig. 6; Fig. 7). There was a lack of deep structure in the accessory-genome network, which indicated high rates of horizontal exchange and/or a lack of phylogenetic signal.

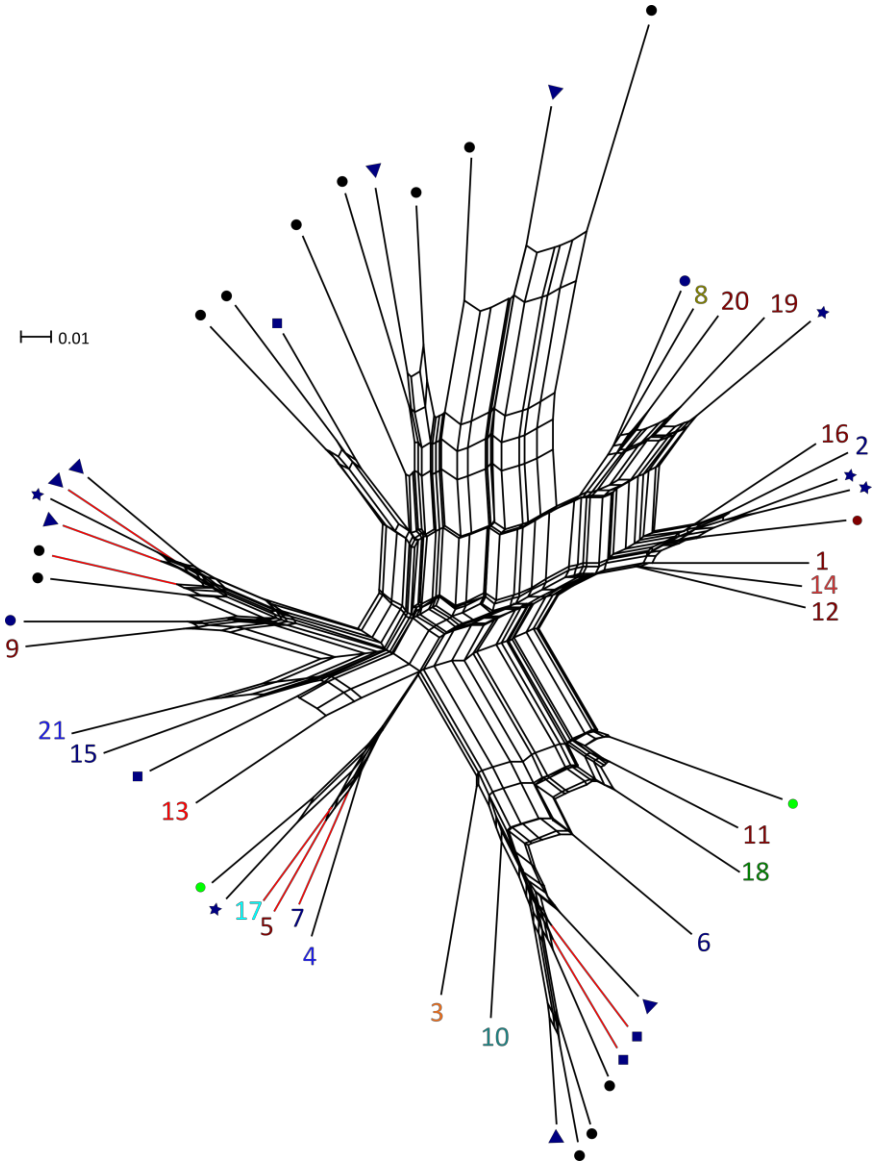


Figure 7. Accessory-genome Neighbor-Net of Czech Republic ET-15 isolates. A binary distance matrix was generated from presence/absence of accessory homologous blocks (on average 148,263bp per genome) from a progressiveMauve multiple-genome alignment. Edges in red are those that were also adjacent in the core-genome network (Fig. 4b). Taxa are annotated as in Figure 4.

Discussion

Lineage 11 (cc11) meningococcal disease is of increasing concern with recent outbreaks in communities of 'men who have sex with men' and rising prevalence in England and Wales (Weiss and Varma 2013, Ladhani *et al.* 2015). Epidemiological surveillance and understanding the population biology of these organisms is essential for limiting their public health impact, particularly in the context of imminent introduction of ACWY vaccines to the UK. In this study a well characterised lineage 11.2 (cc11, ET-15) epidemic was retrospectively investigated with the high genetic resolution afforded by whole-genome sequence (WGS) data. Epidemic cases were indistinguishable with conventional multilocus methods, however, WGS data were fully discriminatory, facilitating insight to the transmission dynamics of the strain during the epidemic and speculation on the biology of ET-15 organisms.

Evolution of lineage 11.2 (the ET-15 clone)

Contrary to suggestions that ET-15 meningococci belong to a recently evolved clone (Ashton *et al.* 1991), population genomic analyses indicated that these organisms belong to a discrete lineage 11 sub-lineage, lineage 11.2, that diverged from lineage 11 considerably earlier than 1964 (Fig. 2; Fig. 3). It is plausible that extensive use of sulphonamides between the Second World War and the early 1960s facilitated the expansion of lineage 11.2 around 1986 when the first isolates were collected (Millar *et al.* 1963, Caugant 1998). In this study all lineage 11.1 (non-ET-15) isolates were predicted to be sulphonamide resistant with the exception of a USA isolate from 1964, which shared a predicted sulphonamide sensitive *folP* allele with lineage 11.2 organisms (Qvarnstrom and Swedberg 2000, Fiebelkorn *et al.* 2005). Mass prophylactic administration of the drug in the 1960s in response to USA military outbreaks caused by lineage 11.1 meningococci was concurrent with the first recorded resistance to sulphonamides (Millar *et al.* 1963) and sulphonamide administration was subsequently halted. It is possible that from this point lineage 11.2 expanded following reduced

competition from lineage 11.1. Indeed, ET-15 meningococci in Quebec in the early 1990s had not developed sulphonamide resistance (Ringuette *et al.* 1995).

Dissemination of ET-15 meningococci during the Czech Republic epidemic

The gene-by-gene approach to population genomics and the understanding of cc11 epidemiology gained in previous studies permitted speculation on the dissemination of the epidemic strain throughout the Czech Republic (Maiden *et al.* 2013). It was likely that patterns of epidemic disease incidence resulted from: co-colonization, transmission, and localised circulation of clouds of ET-15 variants; rapid transmission through the Czech Republic population; and increased transmission and disease in densely populated areas.

Genomic variation of the epidemic strain

Distinct strains caused the 1993 Czech Republic epidemic and the 1997 Southampton outbreak (Fig. 3). Although lineage 11.2 (ET-15) was highly conserved with an average pairwise diversity (pwd) of only 45.8 locus differences, clustering of Czech Republic epidemic isolates at the exclusion of English isolates indicated that the Czech Republic and UK lineage 11.2 strains diverged before 1993 and that they did not diverge within either country (Table 1; Fig. 3). The great similarity of the epidemic isolates indicated that the epidemic strain was introduced once to the Czech Republic from a single source (a single individual or several individuals carrying the same strain), resulting in bottlenecking of lineage 11.2 diversity on introduction (Fig. 1a) (Achtman 1997). Although it was not possible to date introduction given the lack of temporal signal in the WGS data, the speed with which the epidemic became established suggests that introduction was not long before the first case in Prague on 21st January 1993 (Fig. 1b). The lifting of travel restrictions following the Velvet Revolution in 1989 and dissolution of Czechoslovakia on 1st January 1993 may have facilitated carriage of the strain into a naïve population via travel. Many ET-15 outbreaks in the 1990s are suggested, based on homogeneous within-country compared to among-country PFGE patterns, to have resulted from single introductions of ET-15 strains (Jelfs *et al.* 2000).

The extent to which the observed genomic variation was generated within the country versus its presence in initial introduction is unclear. The diversity of one year's epidemic isolates from across the Czech Republic (54 isolates) was comparable to that of Southampton outbreak strain (4 isolates) (Table 1) and appeared to be less than that of cc32 meningococci spanning five months from a single US state (80% versus 70% identical loci respectively) (Krauland *et al.* 2012). The lack of correlation between temporal and genetic distances during the epidemic is suggestive of variation generated prior to the first case. Together, it is likely that there was little generation of new ET-15 variation via mutation and recombination within the Czech Republic in 1993; however, there were no comparable estimates of meningococcal diversity or genome-wide mutation rates in the literature to permit further inference. Experimental evolutionary studies and quantification of meningococcal diversity, particularly within hosts, will be of great benefit in understanding outbreaks as WGS data becomes routine in public health surveillance.

Spatial transmission patterns

There was evidence of geographical subdivision of the epidemic strain in the existence of east and west ET-15 variants and in clustering of meningococci isolated from classrooms (Fig. 4). The lack of discrete variant clusters associated with districts and the diversity of meningococci in adolescent classmates at single time points, however, suggested that multiple variants were circulating concurrently in carriage (Table 1; Fig. 4). This was indicative of localised circulation of clouds of particularly similar ET-15 variants, or genoclouds (Zhu *et al.* 2001). Dissemination of the epidemic strain through the country was likely to have occurred more frequently via transmission on a local scale than via transmission events among individuals traveling long distances.

Nonetheless, soon after introduction it is likely that a subset of ET-15 variants was translocated to the east of the Czech Republic from the west. The greater diversity of meningococci in the west, location of the capital in the west (AX district), and occurrence of the first epidemic case in Prague was consistent with establishment of the ET-15 strain in this region before it was introduced to the

east (Table 1; Fig. 4). The occurrence of the second case on the 21st April in a particularly populated city in the east of the country, Olomouc (OC district), preceded an increase in the rate with which new cases occurred (Fig. 1). It was unlikely that ET-15 meningococci arrived in the east via passive diffusion across the country, but rather that they were translocated from west to east during travel: east and west variants had not hybridised by the end of 1993 and there were no cases in the centre of the country (Fig. 4; Fig. 1a). This resulted in a reduction of ET-15 diversity in an effect similar to that of sequential bottlenecking (Achtman 1997) (Fig. 4).

Epidemiological characteristics of ET-15 meningococci

Theoretically, for the ET-15 strain to have successfully spread through the Czech Republic following introduction, it must have had a large basic reproductive number (R_0: the average number of secondary carriers caused by a single primary carrier in a susceptible population) (Moxon and Jansen 2005). This quantity is determined by the transmissibility of the organism, the rate of invasion (with no transmission following invasion), the duration of carriage, and the number of susceptible hosts. Here, it is likely that the characteristics of the Czech Republic epidemic resulted from very low natural immunity in a population that had not previously encountered this strain-type (i.e. high number of susceptibles) (Buckee *et al.* 2010) combined with low carriage durations and high transmissibility. These led to particularly rapid transmission among members of the Czech Republic population. Rapid transmission was evident from the rate with which new epidemic cases occurred following introduction of the strain to the country (Fig. 1b); since invasive disease is very rare, new carriers must have acquired the strain at a far greater rate than invasive disease. The lack of district level subdivision suggests rapid homogenisation of variation following introduction of ET-15 meningococci to the east and west of the country (Fig. 4).

To permit rapid transmission, the rate of acquisition of ET-15 meningococci in carriage must have been high. However, as with other hyperinvasive meningococci, low point prevalence of cc11 meningococci in human carriage relative to disease incidence has been frequently observed

(Yazdankhah *et al.* 2004). For example: disease cc11 isolates were 6.6 times as frequent as carriage cc11 isolates in the PubMLST database; twice as many cc11 meningococci were isolated from disease than from carriage in Czech Republic adolescents in 1993 (Jolley *et al.* 2000, Caugant and Maiden 2009); ET-15 meningococci were not identified in carriage during the 1989/9 Ontario ET-15 outbreak or during the 1997 Southampton University ET-15 outbreak (except the direct contact of case 1); and ET-15 was identified in only 0.3% of carriers during a USA University ET-15 outbreak (Le Saux *et al.* 1992, Imrey *et al.* 1996, Gilmore *et al.* 1999). It may be that cc11/ET-15 is particularly virulent: ET-15 was associated with particularly severe symptoms and as many as 1 in 20 acquisitions of cc11 resulted in invasive disease (Roznovsky *et al.* 1994, Krizova and Musilek 1995, Stephens 1999). However, low point prevalence may also have resulted from low carriage duration within individuals, resulting in underestimation of meningococcal prevalence (Le Saux *et al.* 1992, Jolley *et al.* 2005).

According to epidemiological theory, the life history trade-off of transient carriage is high transmissibility: to maintain an R_0 sufficient for persistence, high transmissibility is essential to achieve infection of new hosts during short episodes of carriage (Jolley *et al.* 2000, Lappann *et al.* 2010). Previous observations are consistent with high transmissibility: transmission of cc11 between two individuals sharing an ambulance for just one hour led to invasive disease within 10 days and following acquisition of ET-15 meningococci during single bar and nightclub visits in Southampton, Sydney, and the USA, cases were reported after only 3-7 days (Edmond *et al.* 1995, Imrey *et al.* 1996, Jelfs *et al.* 1998, Gilmore *et al.* 1999, Puleston *et al.* 2012). In the Czech Republic, carriage isolates sampled from adolescent classmates at single time points were not necessarily more similar to each other than to those from other classrooms at different times (Fig. 4), and, compared to disease rates in this age-group, carriage rates were low (Jolley *et al.* 2005). This is consistent with concurrent circulation of multiple variants resulting from high rates of carriage acquisition and loss. Although ET-15 carriage rates were low compared to disease rates during the Czech Republic epidemic, this

ratio was not as large as observed in other ET-15 outbreaks and is likely to have resulted from the lack of natural immunity in the Czech Republic population (Jolley *et al.* 2000, Jolley *et al.* 2005).

Invasiveness of ET-15 meningococci

Rates of disease are thought to be determined by the combinatorial effects of expansion and contraction of strains with particular transmission fitnesses, population immunity, and perhaps the virulence of some strains (Buckee *et al.* 2008). Here, disease rates were higher in densely populated locations. All but two cases of ET-15 disease occurred in Czech Republic districts with populations greater than 100,000 individuals (Fig. 1a). Also, disease caused by both the east and west ET-15 variants occurred at higher frequency in districts containing large cities: the AX district surrounding Prague and the OC district surrounding Olomouc (Fig. 1a) (Kriz *et al.* 1995). AX was densely populated and army campuses that experienced increased incidence were situated in the north-east (Krizova and Musilek 1995) (Fig. 1a). It is possible that the high density of susceptible individuals in these locations was sufficient to sustain large transmission networks, leading to increased likelihood of invasive disease (Yazdankhah and Caugant 2004, Jolley *et al.* 2005). Only single sporadic cases occurred in the less populated districts surrounding AX and OC but it is unclear whether these cases occurred due to increased human movement through local cities or radial diffusion from these cities acting as disease foci. From 1993-1995 ET-15 meningococci came to spread across most districts of the Czech Republic (Krizova and Musilek 1997); a similar pattern of dissemination occurred in the first ET-15 outbreak where a focal outbreak expanded across the rest of the province after a year in Canada (Heimann *et al.* 1989, Ashton *et al.* 1991).

High host density scenarios are a common feature of ET-15 outbreaks. Educational institutes, army campuses, nightclubs, or bars have often been identified as an epidemiological link between cases; in one study a cluster of ET-15 cases occurred only in the most congested bar among individuals only transiently in close proximity (Pinner *et al.* 1991, Wang *et al.* 1993, Edmond *et al.* 1995, Imrey *et al.* 1996, Jelfs *et al.* 1998, Gilmore *et al.* 1999, Round *et al.* 2001, Puleston *et al.* 2012). Close proximity

of susceptible hosts is particularly important for pathogens with transient infective periods and the significance of host contact structure on meningococcal strain dynamics has been illustrated using mathematical models (Anderson and May 1979, Buckee *et al.* 2004, Bharti *et al.* 2012). Future lineage 11 studies of carriage and disease in populations with a range of host densities will be of great utility in identifying the factors predictive of outbreaks.

No loci of the core- or accessory-genomes were identified that consistently differed between carriage and diseases isolates (Fig. 4; Fig.7) in line with studies of variation during serogroup A meningococcal epidemics (Crowe *et al.* 1989, Achtman 1997). To date, no consistent differences have been found in the genomic content of disease and carriage isolates, however, it has been suggested that lineage level metabolic characteristics affect likelihood of invasion and that the outcome of colonisation may be mediated by contingency locus control of gene expression (Hotopp *et al.* 2006, Moxon *et al.* 2006, Buckee *et al.* 2008, Schoen *et al.* 2008, Joseph *et al.* 2011). Public health investigation at the time of the Czech Republic epidemic and Southampton outbreak identified no secondary cases of disease (Gilmore *et al.* 1999). Although allelic variation in outer-membrane protein loci and a helicase-like enzyme was specific to the different ET-15 strains responsible for the Czech Republic and Southampton outbreaks, it seems most plausible that the different epidemiological characteristics of these outbreaks were attributable to factors such as host immunity and behaviour.

WGS data in outbreak management

These insights are useful in understanding the extent to which WGS data will be beneficial in management of future lineage 11 outbreaks. The gene-by-gene approach to population genomics, implemented within the PubMLST database, permitted straightforward evaluation of the relationships among meningococci by increasing the number of loci needed for resolution as the variation among isolates decreased (Table 1). This method will provide an accessible way in which to

investigate isolate relationships at the lineage level and the transmission level as part of public health management (Jolley *et al.* 2012b, Maiden *et al.* 2013).

Populations

There were not enough ET-15 isolates in this study to estimate the source of the Czech Republic strain, but as whole-genome sequencing of isolates collected in surveillance becomes routine, it is very likely that this will be possible (Lucidarme *et al.* 2015). Within-country geographical subdivision was identified using network diagrams generated from allele data extracted using Genome Comparator (Fig. 4). This was not evident in: network displays of nucleotide data (Fig. 6a); more conventional Maximum-Likelihood genealogies that enforce a bifurcating tree structure (Fig. 6b-c); or in a tree generated using ClonalFrame, an algorithm developed to account for the effects of homologous recombination in meningococcal multilocus data (Fig. 6d) (Didelot *et al.* 2009).

It was unlikely that strong, statistically supported clustering of genome-wide variation by geographical location would occur by chance alone. Neighbor-Net graphs are a data display method and thus provide clear visualisation of complementary and contradictory splits in distance data without assumptions on evolutionary processes (Morrison 2011). Branches nearer the root in Maximum-Likelihood trees are particularly likely to be affected by high recombination rates in exponentially growing pathogen populations (Hedge and Wilson 2014). This demographic scenario is in line with the introduction of a novel ET-15 strain to a naïve population. High rates of recombination in these data were evidenced by significant PHI tests and in particular, the allele Neighbor-Net graph was less reticulated than the nucleotide graph, indicating that use of allele data accounted for some homoplasy from polymorphic sites introduced in single recombination events (Fig. 4a; Fig. 6a). High rates of recombination relative to mutation in ET-15 populations have been identified in previous studies (Kriz and Musilek 1996, Swartley *et al.* 1997, Musilek *et al.* 1998, Kriz *et al.* 1999a, Jolley *et al.* 2000, Jolley *et al.* 2005). Therefore, lack of deep structure in the phylogenetic

trees was likely to result from high rates of homologous recombination and possibly from contemporaneous ancestors and descendants in the epidemic sample.

Statistical methods such as BEAST that reconstruct genealogies and demographic processes (Drummond *et al.* 2012) are also affected by high rates of recombination (Hedge and Wilson 2014). The recent development of the ClonalFrameML algorithm, which applies the principals of ClonalFrame to whole-genome sequence alignments whilst reducing MCMC convergence issues (Didelot and Wilson 2015), could provide a targeted approach to build on initial network analysis.

Direct transmission

Identification of short transmission chains during outbreaks is dependent on understanding the relationships among isolates from cases. For cc11 meningococci, this is complicated by possible co-colonization by multiple variants, transmission of subsets of within-host diversity, unsampled diversity in carriage, and high rates of recombination. This inherent uncertainty means that defining a diversity cut-off for identification of transmission events in outbreaks will be complex. For the most part, however, all the core-genome methods used here identified the same highly similar isolate pairs (Fig. 4; Fig. 6). The gene-by-gene approach will be particularly useful in examination of the loci that vary among putative transmission pairs and their comparison with strain-level variation (Fig. 5). Establishing the necessary cut-off value for investigation of future lineage 11 outbreaks may in particular require understanding of within-host genomic diversity, within-host mutation rates, and lineage 11 carriage dynamics. In the meantime, genomic diversity measures could be obtained from multiple meningococcal colonies from single samples, or, statistical approaches for the detection of mixed infection in single samples could be developed, as for *Clostridium difficile* (Eyre *et al.* 2013a). It is clear, however, that detailed epidemiological data and meningococcal isolates from close contacts, as currently collected in public health outbreak investigations, remain indispensable.

Chapter 4. Variation and distribution of phase-variable Type III methyltransferases among meningococcal lineages

Abstract

A phase-variable regulon, or 'phasevarion', is a suite of genes whose expression is altered by a phase-variable methyltransferase (Mod) belonging to a Type III restriction-modification system. Some restriction-modification systems are thought to be specific to meningococcal lineages but the evolutionary forces shaping their distribution, and the distribution of phasevarions in particular, are unclear. Analysis of *mod* genes in whole-genome sequence data from 1,674 isolates indicated that there is diversifying selection on the three meningococcal phasevarions and that their proliferation takes place via horizontal gene transfer. Bioinformatic analysis of the genomic distribution of Mod epigenetic methylation suggested that phasevarions may affect the metabolic characteristics of certain lineages. There was an association between Mod variants and infection outcome that depended on whether they were found in hyperinvasive or non-hyperinvasive lineages.

Introduction

Reasons for high incidence of 'hyperinvasive' meningococcal lineages in disease relative to carriage, and for persistence of such lineages despite the lack of fitness advantage from disease, remain unclear (Caugant 1998, Yazdankhah *et al.* 2004, Buckee *et al.* 2008, Brown *et al.* 2012). Twenty years into the genomics era, the meningococcal determinants of invasion are still being elucidated. Alternative phenotypes result from alternative gene contents, sequence variation in shared content, or from gene expression changes, yet, few genomic loci have been identified that consistently differ between meningococci isolated from carriage and disease or that are lineage specific (Medini *et al.* 2008, Schoen *et al.* 2008, Joseph *et al.* 2010). Epidemiological approaches have identified associations between certain genomic loci and groups of lineages; the MDA phage, some iron-acquisition systems, and variants of the NadA adhesin are more frequent in certain hyperinvasive lineages (Comanducci *et al.* 2004, Bille *et al.* 2005, Hotopp *et al.* 2006, Bille *et al.* 2008, Budroni *et al.* 2011, Harrison *et al.* 2013a). Investigation of meningococcal gene activity in vivo is hindered by the lack of good animal models. Loci of restriction-modification systems (RMS) have, however, been shown to be heterogeneously distributed among meningococcal lineages (Claus *et al.* 2000, Budroni *et al.* 2011) and to mediate changes in the expression of genomic loci (Srikhanta *et al.* 2009, Seib *et al.* 2011).

RMS activity consists of sequence-specific methylation by a methyltransferase and cleavage of un-methylated DNA by a restriction-endonuclease (Boyer 1971, Srikhanta *et al.* 2010). Upon discovery that these systems were particularly prevalent in naturally competent organisms, it was proposed that RMS have a primary role in protection of bacterial cells from invading phage DNA (Anderson and Felix 1952). This view is now generally accepted as a mechanism that may result in sexual isolation of particular strains (Dupuis *et al.* 2013, Vasu and Nagaraja 2013, Murphy *et al.* 2014). However, variation in the distribution and mechanisms of RMS among bacterial species has led to a variety of additional proposals for the adaptive nature of these systems. For example, it has been

suggested that restriction-modification systems: (i) are selfish elements, with lack of their propagation in bacterial progeny resulting in postsegregational killing (Naito *et al.* 1995); (ii) aid generation of genetic diversity by creating double-strand breaks on incoming DNA and promoting efficient recombination (Arber 2000); (iii) have a role in posttranscriptional gene regulation, cell cycle control, and DNA mismatch repair as a result of their methylation activity (Low and Casadesus 2008).

The Type III RMS are of particular interest in the meningococcus. As in other human adapted pathogens, the *mod* genes encoding Mod methyltransferases in some meningococcal Type III RMS belong to an unusual class of contingency loci (Moxon *et al.* 2006). Despite being an obligate human inhabitant, the meningococcus experiences a multitude of dynamic environments during transmission, colonization, and invasion, and the phase-variation of contingency loci allows random phenotypic variation that may result in fitness increases. The selective advantages of phase-variable RMS are unclear, however. It has been suggested that phase-variation permits generation of genotypic variation in challenging environments through removal of the barrier to transformation; however, many Type III *mod* genes are associated with inactive *res* components (encoding Res, the restriction enzyme) that therefore do not prevent transformation in the first place (Ando *et al.* 2000, Fox *et al.* 2007, Srikhanta *et al.* 2009, Srikhanta *et al.* 2010). Second, it has been suggested that phase-variable RMS have evolved through kin selection, since autolysis following self-restriction releases DNA for uptake by sister cells (Dybvig *et al.* 1998, Saunders *et al.* 1998, Srikhanta *et al.* 2010); however, since Res is only active when in complex with Mod and since methylation activity is favored over restriction, this theory is disputed (Dryden *et al.* 2001). Most convincingly, it has been suggested that phase-variable methylation of genomic loci randomly alters gene expression in a shot-gun approach to survival in changing environments, a function that would render the activity of Res irrelevant (Srikhanta *et al.* 2010).

Three phase-variable *mod* loci have been identified in meningococcal genomes that alter the expression of up to 80 genes each. These genes encode pathogenicity factors such as lactoferrin-binding protein and heat-shock proteins, affect biofilm formation, and confer antimicrobial resistance (Srikhanta *et al.* 2005, Srikhanta *et al.* 2009, Jen *et al.* 2014). The specific DNA sequence recognised by Mod is determined by its DNA recognition domain (DRD): the *modA*, *modB*, and *modD* genes each contain different DRDs and are thought to regulate different suites of genomic loci. In addition, sequence comparisons have found that individual *mod* genes may possess one of a number of DRD 'alleles'; the particular combination of *mod* loci possessed by a meningococcal cell could therefore affect expression of hundreds of genes. To date, however, gene expression studies of some phasevarions have been hindered by undetermined Mod recognition sites.

In combination with observations that *modD* is associated with hyperinvasive cc41/44 isolates and that a range of Mod alleles have been identified among clonal complexes (Srikhanta *et al.* 2009, Seib *et al.* 2011), it is possible that particular phasevarions characterise particular meningococcal lineages. The prevalence of meningococcal whole-genome sequence data now permits unification of population studies with functional studies. In particular, PacBio SMRT sequencing permits identification of the epigenetic signatures of methyltransferases and differentiation of meningococci at the highest possible level of genetic resolution (Flusberg *et al.* 2010, Korlach and Turner 2012). In this study, the prevalence, diversity, and lineage association of the three *mod* genes were investigated among diverse meningococcal genomes. New DRD alleles were discovered and *mod* genes and their DRD alleles were found to be highly lineage associated, despite evidence of frequent horizontal gene transfer. Certain alleles were particularly associated with hyperinvasive lineages, but their relationship with invasive disease interacted with lineage invasiveness. In addition, the recognition site of the most prevalent meningococcal Mod allele was discovered for the first time; the focus of this enzyme's activity was metabolic loci.

Methods

Isolates

In total, 1,674 meningococcal isolates were analysed (Table 1). The majority of disease isolates were from the MRF Meningococcus Genome Library (MRF-MGL; n=1,380) and from a collection of well characterised genomes ('Global collection'; n=20) (Parkhill *et al.* 2000, Tettelin *et al.* 2000, Bentley *et al.* 2007, Peng *et al.* 2008, Schoen *et al.* 2008, Budroni *et al.* 2011). The majority of carriage isolates were from the Czech Republic in 1993 (n=263) (Jolley *et al.* 2000, Jolley *et al.* 2005). Genomes from meningococci (n=11) isolated during a lineage 11 outbreak at Southampton University in 1997 were also included (Feavers *et al.* 1999, Gilmore *et al.* 1999, Jolley *et al.* 2012b).

Table 1. Isolates surveyed for presence of *mod* genes.

Isolate collection	Temporal range	# carriage isolates (%)	# disease isolates (%)	Total isolates (%)
MRF-MGL (UK)	2010-2013	0	1380 (82.44)	1380 (82.44)
Czech Republic	1993	209 (12.49)	54 (3.23)	263 (15.71)
Global	1976-2005	2 (0.12)	18 (1.08)	20 (1.19)
Southampton outbreak	1997	5 (0.30)	6 (0.36)	11 (0.66)
Total	**1976-2013**	**216 (12.90)**	**1458 (87.10)**	**1674 (100)**

%: percentage of total dataset.

Isolates were assigned to whole-genome derived meningococcal lineages using allelic distances among rMLST and cgMLST loci, as described in Chapter 2.

Bioinformatic curation of *mod* loci in PubMLST

Locus records for *modA* (NEIS1310), *modB* (NEIS1194) and *modD* (NEIS2364) were created in the PubMLST *Neisseria* sequence definitions database (http://pubmlst.org/neisseria) (Jolley and Maiden 2010). Records were seeded with *modA* and *modB* reference sequences from the MC58 genome (Tettelin *et al.* 2000, Srikhanta *et al.* 2009) and with the *modD* reference sequence from the M0579

genome (Budroni *et al.* 2011). The PubMLST BLAST tool (Jolley and Maiden 2013) was used to identify these reference sequences within whole-genome sequence (WGS) data (parameters: word size 11; reward 2; penalty -3; gap open 5; gap extend 2; return hits >30% alignment and >50% identity). Since *mod* gene sequences were highly diverse, hits were manually aligned against previously identified *mod* sequences (Fox *et al.* 2007, Gawthorne *et al.* 2012) in MEGA6 (Tamura *et al.* 2013). Out-of-frame sequences were edited to obtain in-frame phase-variable repeat tracts, allowing trimming to the distal initiation codon and identification of internal stop codons (Srikhanta *et al.* 2009), incorrect BLAST hits were removed, and DNA recognition domain (DRD) 'alleles' were identified within the gene sequence.

Unique gene sequences were uploaded to the appropriate locus record for storage as database alleles and were assigned to *mod* locus 'variants' based on their DRD allele (Fox *et al.* 2007, Srikhanta *et al.* 2009, Seib *et al.* 2011, Gawthorne *et al.* 2012). The 'autotagger' and 'autodefiner' tools were then used to automatically tag *mod* genomic positions and assign *mod* allele numbers in WGS data (Jolley and Maiden 2013). *mod* loci were frequently present on different contiguous sequences (contigs) of a genome assembly due to break points within phase-variable repeat regions: allele numbers were not assigned and database flags (orange 's') were applied to individual isolate records to indicate partial assemblies. This process was performed iteratively until no new alleles of the three *mod* loci could be identified, a process taking several months for the three *mod* loci.

Analysis of nucleotide sequence data

Insertion sequences (IS) were identified using the search tool at ISFinder (Siguier *et al.* 2006). New DRD variants of *modA*, *modB*, and *modD* were defined when the DRD was >95% identical to existing DRD alleles (Gawthorne *et al.* 2012). Between- and within- variant DRD nucleotide *p*-distances were calculated in MEGA6 for comparison of sequence variation: alignments were trimmed to the longest DRD, gaps were deleted in pairwise comparisons, and all substitution types were included. Variants

were named in accordance with 'allele' nomenclature with new variants assigned the next available number.

Unique sequences of each *mod* gene were aligned in MEGA6 using Muscle with default parameters (Edgar 2004). All but the first repeat unit of phase-variable tracts were trimmed from alignments and unique trimmed sequences were identified using the Non-redundant databases (NRDB) program written by Warren Gish, Washington University. Neighbor-Net graphs were created in SplitsTree4 (Huson and Bryant 2006) using distance matrices calculated at http://PubMLST.org/analysis from allele profiles, or distances matrices calculated by SplitsTree4 from nucleotide alignments. Networks generated from core-genome (cg) data made use of the cgMLST version 2 scheme (1,592 loci) in PubMLST. Trimmed alignment figures were generated using JalView 2.8.2. (Waterhouse *et al.* 2009).

Statistical analysis

Statistics were performed in R version 3.0.2 (R Development Core Team 2012) except for Simpson's Diversity Indices (*D*) calculated manually in Excel. Odds ratios (conditional maximum likelihood estimation method) and confidence intervals (Fisher exact method) for associations of lineages with *mod* variants were calculated using the 'oddsratios' function in the 'epitools' package (Rothman 2002, Aragon 2012) on manually constructed contingency tables.

Associations between *mod* variants and 'lineage invasiveness' were measured as odds ratios in a binomial logistic regression model for each *mod* gene. Lineages were grouped into recognized 'hyperinvasive lineages' (lineages 2, 3, 4, 5, 6, 10, 11, 14, 18, 23, 39) (Caugant 1998, Yazdankhah *et al.* 2004, Caugant and Maiden 2009) and 'non-hyperinvasive' lineages for modeling as a dichotomous 'lineage invasiveness' response variable. *mod* variant was modeled as a categorical predictor variable with minor variants grouped into a single category (*modA18*, *modA19*, *modA2*, *modA4*, *modA6* and 'null'; *modB3*, *modB5*, *modB6*, 'null' and unknown; *modD2*, *modD7*, *modD3*).

Associations of *mod* variants with 'infection outcome' (carriage or disease) were measured as odds ratios in a multivariable logistic regression model for each *mod* gene. The predictor variable was *mod* variant (grouped as above) with 'lineage invasiveness' included as an additional predictor to account for the increased likelihood that meningococci from hyperinvasive lineages were disease isolates. Two multivariable 'infection outcome' logistic regression models were run for each *mod* gene: once with 'lineage invasiveness' as an independent predictor of infection outcome and once with 'lineage invasiveness' and *mod* variant as an interaction term. In all cases, the standard R glm function was used with the binomial logit.

Methylome characterisation

Two Czech Republic isolates and two Southampton University isolates belonging to lineage 11.2 (cc11, ET-15) were chosen to undergo PacBio SMRT sequencing at Pacific Biosciences, California, for identification of methyltransferases and their recognition sites (Table 1). Isolates were chosen because each pair was very closely related based on core-genome similarity and epidemiological data (Ch. 3) (Gilmore *et al.* 1999, Jolley *et al.* 2012b) and each comprised a carriage and disease isolate, allowing methylome comparison. Meningococcal culture and genomic DNA preparation were carried out using standard methods (Jolley *et al.* 2012a, Harrison *et al.* 2015).

Generation of PacBio SMRT and methylation data was carried out by Tyson Clark and colleagues at Pacific Biosciences California. SMRT sequencing was performed as previously described (Clark *et al.* 2012) and sequence reads were hybrid assembled using the Celera assembler (Koren *et al.* 2012) and de-novo assembled with the Hierarchical Genome Assembly Process (HGAP) (Chin *et al.* 2013), resulting in genomes composed of one to 62 contigs. Polymerase kinetics from SMRT sequencing were analysed by PacBio to identify modified sequence motifs using the SMRT analysis pipeline version 1.3.3 (http://www.pacb.com/devnet) as previously described (Lluch-Senar *et al.* 2013, Schadt *et al.* 2013). Briefly, interpulse durations (IPD) were measured at each genomic position and compared to those of an in-silico control (model predicted IPDs for each position based on local

sequence context) using a t-test with a log-normal distribution model and associated p-values. A given position was assigned 'methylated' when the log-transformed p-value was >50. Sequence motif cluster analysis was carried out using custom scripts.

Methyltransferase genes within each of the four genomes were bioinformatically predicted by Richard Roberts at New England Biolabs (NEB), New Jersey, using the SEQWARE program (O'Connor *et al.* 2010) and REBASE (Roberts *et al.* 2015) as previously described (Fang *et al.* 2012). Briefly, SEQWARE was used to: scan WGS data for homology to REBASE restriction modification systems (RMS); predict new RMS components based on criteria such as conserved functional motifs; and to predict possible recognition sequences.

Cloning experiments were performed by Rick Morgan and Yvette Luyten at NEB to confirm the recognition site of ModA12 (Clark *et al.* 2012). Briefly: *modA12* and its associated restriction gene (NMC1310 and NMC1311 respectively) were amplified from Fam18, and *modA12* was cloned individually and with the restriction gene into the plasmid pRRS (a pUC19 derivative). Both constructs were transformed into the non-methylating *E. coli* host ER2796. To ensure ModA12 was active prior to SMRT sequencing of transformed plasmids, crude DNA extracts from individual isolates were tested for incorporated tritiated methyl groups from S-adenosyl-methionine. Total genomic DNA from ER2796 cells carrying the plasmid construct was then isolated and SMRT sequencing and motif detection was performed as above, but with the ER2796 genome as the control.

Results

mod nucleotide variation

The PubMLST *Neisseria* database, which integrates storage of meningococcal whole-genome sequence (WGS) data with systems for cataloguing and comparing locus variation, permitted a survey of the prevalence and variation of three Type III methyltransferases (MTases) in 1,674 meningococci isolated from carriage and cases of disease (Table 1). The *modA* gene (NEIS1310) was identified in all isolates; in contrast, the *modB* (NEIS1194) and *modD* (NEIS2364) genes were found in 1,311 (78.32%) and 398 (23.78%) isolates respectively. No gene was present more than once in an isolate.

The overall structure of *mod* genes was: N-terminus tandem repeats (phase-variable repeat tract), followed by central highly variable regions (the DNA recognition domain (DRD)), flanked on either side by conserved regions (Fig. 1) (Srikhanta *et al.* 2009). Variation at DRDs grouped sequences into variants of each *mod* gene, described in previous work as 'alleles' (Srikhanta *et al.* 2009, Srikhanta *et al.* 2010, Srikhanta *et al.* 2011, Gawthorne *et al.* 2012, Jen *et al.* 2014). The mean nucleotide identity of DRDs within variants was 97.7%, 98.9%, and 99.9%, in contrast to between variants where it was 50.0%, 58.7%, and 49.3%, for *modA*, *modB*, and *modD* respectively (Fig. 1; Fig. 2).

Eight *modA* variants were identified including *modA2* and *modA6,* only previously observed in other bacterial species (Fig. 1A) (Gawthorne *et al.* 2012). Six *modB* variants were identified, including *modB5* and *modB6* that were identified for the first time here: nine (0.69%) *modB* positive isolates possessed *modB5* (50-60% DRD nucleotide identity to other variants) and one (0.08%) isolate possessed *modB6* (49-56% DRD nucleotide identity) (Fig. 1B; Fig 3B). Five *modD* variants were identified: *modD7* was identified for the first time (42-53% identity to other variants) and found in five (1.26%) *modD* positive isolates; modD3 was newly observed in the meningococcus (Fig. 1C; Fig. 2C).

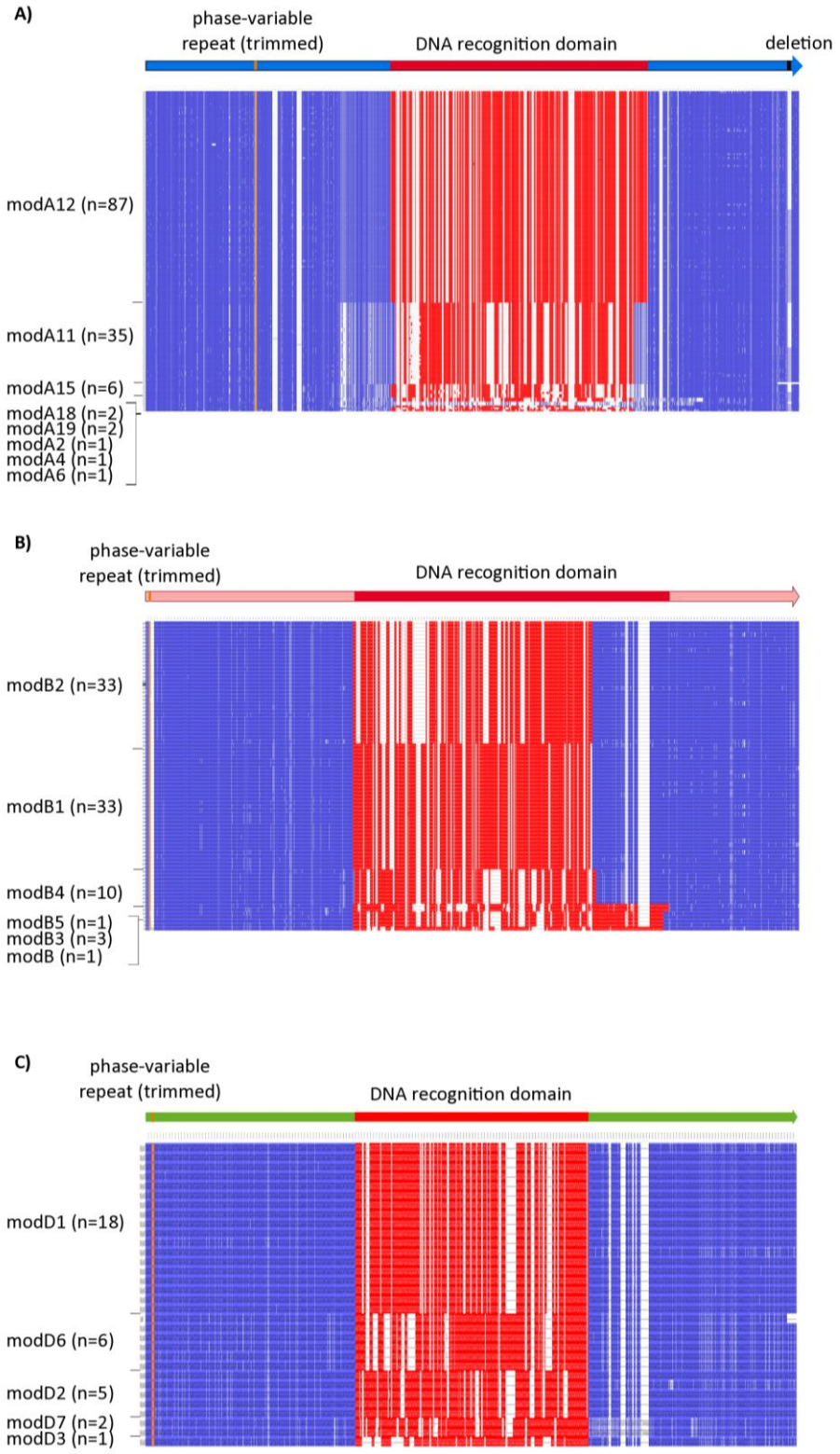
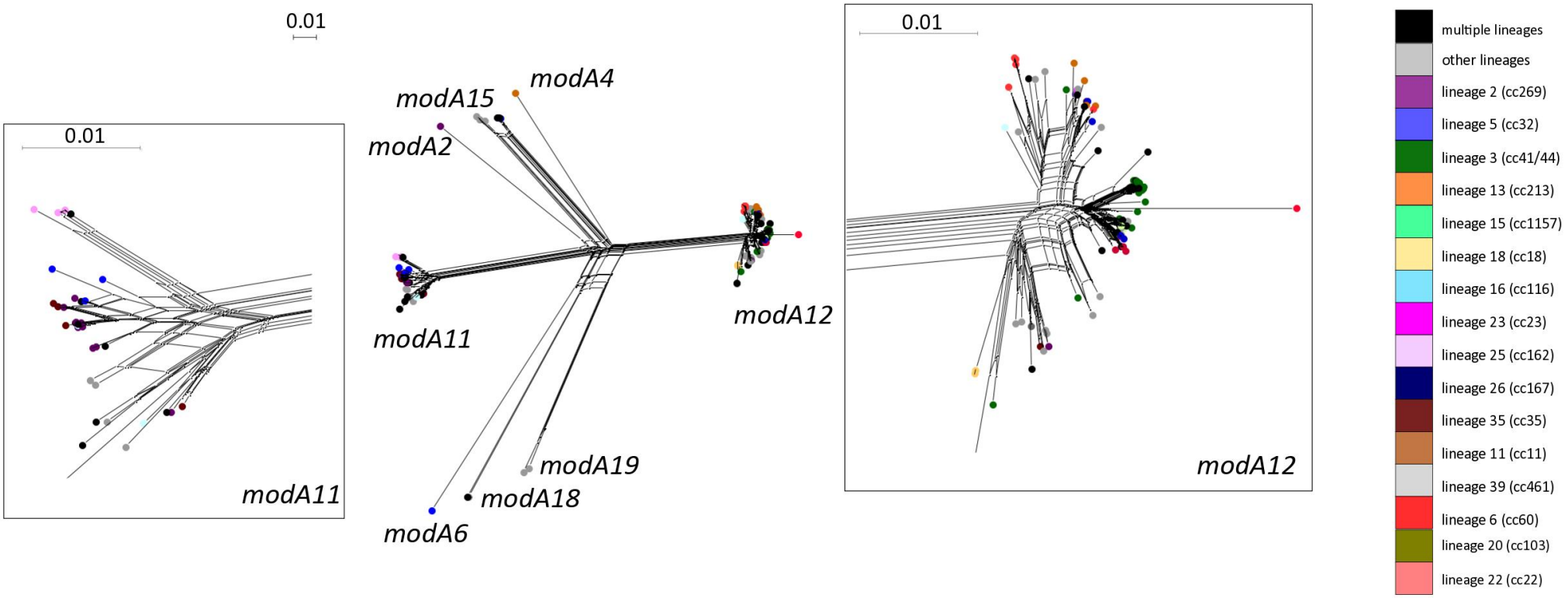
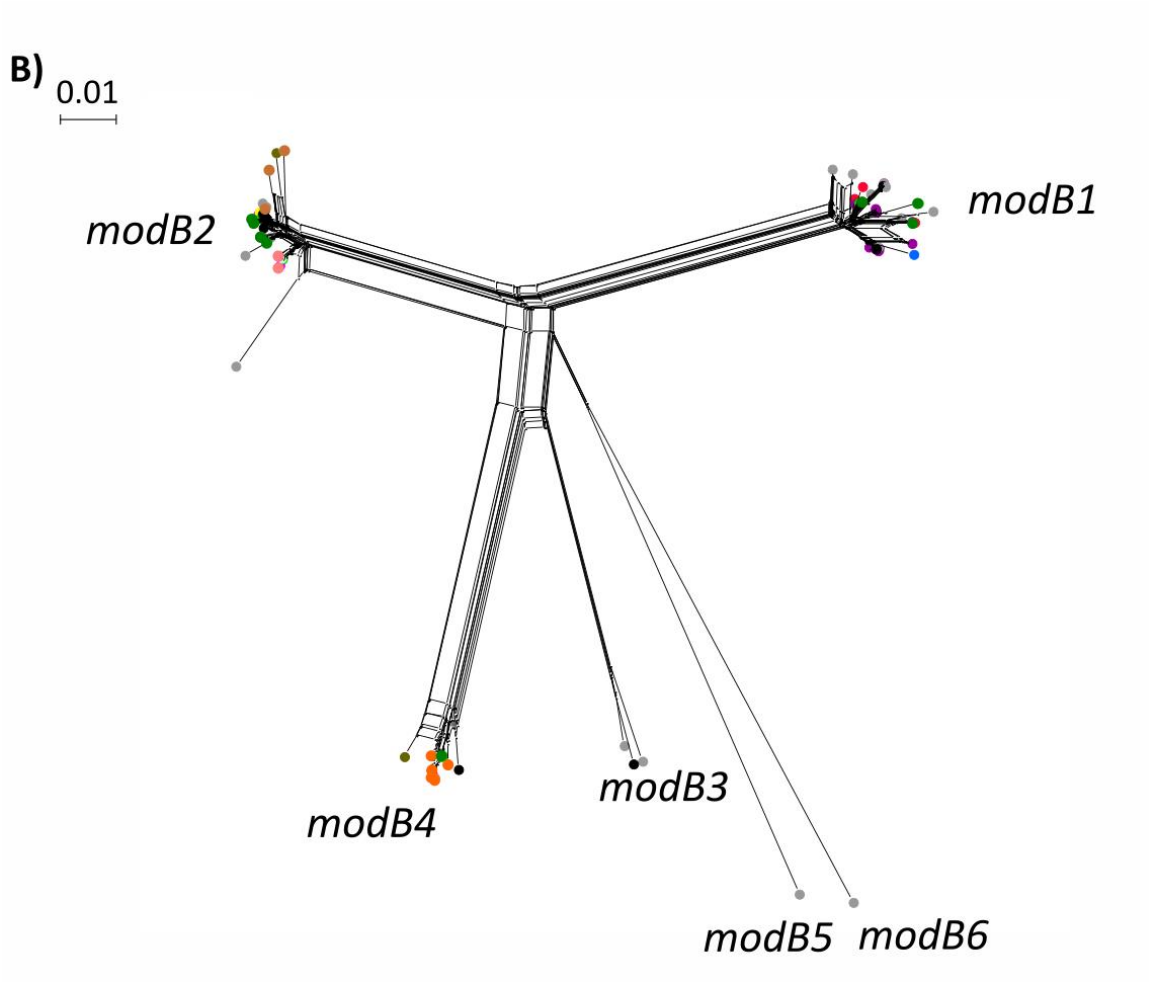


Figure 1. *mod* gene nucleotide diversity. Alignments of unique *mod* gene sequences. Relative positions of phase-variable repeats (orange), DNA recognition domains (red), and conserved residues (shaded by identity: dark blue >80% identity, light blue >50% identity, white <50% identity or a gap) are indicated. DRD-defined variants are labeled. All but the first phase-variable repeat unit were trimmed: n=number unique sequences following trimming. **A)** *modA* sequences with 15bp C-terminus deletion. **B)** *modB* sequence alignment. **C)** *modD* sequence alignment.

A)


0.01
modA15
modA4
modA2
modA11
modA12
modA19
modA18
modA6
0.01
modA11
0.01
modA12
multiple lineages
other lineages
lineage 2 (cc269)
lineage 5 (cc32)
lineage 3 (cc41/44)
lineage 13 (cc213)
lineage 15 (cc1157)
lineage 18 (cc18)
lineage 16 (cc116)
lineage 23 (cc23)
lineage 25 (cc162)
lineage 26 (cc167)
lineage 35 (cc35)
lineage 11 (cc11)
lineage 39 (cc461)
lineage 6 (cc60)
lineage 20 (cc103)
lineage 22 (cc22)


B)
0.01
modB2
modB1
modB4
modB3
modB5
modB6

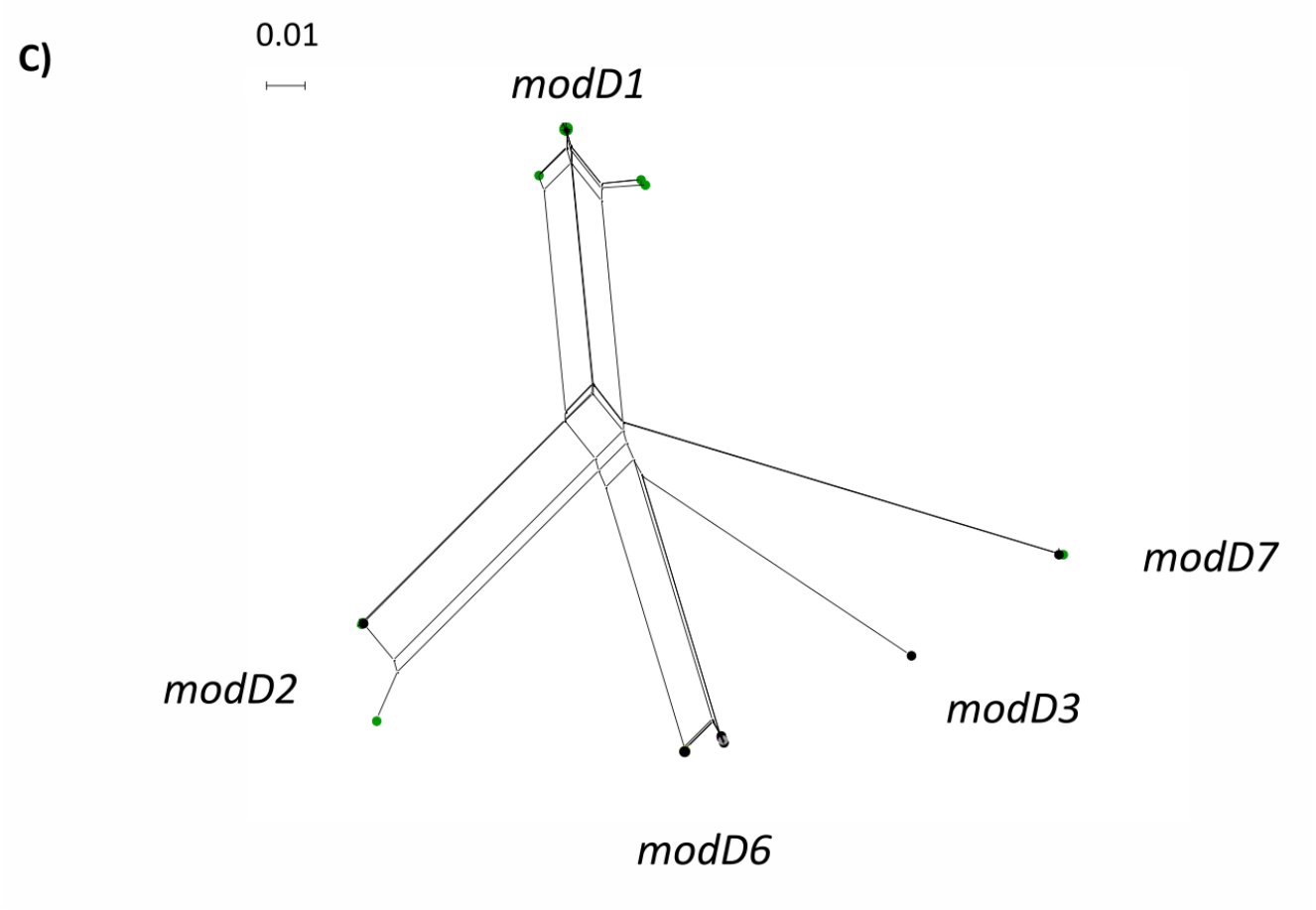


Figure 2. Neighbor-Net graphs of nucleotide *p*-distances among *mod* gene sequences. Unadjusted nucleotide *p*-distances were calculated from unique allele sequences of **A)** *modA*, **B)** *modB*, and **C)** *modD* in SplitsTree4. All but the first repeat unit of the phase-variable tract were trimmed from alignments. Lineages associated with trimmed allele variants indicated by colours: variants found in multiple lineages shown in black. Gene variants defined by the DNA recognition domain are labeled.

Variation to the highly conserved 3'-end of *modA* was observed. A 15bp deletion occurred in 683 (58.38%) *modA12* sequences, in 305 (67.03%) *modA11* sequences, in six *modA15* (15.79%) sequences, and in one (33.33%) *modA19* sequence (Fig. 1A). The biological significance of the loss of five amino acids from the C-terminus of the ModA12 protein was unclear. Additional probable truncations to ModA11 in C-terminal regions were identified in two isolates (premature stop codon caused by a C->A substitution) and also to ModD6 in four isolates (1bp deletion) (Fig. 1).

Although the phase-variable (PV) repeat tract was located at the 5'-end of all three *mod* loci, it was substantially further downstream of the start codon in *modA* (position 412bp) than in *modB* (position 12bp) or *modD* (position 21bp) (Fig.1). In addition, the *modA* PV repeat tract was composed of repeating tetranucleotide units unlike the pentanucleotide units of *modB* and *modD*. Two varieties of the *modA* repeat tract were found: 5'-AGCC-3' was predominant (1,643 (98.15%) *modA* sequences) but 5'-AGTC-3' was present in 28 (1.67% of *modA*) *modA11* and *modA19* sequences (Fig. 3A). *modB* also had two repeat tract varieties: 5'-CCCAA-3' was present in 1,028 (78.41%) *modB* sequences and 5'-GCCAA-3' was found in 280 (21.36%) *modB* sequences, including all *modB4* genes (Fig. 3b). The *modD* gene only possessed 5'-ACCGA-3' repeat tracts.

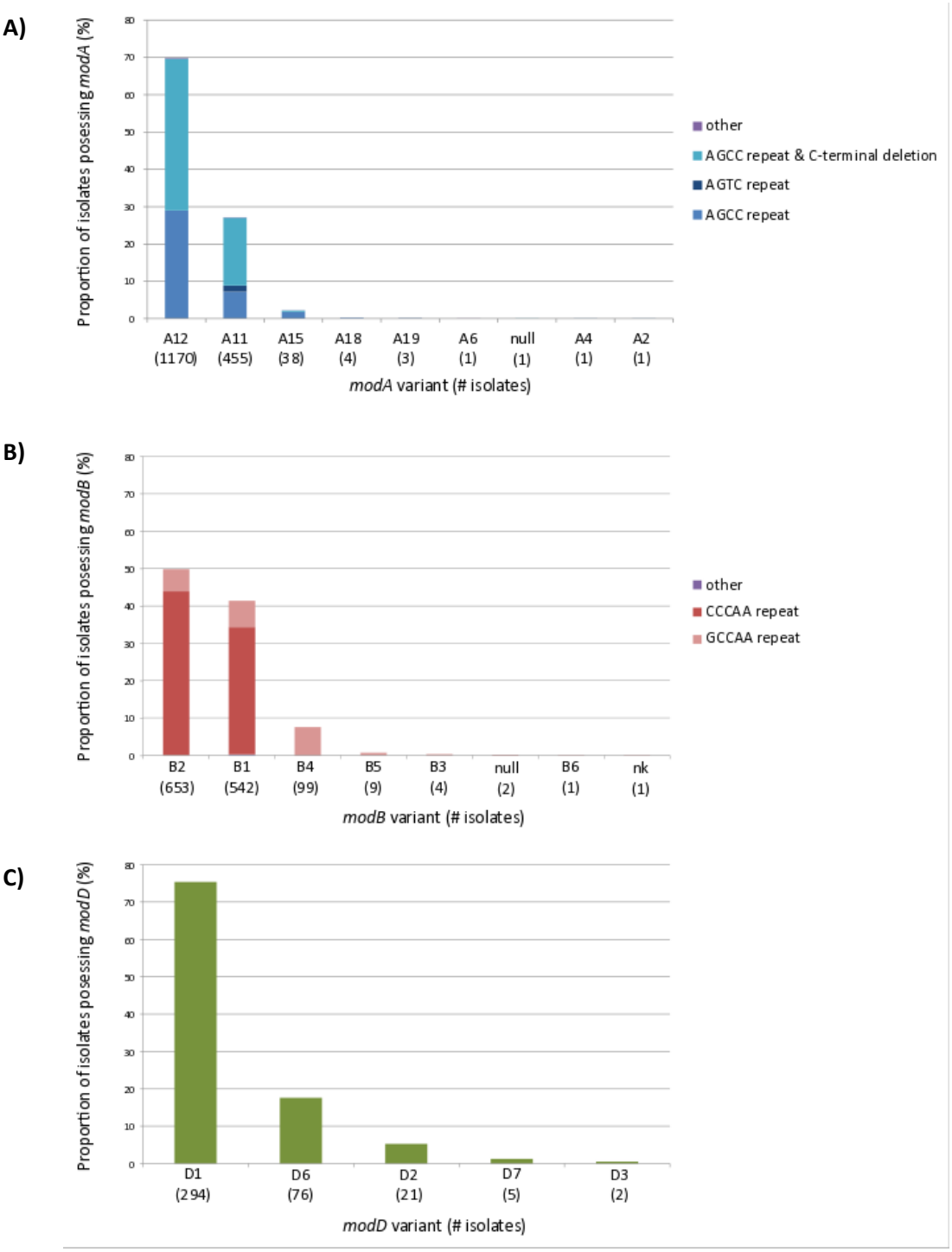


Figure 3. Variation to *modA*, *modB*, and *modD* coding sequences. A) Frequency of *modA* variants (defined by the DNA recognition domain (DRD)), phase-variable repeat types, and 15bp C-terminus deletion in 1,674 meningococcal isolates. **B)** Frequency of *modB* variants and phase-variable repeat type in 1,311 isolates. **C)** Frequency of *modD* variants in 398 isolates. 'other' refers to isolates for which there was no repeat and/or C-terminus information available. 'null' refers to isolates that possessed *mod* genes without DRDs. 'nk' refers to isolates without DRD information available.

Cataloguing mod variation in the PubMLST database

Unique gene sequences were uploaded to the PubMLST sequence definitions database for storage as alleles in the locus records NEIS1310 (*modA*), NEIS1194 (*modB*) and NEIS2364 (*modD*). Alleles were automatically assigned an arbitrary numeric identifier (allele id) and were manually assigned to a locus specific 'variant' (e.g. *modA12* in NEIS1310) defined by the DRD. Flags were manually applied to allele records to denote, for example, alleles containing frame shift mutations, atypical variation, or premature stop codons that probably encoded truncated Mod proteins.

DNA sequence repeats comprise a large proportion of the meningococcal genome. This presents some difficulty for algorithms assembling WGS reads into contiguous genome sequences (contigs). *mod* genes were frequently assembled onto two contigs (partially assembled): *modA* was partially assembled in 353 (21.09%) isolates, *modB* was partially assembled in 667 (50.77%) isolates, and *modD* was partially assembled in four (1.01%) isolates (Table 2). These sequences could not be added as alleles to PubMLST, however, variants could be identified since in all three *mod* loci, the majority of assembly breaks occurred in the PV repeat tract. Assembly breaks also occurred within insertion sequences integrated within the gene in 27 (7.65%) *modA* sequences**,** 30 (4.50%) *modB* sequences**,** and 1 (25%) *modD* sequence**.** This information was flagged in the relevant individual isolate records.

Table 2A-C. Prevalence and allelic diversity of *mod* genes.

A) *modA* variants

Variant	# isolates (% total)	# complete assemblies (% variant)	# PubMLST nucleotide alleles (%)	# unique trimmed alleles (%)	# unique trimmed peptides (%)	Diversity (*D*) [95% CI]
modA12	1170 (69.89)	948 (81.03)	255 (65.72)	87 (64.44)	71 (63.39)	0.85 [0.83, 0.87]
modA11	455 (27.18)	336 (73.85)	105 (27.06)	35 (25.93)	28 (25)	0.66 [0.61, 0.72]
modA15	38 (2.27)	27 (71.05)	19 (4.9)	6 (4.44)	6 (5.36)	0.71 [0.58, 0.84]
modA18	4 (0.24)	4 (100)	4 (1.03)	2 (1.48)	2 (1.79)	0.50 [0.07, 0.93]
modA19	3 (0.18)	2 (66.67)	2 (0.52)	2 (1.48)	2 (1.79)	1.00 [1.00, 1.00]
modA6	1 (0.56)	1 (100)	1 (0.26)	1 (0.74)	1 (0.89)	NA
modA4	1 (0.56)	1 (100)	1 (0.26)	1 (0.74)	1 (0.89)	NA
modA2	1 (0.56)	1 (100)	1 (0.26)	1 (0.74)	1 (0.89)	NA
null	1 (0.56)	1 (100)	na	na	na	NA
Total	**1674**	**1321 (78.91)**	**388 (100)**	**135 (100)**	**112 (100)**	**0.90 [0.89, 0.91]**

B) *modB* variants

Variant	# isolates (% total)	# complete assemblies (% variant)	# PubMLST nucleotide alleles (%)	# unique trimmed alleles (%)	# unique trimmed peptides (%)	Diversity (*D*) [95% CI]
modB2	653 (49.81)	321 (49.16)	96 (43.44)	33 (41.25)	32 (40.51)	0.75 [0.71, 0.78]
modB1	542 (41.34)	262 (48.34)	98 (44.34)	33 (41.25)	33 (41.77)	0.78 [0.74, 0.82]
modB4	99 (7.55)	50 (50.51)	19 (8.6)	10 (12.50)	9 (11.39)	0.30 [0.13, 0.47]
modB5	9 (0.68)	6 (66.67)	3 (1.36)	1 (1.25)	1 (1.27)	NA
modB3	4 (0.31)	4 (100)	4 (1.81)	3 (3.75)	3 (3.80)	0.83 [0.58, 1.08]
null	2 (0.15)	0	na	na	na	NA
modB6	1 (0.08)	1 (100)	1 (0.45)	1 (1.25)	1 (1.27)	NA
nk	1 (0.08)	0	na	na	na	NA
Total	**1311 (100)**	**644 (38.47)**	**221 (100)**	**80 (100)**	**79 (100)**	**0.90 [0.88, 0.91]**

C) *modD* variants

Variant	# isolates (% total)	# complete assemblies (% variant)	# PubMLST nucleotide alleles (%)	# unique trimmed alleles (%)	# unique trimmed peptides (%)	Diversity (*D*) [95% CI]
modD1	294 (73.87)	294 (100)	32 (45.07)	18 (56.25)	13 (52.00)	0.18 [0.11, 0.24]
modD6	76 (19.10)	72 (94.74)	22 (30.99)	6 (18.75)	5 (20.00)	0.38 [0.24, 0.51]
modD2	21 (5.28)	21 (100)	12 (16.9)	5 (15.63)	4 (16.00)	0.35 [0.09, 0.61]
modD7	5 (1.26)	5 (100)	4 (5.63)	2 (6.25)	2 (8.00)	0.40 [-0.03, 0.83]
modD3	2 (0.50)	2 (100)	1 (1.41)	1 (3.13)	1 (4.00)	NA
Total	**398 (100)**	**394 (23.54)**	**71 (100)**	**32 (100)**	**25 (100)**	**0.52 [0.46, 0.58]**

'Complete assemblies': *mod* coding sequences assembled on one contiguous sequence of the genome assembly. 'PubMLST alleles': unique allele sequences stored in the PubMLST sequence definitions database (http://PubMLST.org/neisseria/). 'Trimmed alleles': PubMLST alleles with all but the first repeat unit of the phase-variable tract removed. Diversity (*D*): diversity of trimmed alleles as measured by Simpson's Diversity Index with 95% confidence intervals (CI). na: not any. NA: not applicable. 'null': mod genes did not contain a DNA recognition domain (DRD) and could not be assigned to a variant. 'missing': gene absent from isolates. 'nk': variant not known (full DRD information not available).

Disruption to mod coding sequences

Completely assembled *mod* genes (Table 2) were apparently nonfunctional in the large majority of isolates. More than three quarters of sequences contained internal stop codons (ISCs). These were the result of the number of tandem repeats (n) in the phase-variable (PV) repeat tract in: 988 (74.79%) *modA* positive isolates; 406 (63.04%) *modB* positive isolates; and 376 (95.43%) *modD* positive isolates. In *modA*, n ranged from 2-34, in *modB* n ranged from 2-28, and in *modD* n ranged from 2-15: to be phase-variable 'on', where y was a positive integer, *modA* and *modD* required n=3y + 1 repeat units, whereas *modB* required n=3y repeat units. Alleles which were phase-variable 'off' were assigned the flags 'phase-variable: off' and 'frameshift' in the PubMLST database.

Many isolates harboured phase-variable 'on' *mod* genes that were not functional due to ISCs from other mutations, often within homopolymeric tracts (HTs). A different mutation was responsible in each of seven isolates with such *modA* alleles: two isolates possessed single thymidine (T) insertions in poly(T) HTs either side of the DRD, three isolates possessed 14-20bp indels between the PV tract and DRD, one had a 1bp deletion upstream of the PV tract, and one had a point mutation leading to a premature stop codon. In contrast, 129 isolates were identified to possess such *modB* alleles. This was due to one of two mutations causing probable termination of translation: 45 isolates had *modB1* alleles with a point mutation between the PV tract and the DRD and 83 isolates had *modB2* alleles with a single cytosine (C) insertion in a poly(C) HT located immediately downstream of the PV tract. Only two isolates contained such *modD* alleles: each contained frameshift mutations in the DRD due to indels, one being a single base insertion in a poly(A) HT. Phase-variable 'on' alleles containing ISCs because of indels were assigned the flags 'internal stop codon' and 'frameshift'; if the indel was >1bp they were also assigned the flag 'atypical'. Those containing ISCs due to point mutation were assigned the flag 'internal stop codon' only. Compensatory mutations were also identified in *modB:* phase-variable 'off' *modB2* alleles were identified to be complete coding sequences due to insertions to the aforementioned poly(C) HT in 20 isolates. Free text comments were manually entered to these allele records.

One of three insertion sequences (IS) interrupted 27 (1.61%) *modA* genes*,* 41 (3.13%) *modB* genes, and one (0.25%) *modD* gene. The element IS1301 was found: within the DRD in 18 *modA12* genes, 11 *modB1* genes, 8 *modB2* genes, and 1 *modB4* gene; downstream of the DRD in six *modA12* genes; and between the PV tract and DRD in one *modA15* gene. In all cases the consensus IS1301 target site was 5'-AYTAG-3' and the element was inserted in either orientation prior to the central 5'-TA-3' dinucleotide. ISNme3 was associated with *modB1*, interrupting 17 *modB1* genes within the DRD, although it was also observed in one *modA15* gene downstream of the DRD. Its target sequence was deduced to be 5'-ATT(ATTGAA)TTA-3' and insertion occurred in either orientation at the positions indicated by brackets. The element IS1655 was found in the C-terminal region of one *modA19* gene

and in the DRD of one *modB1* gene; its target sequence was 5'-GTTGC-3'. Three *modB1* genes and one *modD6* gene also appeared to be interrupted by an IS that had a second IS within it (ISNme3 and IS1301, and IS1655 and IS1301, respectively).

Prevalence and diversity of *mod* genes in meningococcal isolates

mod allelic diversity

The *modA* gene was universally present unlike *modB* and *modD*. Relative to *modB* (1,311 isolates), *modD* was rare (398 isolates). *modD* was the least diverse gene with a relatively small proportion of unique PubMLST alleles (18.02% of *modD* sequences were unique, compared to an average of 31.85% *modA* and *modB* sequences) (Table 2). To remove uninformative allelic variation produced by random reversible indels in the PV tract, all but the first PV repeat units were trimmed from *mod* alignments to create 'trimmed alleles' (Fig. 1). For each *mod* gene, this reduced the number of unique alleles by around 60%, resulting in 135, 80, and 32 unique trimmed alleles of *modA*, *modB*, and *modD* respectively (Table 2). As a consequence, *modA* was substantially less diverse relative to both *modB* and *modD* (1.02% of *modA* trimmed sequences were unique, compared to 12.42% of *modB* and 8.12% of *modD* sequences) (Table 2). In terms of deduced peptide sequences, *modB* was the most diverse since 79 of 80 alleles (98.75%) encoded a different peptide, compared to only 112 of 135 (82.96%) *modA* alleles and 25 of 32 (78.13%) *modD* alleles (Table 2).

The diversity of the genes among isolates was measured using Simpson's Diversity Index (*D*) on the trimmed alleles. By this measure, *modA* and *modB* were equally highly diverse (*D*=0.90, 95% CI=0.89,0.91) indicating the presence of a large number of rare variants (Table 2). *modD* was relatively conserved (*D*=0.52, 95% CI=0.46, 0.58) and this was because *modD* sequences were dominated by a single allele (Table 2). These patterns reflected the proportions of variants of each gene: single variants of *modA* and *modD* were dominant, with *modA12* and *modD1* each comprising around 70% of sequences, whereas *modB1* and *modB2* were almost equally frequent and present in >40% isolates each (Fig. 3). The remaining, minor, variants of each *mod* occurred in under 10% of

isolates (Fig. 3). In addition, one *modA* gene and two *modB* genes (both MRF-MGL, 2012, lineage 3 (cc41/44) isolates) had lost the central DRD region (referred to as 'null' variants); the remaining portions of these loci were as normal (Fig. 3).

The dominant variants of *modA* and *modB* (*modA12*, *modB1*, *modB2*) were significantly more diverse than others of the same gene (95% confidence intervals of *D* were non-overlapping), in contrast to the dominant variant of *modD*, *modD1*, which was less diverse (Table 2). In part, these patterns resulted from the distribution of the 34 meningococcal lineages in this study (Fig. 4). For instance, the greater the association of a lineage with a variant (Table 3), the less diverse the variant was (Table 2): e.g. *modD1* was almost exclusive to lineage 3 (293 of 294 *modD1* identified in lineage 3) and was of very low diversity (D = 0.18, 95% CI=0.11, 0.24). Despite the association of lineages with *mod* genes (Table 3), there was not clear evidence of vertical descent of *mod* genes within lineages (Fig. 2). The same alleles were found in multiple lineages (grey taxa in Fig. 2), there was little clustering of allele nucleotide diversity by lineage, and there was a great deal of homoplasy (reticulation) among alleles (Fig. 2). In addition, *modA* and *modB* genes within a lineage commonly had either of the two PV repeat tract types, and *modA* genes had either C-terminus type (Fig. 1) with the exception of *modA11* alleles with the repeat type 'AGTC' that were strongly associated with lineage 22 (cc22) (23 of 27 *modA11*:AGTC sequences).

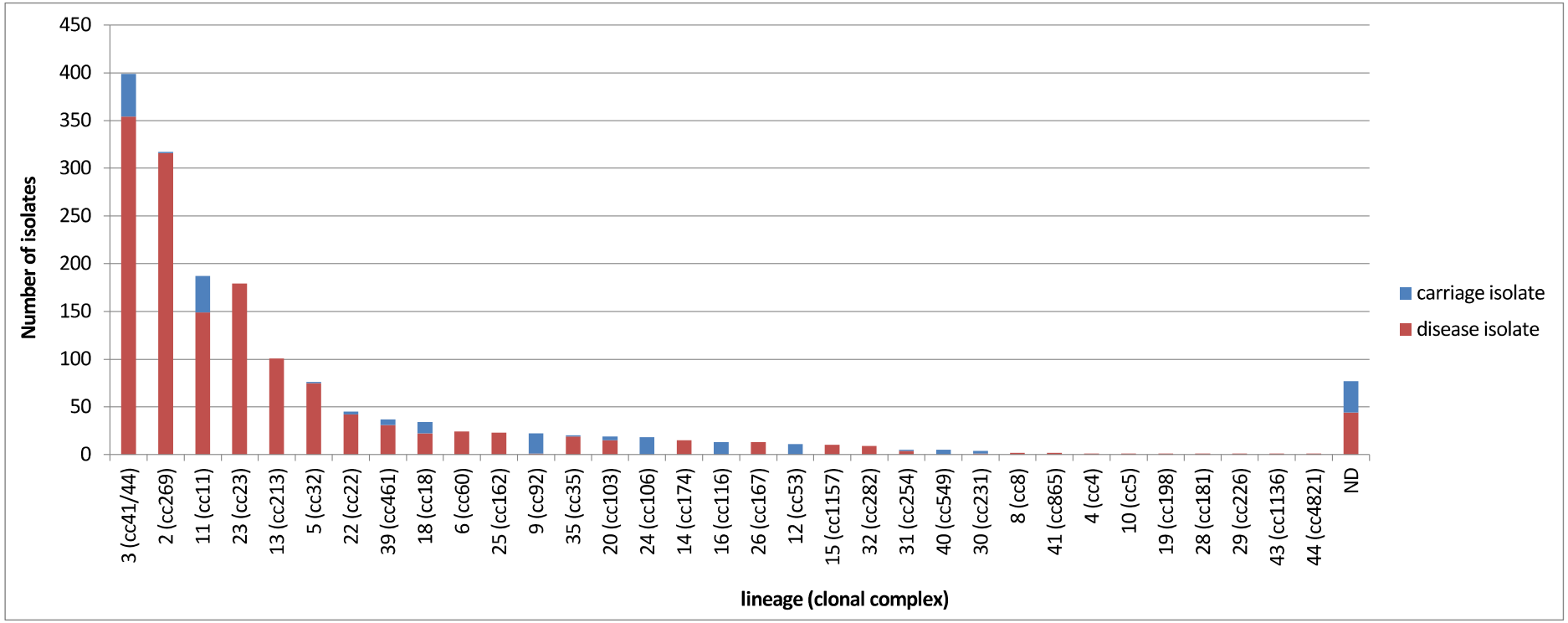


Figure 4. Meningococcal lineages of 1,674 isolates. The majority of disease isolates were from England, Wales, and Northern Ireland in the epidemiological years 2010/11-2012/13 (MRF meningococcus genome library; n=1380). The majority of carriage isolates were from the Czech Republic in 1993 (n=209), at the time of a lineage 11 epidemic. Additional isolates were obtained from a global collection (Budroni *et al.* 2011) and from the Southampton University 1997 cc11 meningococcal disease outbreak (Jolley *et al.* 2012b). Lineage ND: lineage not designated.

Table 3. Association of lineages with *mod* gene variants.

Gene	Variant	Associated lineage (# variant in lineage/total variant)	Odds ratio [CI], significance
modA	*modA12*	3 (395/1170)	63.59 [24.41, 234.94]***
	modA11	2 (300/455)	136.26 [80.87, 241.99]***
	modA15	9 (21/38)	1854.85 [278.20, 4.50e+15]***
	modA18	lineage ND (3/4)	64.00 [5.08, 3314.47]**
	modA19	na	na
	modA6	2 (1/1)	na
	modA4	12 (1/1)	na
	modA2	5 (1/1)	na
	null	13 (1/1)	na
modB	*modB2*	3 (307/653)	8.95 [6.84, 11.79]***
	modB1	2 (300/542)	81.04 [48.60, 143.37]***
	modB4	13 (91/99)	164.63 [77.21, 401.52]***
	modB5	32 (9/9)	na
	modB3	na	na
	null	3 (1/1)	na
	modB6	lineage ND (1/1)	na
	nk	3 (1/1)	na
	missing	23 (179/363)	Inf [334.88, inf]†
modD	*modD1*	3 (293/294)	3632 [608.80, 4.50e+15]***
	modD6	22 (44/76)	2251.43 [340.18, 4.50e+15]***
	modD2	3 (20/21)	88.07 [13.99, 3585.99]***
	modD7	3 (4/5)	2.88 [1.27, 633.56]*
	modD3	na	na
	missing	2 (316/1276)	130.95 [23.08, 5037.19]***

Associated lineage: lineage with most occurrences. CI: 95% confidence intervals. Significance from 2-tailed Fisher's exact tests: *** p-value<0.0001 **p-value<0.001 *p-value<0.0. †Contingency table contained 0 values (all lineage 23 missing *modB*) leading to infinite value. 'null': genes without DNA recognition domains. 'missing': gene absent from isolate. 'nk': variant not known. na: not applicable.

Association of mod genes with meningococcal lineages

The accessory mod genes, *modB* and *modD*, and variants of all three genes, were strongly associated with particular lineages. Single *modA* variants were observed in 15 lineages (with more than one isolate member) and otherwise one *modA* variant predominated (Fig. 5A). *modA12* was prolific and associated with all hyperinvasive lineages except for lineages 2 (cc269) and 5 (cc32), which were associated with *modA11*; all 15 isolates comprising sub-lineage 2 of lineage 2 (lineage 2.2) possessed *modA12*, however (Fig. 5A; Fig. 6A-B). *modA15* was associated with two carriage-associated lineages: lineage 9 (cc92), composed of serogroup W and Y isolates and isolates from carriage in the Czech Republic, and lineage 31 (cc285), although the latter association was weaker (Table 3; Fig. 5A). *modA15* was also associated with genogroup X carriage isolates without a designated lineage. Minor *modA* variants were dispersed among lineages (Fig. 5A).

modB was observed only in certain lineages (Fig. 5B). All hyperinvasive lineages possessed *modB* except lineages 23 (cc23) and 39 (cc461) within which no isolates had acquired the gene (Fig. 5B). Occasionally, single isolates of a *modB*-associated lineage were missing *modB*. Single variants predominated within *modB*-associated lineages, although distributions were more heterogeneous than for *modA* (Fig. 5). As with *modA,* however, all hyperinvasive lineages were associated with the same variant (*modB2*) except for lineages 2 and 5, which were associated with *modB1*. Lineage 2.1 was heterogeneous with 16 isolates, all from rMLST cluster 3, in possession of *modB2* (Fig. 5B; Fig. 6A-B). Similarly, lineage 3.2 of lineage 3 (cc41/44) contained mixed *modB* variants: 84 lineage 3.2 isolates had *modB1*; 10 carriage isolates from the Czech Republic had *modB2* in common with lineage 3.1; and five MRF-MGL isolates contained *modB4* (Fig. 5B; Fig. 6A; Fig. 6C). *modB4* was otherwise strongly associated lineage 13 (cc213) (Table 3). Hyperinvasive lineage 18 (cc18) also had mixed variants: all *modB1* isolates were from the Czech Republic collection and most *modB2* isolates were from the MRF-MGL (Fig. 5B; Fig. 6B). *modB5* was exclusive to lineage 32 (cc282) (Table 3; Fig. 5B).

modD was entirely absent from the majority of lineages (29 of 34) and even in lineages associated with the gene (lineages 3, 22, and 18), it was not universal (Fig. 5C). Particular lineage 3 sub-lineages were associated with the *modD* gene and its variants: *modD1* was associated with lineage 3.1; *modD* absence was associated with lineage 3.2; and where lineage 3.2 isolates possessed the gene, it was mainly variant *modD2*, although *modD6* and *modD7* were also present in lineage 3.2 isolates from the Czech Republic**.**

The existence of multiple *mod* genes, and of multiple variants of each gene, means that meningococci could have multiple independent phasevarions. Most lineages were associated with a particular *mod* combination that was shared by other lineages and where sporadic combinations occurred this was due to an altered *modB* or *modD* variant (Fig. 6A). Lineages sharing a major *mod* combination with other lineages were either: (i) lineages that did not possess the accessory *mod* genes; (ii) lineages associated with *modA12* and one of the major *modB* variants; or (iii) lineages that appeared to be particularly similar (Fig. 6A). For example, *'modA12*, *modB* missing, *modD* missing' was the prevalent combination in the non-hyperinvasive lineages 24 (cc106) and 26 (cc167) and in the hyperinvasive lineages 23 (cc23) and 39 (cc461), and **'***modA11*, *modB1*, missing**'** was associated with lineages 2 and 5, which are particularly similar as measured by rMLST (Fig. 6A). Lineage 3 did not share its major *mod* combination, '*modA12, modB2, modD1*', with another lineage. Lineages containing more than one prevalent combination, for example lineages 18, 2, and 3, tended to be diverse as measured by rMLST (Fig. 6A).

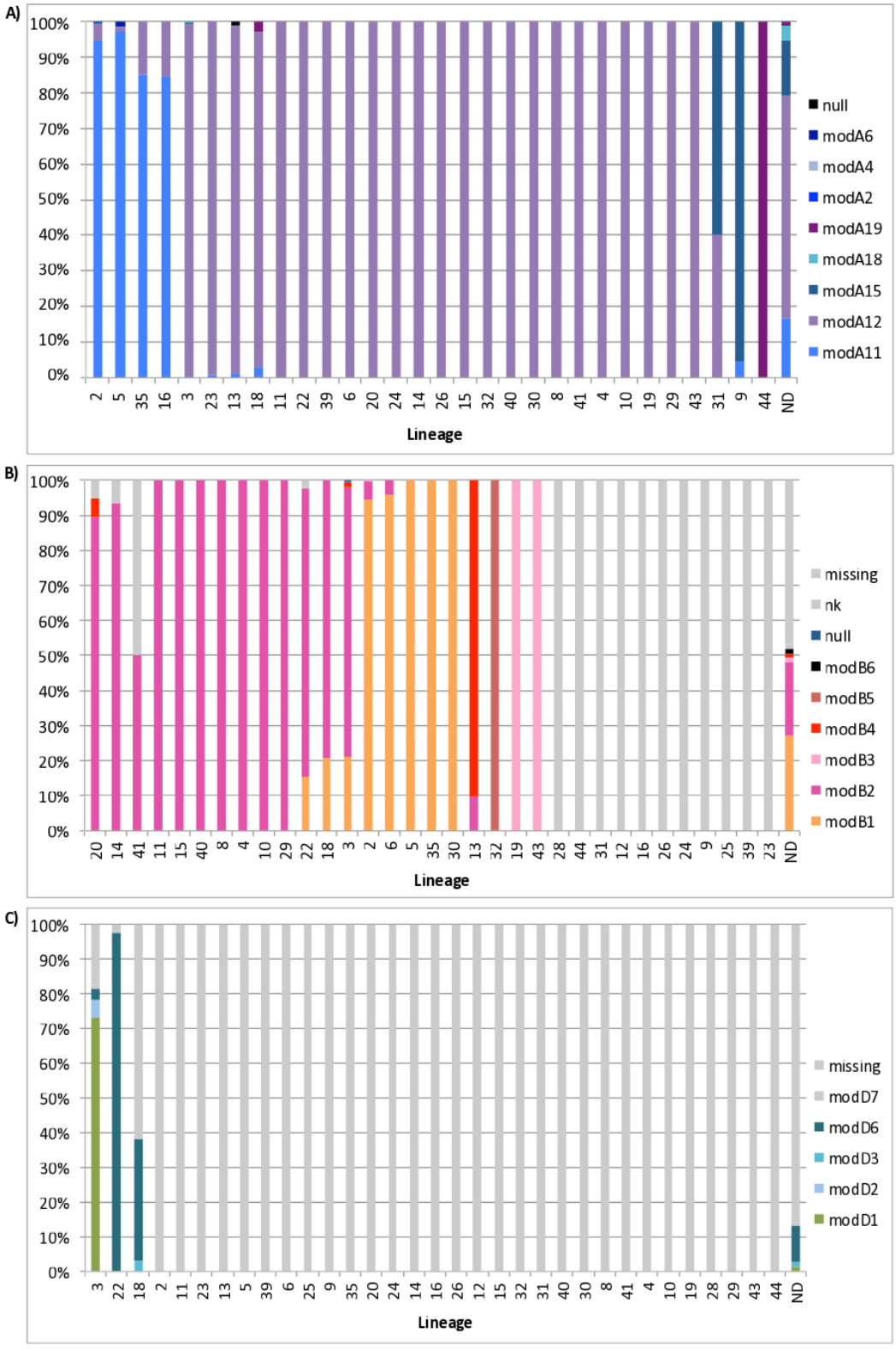
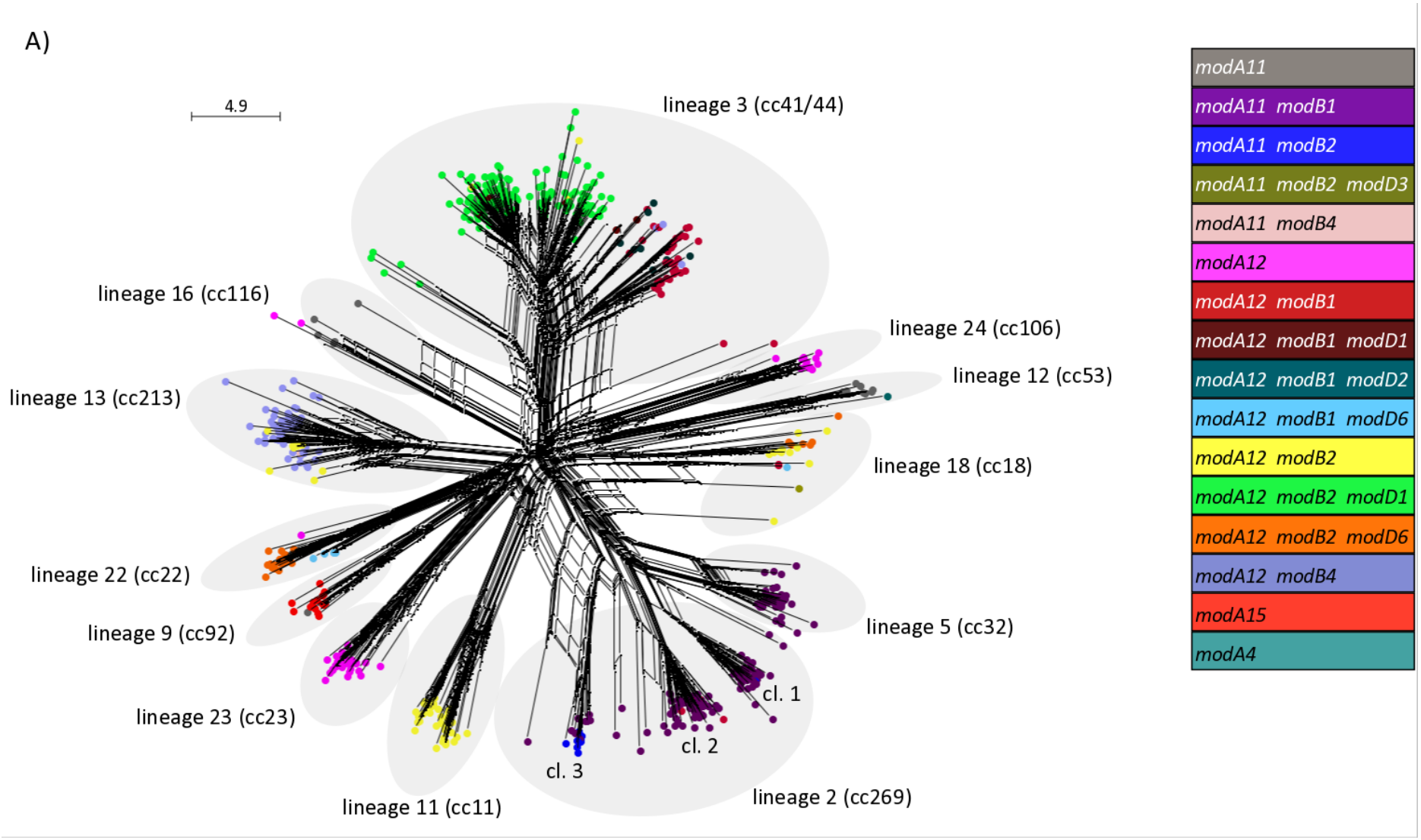


Figure 5. Prevalence of *mod* genes and variants among 1,674 isolates from 34 meningococcal lineages. A) *modA*. **B)** *modB*. **C)** *modD*. ND: lineage not designated. 'null': *mod* genes did not contain a DNA recognition domain (DRD) and could not be assigned to a variant. 'missing': gene absent from isolates. 'nk': variant not known (full DRD information not available).


A)
4.9
lineage 3 (cc41/44)
lineage 16 (cc116)
lineage 24 (cc106)
lineage 12 (cc53)
lineage 13 (cc213)
lineage 18 (cc18)
lineage 22 (cc22)
lineage 9 (cc92)
lineage 5 (cc32)
cl. 1
lineage 23 (cc23)
cl. 2
cl. 3
lineage 11 (cc11)
lineage 2 (cc269)
modA11
modA11 modB1
modA11 modB2
modA11 modB2 modD3
modA11 modB4
modA12
modA12 modB1
modA12 modB1 modD1
modA12 modB1 modD2
modA12 modB1 modD6
modA12 modB2
modA12 modB2 modD1
modA12 modB2 modD6
modA12 modB4
modA15
modA4

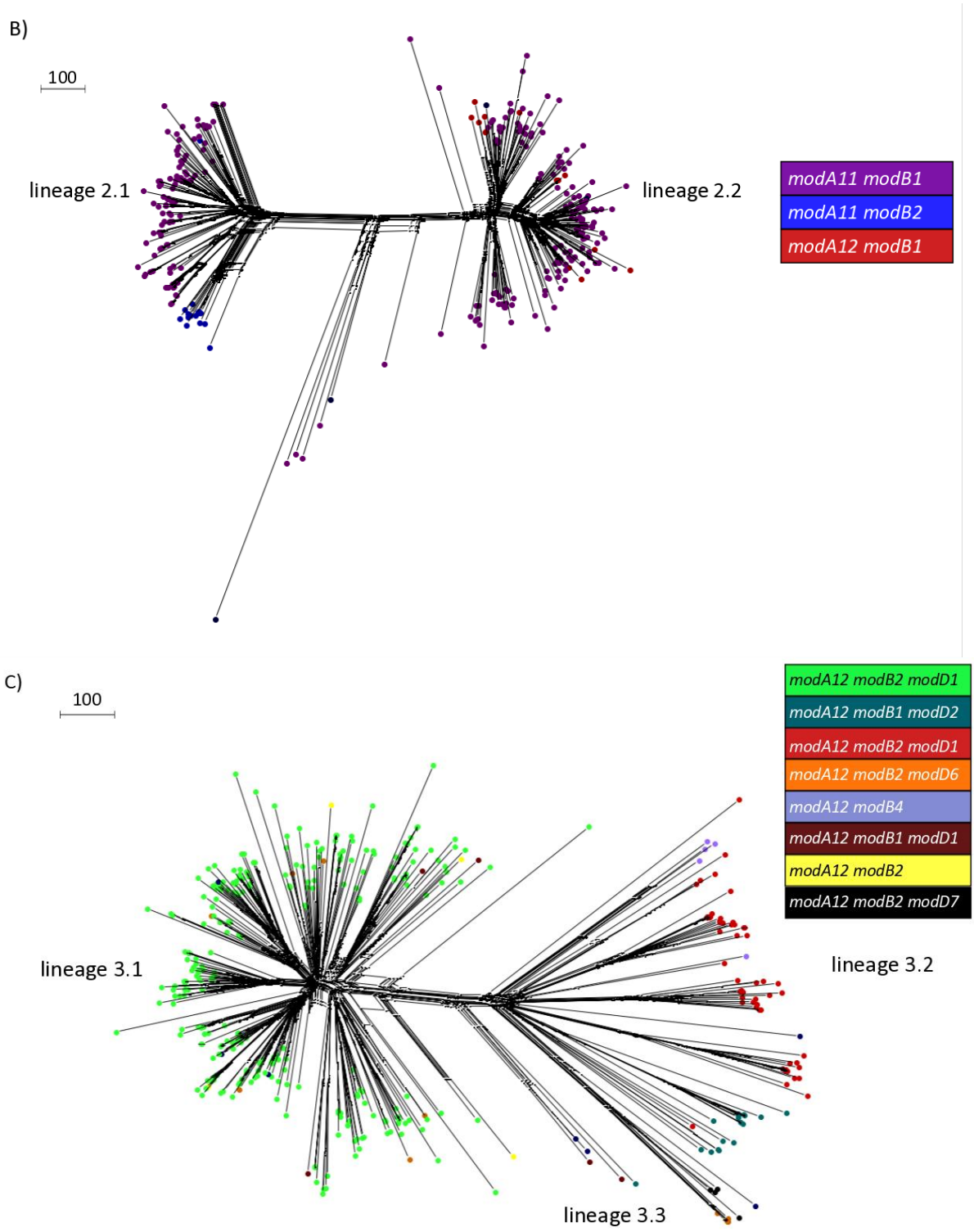


Figure 6. Associations between *mod* combination and meningococcal lineages and sub-lineages. A) rMLST Neighbor-Net graph of unique rSTs (n=639) from meningococcal lineages comprising >10 isolates. **B)** Neighbor-Net graph of allelic distances among core-genome loci (n=1,592) in lineage 2 isolates (n=317); *mod* combinations identified once only are coloured navy (n=4). **C)** Neighbor-Net graph of allelic distances among core-genome loci (n=1,592) in lineage 3 isolates (n=398); *mod* combinations identified once or twice in lineage 3 (n=4) are coloured navy. rSTs are coloured by *mod* combination unless associated with >1 combination (n=9).

Association of *mod* genes with meningococcal invasiveness

It is plausible that ModA, ModB, or ModD phasevarions affect the invasive potential of lineages (Srikhanta *et al.* 2009, Seib *et al.* 2011). Isolates belonged to 11 hyperinvasive lineages (n=1,248) and 23 non-hyperinvasive lineages (n=426); this variable, 'lineage invasiveness', was included in binomial logistic regression models to measure the odds that isolates possessing *mod* gene variants were members of hyperinvasive lineages relative to non-hyperinvasive lineages. All *modA* variants were significantly associated with either (Table 4). *modA12* and *modA11* were associated with hyperinvasive lineages, although the relative odds ratio (OR) was significantly greater for *modA12* (OR=2.85 [95% CI 2.47-3.20] and OR=1.75 [1.33-2.32] respectively). *modA15* was associated with non-hyperinvasive lineages (OR=0.031 [0.0073-0.086]), as were minor *modB* variants to a lesser degree (0.20 [0.053, 0.68]) (Table 4).

Associations between variants and infection outcome (disease relative to carriage) were also investigated, with separate multivariable regression models for each gene (Table 4). 'Lineage invasiveness' was included as a predictor to account for the confounding effect of lineage on the likelihood of disease; as expected, non-hyperinvasive lineages were significantly associated with carriage (OR<1), regardless of variant: e.g. in the *modA* model, OR=0.34 [0.25, 0.46]. Isolates in possession of *modA12* or *modA11* were significantly associated with disease, although *modA12* isolates were considerably more so (*modA12* OR=9.30 [7.53-11.62], *modA11* OR=1.99 [1.33-3.05]). The opposite was true for *modA15*, which was carriage-associated (OR=0.10 [0.05-0.21]). Minor *modA* variants were not significantly associated with either outcome (Table 4). The relationship between *modA11* and infection outcome was altered depending on whether it was found in hyperinvasive or non-hyperinvasive lineages: 'lineage invasiveness' and *modA11* had a significant interaction, with the association of *modA11* with disease greater in hyperinvasive lineages (OR=16.55 [6.18-67.58]; p<0.0001).

Absence of *modB* was significantly associated with non-hyperinvasive lineages (OR=0.26 [0.19-0.35]) (Table 4). Minor variants and *modB4* also had greater odds of being identified in non-hyperinvasive lineages relative to hyperinvasive lineages (OR=0.038 [0.0086-0.12] and OR=0.011 [0.0043-0.025] respectively). In contrast, *modB2* was strongly associated with hyperinvasive lineages (OR=5.68 [4.61-7.08]). Once the effect of hyperinvasive lineages was taken in to account, absence of the *modB* gene was associated with carriage (OR=0.49 [0.34-0.71]) and *modB2* was strongly associated with disease (OR=12.87 [9.79-17.23]). *modB4* also had greater odds of being observed in disease relative to carriage (OR=17.40 [5.21-108.14]); however, this variant interacted strongly with 'lineage invasiveness': in non-hyperinvasive lineages *modB4* had a high relative odds of being identified in disease (relative OR=15.25 [0.089-345.5]), whereas in hyperinvasive lineages it had an insignificant relationship with infection outcome (OR=0.53 [0.06- 4.58]). This probably resulted from the strong association of *modB4* with lineage 13 (cc213), which was rare among carriage isolates in this sample. The effect of *modB* absence on infection outcome also interacted significantly with 'lineage invasiveness': absence became associated with disease (OR=3.70 [1.57-8.73]) when found in hyperinvasive lineages, and with carriage (OR=0.11 [0.016-0.79]) in non-hyperinvasive lineages.

Unlike *modB*, there were high relative odds that hyperinvasive lineages would not possess *modD* (OR=2.45 [2.17-2.76]). Hyperinvasive lineages were likely to possess minor variants (OR=5.32 [1.58-33.11]) and *modD1* (OR=119.85 [26.91-2108.7]) (Table 4). The *modD6* variant of the gene was associated with non-hyperinvasive lineages (OR=0.19 [0.11-0.31]). Once invasiveness of lineages was accounted for, absence of *modD* was twice as likely to be found in disease-causing than carried meningococci (OR =2.82 [2.26-3.56]) and minor variants were more likely to be found in carried meningococci (OR=0.12 [0.05-0.26]). The effect of *modB6* also interacted with 'lineage invasiveness': it was relatively associated with disease in non-hyperinvasive lineages (OR=3.07 [1.37-8.20]) and with carriage in hyperinvasive lineages (OR=0.11 [0.014-1.00]).

The presence of mixed *modB* and *modD* variants in lineage 3 (Fig. 6C) permitted an investigation into whether the combination of variants had an effect on the outcome of infection within a lineage. A binomial regression model for the effect of *mod* combination on outcome was not statistically significant. However, the different sub-lineages were differently associated with disease. Lineage 3.1 (ST-41 associated) contained mainly '*modA12*, *modB2*, *modD1'* and was significantly associated with disease relative to carriage (OR=300 [67.8-5270]). Lineage 3.2 (ST-44 associated), with a diversity of *mod* combinations, was significantly associated with carriage (OR=0.0037 [0.00021-0.018]) (Fig. 6C).

Table 4. Association of *mod* genes and variants with hyperinvasive lineages and meningococcal disease.

Gene	Variant	Hyperinvasive lineages		Non-hyperinvasive lineages		Relative OR HI:non-HI (CI)†	Relative OR disease:carriage (CI)‡
		Carriers	**Patients**	**Carriers**	**Patients**		
modA	*modA12*	99	599	59	415	2.81 (2.47-3.19)***	9.30 (7.53-11.62)***
	modA11	3	374	27	51	1.75 (1.33-2.32)***	1.99 (1.33-3.05)**
	modA15	0	2	28	8	0.03 (0.01-0.09)***	0.10 (0.05-0.21)***
	minor	0	6	1	6	0.20 (0.05-0.68)*	2.37 (0.44-44.04)
modB	*modB2*	56	516	11	72	5.68 (4.61-7.08)***	12.87 (9.79-17.23)***
	modB1	44	423	11	64	1.10 (0.79-1.52)	0.98 (0.67-1.45)
	modB4	2	5	0	92	0.01 (0.00-0.02)***	17.40 (5.21-108.14)***
	missing	0	32	93	238	0.26 (0.19-0.35)***	0.49 (0.34-0.71)***
	minor	0	3	0	14	0.04 (0.01-0.12)***	NA
modD	missing	78	660	108	432	2.45 (2.17-2.76)***	2.82 (2.26-3.56)***
	modD1	0	293	0	1	119.85 (26.91-2108.70)***	NA
	modD6	11	13	6	46	0.19 (0.11-0.31)***	0.92 (0.52-1.69)
	minor	1	1	13	13	5.32 (1.58-33.11)*	0.12 (0.05-0.26)***

Meningococci were grouped into hyperinvasive (HI) and non-hyperinvasive lineages for logistic regression models that measured the association of *mod* variants with: HI lineages relative to non-HI lineages (†); and disease relative to carriage whilst taking into account the effect of hyperinvasive lineages (additive model) (‡). OR=odds ratios relative to first variant listed for each gene. CI=95% confidence intervals. Statistical significance is shown as p-values: *** p<0.0001, **p<0.001, *p<0.01. 'NA': ORs not applicable since all isolates were from disease.

modA12 epigenetic activity

Identification of Mod DNA recognition sites is a precursor to identifying the genes they regulate. Pacific Biosciences SMRT methylome analysis was used to identify methylated DNA motifs within the genomes of four isolates belonging to lineage 11 (Table 5); this lineage had an absolute association with disease-associated variants *modA12* and *modB2* for which recognition sites have not been elucidated (Fig. 5A). Two different methylation patterns were identified (Table 5): Czech Republic genomes had six consistently methylated DNA motifs and Southampton genomes had five different consistently methylated DNA motifs. There was no difference between disease and carriage isolates.

Table 5. Methyltransferases (MTases) and recognition sites in four lineage 11 isolates.

Motif	Methylation	# modified motifs per isolate (%)				MTase name [locus]
		CR disease	CR carriage	SH disease	SH carriage	
5'-G**A**TC-3'	m6A	4,454 (99.8)	4,584 (95.8)	4,510 (98.9)	4,697 (99.5)	ND (M1.NmeM393V or M2.NmeM393V)
5'-GA**A**GG-3'	m6A	4,561 (99.8)	4,730 (96.2)	4,643 (98.6)	4,807 (99.6)	*M.NmeM393ORFIP (**ModB2**) [NEIS1194] or M.NmeM393ORFKP [NEIS0771]
5'-AC**A**CC-3'	m6A	4,056 (99.8)	4,122 (95.0)	4 (0.1)	6 (0.1)	M.NmeM393IV (**ModA12**) [NEIS1310]
5'-G**C**GCGC-3'	m5C	192 (49.2)	142 (36.8)	222 (58.4)	230 (59.6)	M.NmeM393II [NEIS1520]
5'-C**C**GGY-3'	m5C	756 (6.2)	487 (3.8)	1348 (11.0)	1279 (10.3)	M.NmeM393III

Methylated motifs in two Czech Republic (CR) and two Southampton (SH) isolates were identified by Pacific Biosciences SMRT genome sequencing: bases in bold are methylated on the top strand, underlined bases are methylated on the bottom strand. 'm6A': 6-methyladenosine; 'm5C': 5-methylcytosine. Methyltransferases (MTases) were assigned to methylated motifs (deduced recognition sites) through homology to enzymes with known recognition sites, for ModA12 confirmed by cloning experiments. ND: Not determined since two candidate enzymes. *Cloning required to distinguish activity of ModB2 from that of Type II fusion MTase (homology to NEIS0771). Dark to light grey shading indicates high to low proportions of motifs modified in a genome.

Methyltransferase (MTase) genes were bioinformatically predicted in genomes and cross-referenced with methylated DNA motifs to derive their recognition sites (Table 5). The motifs containing methylated cytosine residues (Table 5) were deduced to be the recognition sites of two predicted MTases on the basis of sequence homology to published cytosine-methylating enzymes. Recognition sites of the enzymes with m6A activity could not be deduced from these data (Table 5). First, there were two partner Type II MTases in the genomes that were likely to recognise the sequence GATC and neither partner could be unambiguously assigned to the motif 5'-G**A**TC-3' (Table 5). Second, although the two pentanucleotide motifs containing m6A modifications (5'-GA**A**GG-3' / 3'-CTTCC-5' and 5'-AC**A**CC-3' / 3'-TGTGG-5') were likely to result from activity of Type III enzymes, there was a third MTase in the genomes also likely to recognise GAAGG: this was a bifunctional MTase that modifies this sequence on both strands with either an m6A or an m5C modification (i.e. 5'-GA**A**GG-3'/3'-CTT**C**C-5' or 5'-C**C**TTC-3'/3'-GG**A**AG-5'). Since m5C modification is not detected by SMRT sequencing and since *modB2* was 'phase-variable: off' in the genomes, 5'-GA**A**GG-3' could not be unambiguously assigned to either the bifunctional MTase or to ModB2 without experiments to assess the expression and activity of these enzymes.

The final methylated motif, 5'-AC**A**CC-3', was likely to be ModA12's recognition site: there were no other candidate enzymes and *modA12* was 'phase-variable: off' in the Southampton genomes and 'phase-variable: on' in the Czech Republic genomes, in accordance with the proportion of motifs modified in the genome (>95% and 0.1% respectively) (Table 5). Cloning experiments confirmed that ModA12 activity is phase-variable, that its recognition site is ACACC, and that methylation is comprehensive (average 4089 of 4202 sites, >97%) (Table 5). Most of the 1,033 meningococcal coding sequences containing ModA12 modifications were involved in metabolism; when compared to the functional composition of the lineage 11 reference genome, Fam18 (Bentley *et al.* 2007), genes functioning in metabolism of co-factors, translation and transcription, replication and repair, membrane transport, cell growth/death, and cell motility were over-represented among modified genes. This suggested that a wide range of phenotypic alterations could result from ModA12 activity.

Discussion

Restriction-modification systems (RMS) may maintain meningococcal lineage structure and phase-variable methyltransferases (Mods) of Type III RMS mediate alterations in expression of multiple genes. The wealth of information provided by whole-genome sequence (WGS) data is ideal for comparing gene content among meningococcal lineages; epigenetic DNA modifications in PacBio WGS data provides additional genetic resolution among meningococci that may have distinct phenotypes. This study is the largest survey of the prevalence and sequence diversity of *mod* genes in any bacterium. Using the PubMLST database, *mod* gene variation was indexed in WGS data from 1,674 meningococcal isolates.

Nucleotide diversity across *mod* genes was similar to previous studies, with highly variable DNA recognition domains (DRD) within each gene (Fig. 1) (Fox *et al.* 2007, Srikhanta *et al.* 2009, Gawthorne *et al.* 2012). In the past, DRDs have been categorised into 'groupings' and 'alleles' (Bayliss *et al.* 2006, Fox *et al.* 2007) with 20 *modA* alleles (*modA1-modA20*), four *modB* alleles (*modB1-modB4*), and six *modD* alleles (*modD1-modD6*) described to date. The PubMLST sequence definitions database categorises unique gene sequences as alleles, therefore, in this study unique sequences of each *mod* gene are referred to as alleles and alleles are grouped into variants based on their DRDs.

The particular gene sets affected by each Mod and their regulatory mechanisms have not been fully explored (Srikhanta *et al.* 2005, Fox *et al.* 2007). This is partly because functional studies are limited by unknown Mod recognition sites. Plasmid restriction experiments in *Haemophilus influenzae* demonstrated that the DRD confers recognition site specificity and microarrays of *N. meningitidis* strains possessing *modA11*, *modA12*, and *modD1* showed regulation of different gene sets compared to knockout mutants, but they were unable to demonstrate that these were the direct effects of the Mod enzyme (Humbelin *et al.* 1988, Bayliss *et al.* 2006, Srikhanta *et al.* 2009). Here, a

combination of SMRT sequencing and bioinformatic analyses showed that ModA12 recognises the non-palindromic sequence 5'-ACACC-3', methylating the central adenine on one strand only (Table 5) (Seib *et al.* 2015a). This was typical of other known Type III methyltransferase recognition sites: ModD1 recognises 5'-CC**A**GC-3' and ModB1 recognises 5-CC**A**CC-3' (Adamczyk-Poplawska *et al.* 2009, Seib *et al.* 2015a). ModA12 activity was directly proven to be phase-variable (PV) as a result of 5' tandem repeats and when in-frame, ModA12 activity was comprehensive with >95% of ACACC motifs modified, indicative of a specific role for Mod enzymes in the meningococcus. The isolates investigated also possessed ModB2, however, cloning experiments would be required to distinguish the activity of this enzyme from that of a Type II MTase that modifies the same sequence (Table 5).

As expected from a short DNA sequence motif, the ModA12 recognition site was widespread in the lineage 11.2 (ET-15) genome and most modified coding sequences were metabolic loci in line with the general composition of meningococcal genomes (Tettelin *et al.* 2000, Bentley *et al.* 2007, Bratcher *et al.* 2014). Of particular interest was the increased proportion of modifications in genes functioning in co-factor metabolism. These included *brfB*, whose product, Bacterioferritin, is involved in iron storage and oxidative stress responses and whose expression was increased in a ModA12 deficient meningococcal strain compared to a ModA12 expressing strain; this suggests that modification of *brfB* leads to lower expression of Bacterioferritin than otherwise (Perkins-Balding *et al.* 2004, Seib *et al.* 2004, Srikhanta *et al.* 2009). Genes involved in a wide variety of essential cellular functions were also over-represented among modified genes, suggesting that populations of cells in possession of *modA12* have either of two rather different phenotypes depending on the expression status of the enzyme. Future bioinformatic work should assess ModA12 methylation within intergenic and promoter regions since studies have found that only a quarter of modifications occur in annotated coding sequences, that methylation sites occur upstream of differentially expressed genes, and that ModA12 phasevarions include only 26 genes, far fewer than the 1,033 modified coding sequences here (Srikhanta *et al.* 2009, Fang *et al.* 2012, Seib *et al.* 2015a).

As anticipated for a functional domain, the DRD was highly conserved within variants of all three genes, appearing to be under negative selection (Fig. 1; Fig. 2). Despite this, a number of highly diverse variants of each *mod* were identified including three novel variants, *modB5*, *modB6*, *modD7*, (Fig. 2). Variants are likely to regulate the expression of particular gene sets and the existence of three different *mod* genes, each with a range of variants, is indicative of diversifying selection on alternative phase-variable phenotypes. Since variant distributions among lineages were heterogeneous, the almost universal prevalence of *modA12* suggests that ModA12-mediated phenotypes are selectively maintained (Fig. 5). There was also evidence of negative selection in the N- and C- terminal regions of *mod* genes from the lack of allelic diversity within variants (Table 2; Fig. 2). In contrast to *modA* and *modD*, however, almost every *modB* allele encoded a distinct peptide (Table 2), which suggested either recent, rapid proliferation of *modB* and concomitant mutation events not yet acted on by selective processes, or synonymous changes fixed by diversifying selection.

Longer PV tracts in *mod* are associated with increased switching rates (De Bolle *et al.* 2000). Here, PV tracts varied greatly in length, with up to 34 repeat units in a single tract. Combined with the additional homopolymeric tract frameshift mutations and insertion elements observed, the rate of switching between phase-variable 'on' and 'off' Mod-mediated phenotypes is likely to be great. In accordance with previous observations of a 1.75 fold increase in *modA* PV tract indels creating 'off' phenotypes (De Bolle *et al.* 2000), the repeat unit frequencies observed here (e.g. to be 'on' $n=3y+1$, where y is a positive integer, but $1<n<35$) predicted an increased likelihood of PV 'off' genes, and most isolates possessed out-of-frame *mod* sequences. Two different PV tract types were observed within *modA* and *modB* sequences (Fig. 3): it is likely that PV tracts evolved independently at each locus since there were no observations of mixed type tracts. Insertion elements were frequently inserted near the PV tract and it is plausible that duplication of initial tetra- or penta-nucleotide sequences by insertion leads to repeat tract evolution (Zhu *et al.* 2000, Moxon *et al.* 2006). There were no occurrences of *mod* sequences with fewer than two repeat units, indicating that phase-

variation is selectively maintained despite increased mutation rates in this region and wide-ranging phenotypic consequences. Observations such as these are important for understanding the role of PV Type III RMS within meningococci. Studies of switching rates may provide insights into why *mod* genes are phase-variable and into the selective advantages of particular switching rates: for example, to aid survival in variable environments, such as on invasion of the blood stream, the rate of switching should theoretically be greater than the rate of within-host environmental change (Moxon *et al.* 2006, Bayliss 2009, Palmer *et al.* 2013).

Proliferation of *mod* genes among meningococci appeared to have occurred via horizontal gene transfer (HGT) from other bacterial species and among lineages. The variants *modA2*, *modA6*, and *modD3* were originally discovered in *H. influenzae* and are here reported in the meningococcus for the first time. Other variants had been observed in additional bacterial species that colonise the upper respiratory tract (Fox *et al.* 2007, Srikhanta *et al.* 2009, Gawthorne *et al.* 2012). The sequence conservation of the DRD and the range of loci affected by ModA12 activity suggested that these enzymes are selective for niche specific functions. It is likely that *modA* was acquired before diversification of *Neisseria* from its common ancestor given that this gene has spread to fixation in meningococcal population, and that it has been identified in bacteria from other genera (Fox *et al.* 2007). *modB* is likely to have been acquired after the diversification of *Neisseria*, since this gene has only been observed in *Neisseria* species and is not universally present in the meningococcus. *modD* is likely to be relative recently acquired; it was not fixed in any lineage, it was identical in the majority of isolates (Table 2), and the gene was previously only observed in cc41/44, although here it was identified in two further lineages (Seib *et al.* 2011).

Multiple variants were often seen in the same lineage and the same variant was often shared by isolates of different lineages (Fig 5). This, and the lack of similarity among variants (~50% identity), indicate horizontal acquisition of variants rather than convergent evolution within lineages. The data presented here, in agreement with previous studies, indicate that the DRD is transferred as a unit

among *mod* locus backgrounds (Gawthorne *et al.* 2012). First, the majority of insertion elements were inserted within the DRD. The hypothesis that RMS are selfish elements was prompted by studies showing chromosomal associations between mobile elements and RMS and by experimental observations of linked movement of insertion sequences and RMS (Kobayashi 2001, Fukuda *et al.* 2008, Furuta *et al.* 2010, Takahashi *et al.* 2011); a role for insertion elements in the transfer of genetic material between *N. meningitidis* and *H. influenzae* has also been described (Kroll *et al.* 1998). Second, several isolates possessed 'null' *modA* and *modB* loci that were missing the DRD (Fig. 3); this is unlikely to have resulted from assembly errors since *mod* genes were not paralagous. Third, distinct *modA* and *modB* variants often shared a PV tract type and C-terminus type, despite different tracts and C-termini within variants (Fig. 3); HGT of the DRD seems more likely than convergent evolution of tract types and repeated 15bp deletion/insertion at the C-terminus. Finally, there was evidence for HGT in the regions of the gene outside the DRD: homoplasy (network reticulation) and lack of allelic clustering was observed within *modA* and *modB* variants (Fig. 2), which, combined with previous characterisation of DNA uptake sequences in the N- and C-terminal regions of *mod* genes suggests that recombination occurs at breakpoints either side of the DRD (Kroll *et al.* 1998, Bayliss *et al.* 2006, Gawthorne *et al.* 2012). Multiple DRD transfers are likely to have occurred within each gene, especially within *modB* whose variant distribution was particularly heterogeneous among lineages (Fig. 5).

Despite the evidence for frequent HGT, the association of *mod* variants with meningococcal lineages was non-random and most lineages were associated with single variants (Fig. 5; Table 3). There was some evidence of association between variants and population structure: although relationships among meningococcal clonal complexes are difficult to deduce since phylogenetic signal is obscured by high rates of recombination at these evolutionary distances (Didelot *et al.* 2009), relatedness of cc11 (lineage 11) and cc8 (lineage 8), and of cc269 (lineage 2) and cc32 (lineage 5), is consistently reconstructed and these lineages did mostly share the same *mod* variants (Fig. 5A) (Budroni *et al.* 2011). In addition, lineages 2 and 5 consistently possessed different variants to lineage 3 (cc41/44),

suggesting that the acquisition of alternative variants accompanied an early split between these lineages: genomic maximum likelihood trees suggest a particularly large distance between lineages 2 and 5 and other meningococci (Ch. 2 Fig. 7). When there were two prevalent variants of a *mod* gene within a lineage, this was often associated with the underlying population structure. For example, lineage 2 and 3 sub-lineages were associated with alternative variants and lineage 18 was diverse genotype characterised by a diversity of variants (Fig. 5). Lineage 2.1 appears to be a hybrid genotype: it includes a subset of isolates which have undergone recombination within the ribosomal genes and these same isolates were in possession of *modB2* rather than the usual lineage 2 *modB1* (Fig. 5; Ch. 2).

It is likely that particular variants are selectively maintained by particular meningococcal lineages. First, despite HGT there were strong associations among lineages and variants. Second, loss of *mod* function is not lethal, since isolates without the DRD and isolates without *modB* and *modD* were observed: it is therefore unlikely that *mods* are part of selfish RMSs since theory suggests that post-segregational killing of cells without RMS should occur if so (Kobayashi 2001). Third, DRDs were highly conserved, causing PV modification of a specific sequence in the case of *modA12*, suggesting strict associations with particular gene sets. Fourth, methylation is energetically costly and modification activity was both comprehensive and widespread across the genome.

Speculation regarding the fitness benefits that *mod* genes confer suggests that phase-variable alteration of meningococcal phenotypes is most plausible. For example, phase-variation of three *mod* genes could induce up to eight alternative phenotypes, increasing survival chances in the face of a variety of different environments, such as the immune system, blood stream entry, or transmission. It is possible that these benefits outweigh rare loss of evolutionary fitness resulting from bloodstream invasion as a by-product of Mod phase-variation (Gould and Lewontin 1979, Levin 1996). Indeed, in iron-limiting conditions, the ModA11 phasevarion influences expression of the

lactoferrin receptor, an accepted virulence determinant, in addition to genes involved in growth and energy metabolism (Perkins-Balding *et al.* 2004, Srikhanta *et al.* 2009, Hedman *et al.* 2012).

ModA12 activity affected multiple genes with a wide range of cellular functions, including metabolism and transcription (Srikhanta *et al.* 2009). That differences in metabolic adaptation may influence meningococcal lineage characteristics is becoming well accepted (Joseph *et al.* 2010, Schoen *et al.* 2014). Lineages may occupy different ecological niches (e.g. hosts of different age groups) or employ subtly different life history strategies (e.g. decreased carriage duration and increased transmission rates). Indeed, transmission fitness differences among clonal complexes, which are determined in part by six metabolic genes, have been proposed (Buckee *et al.* 2008). It is therefore possible that meningococci belonging to hyperinvasive lineages may have different metabolic characters, as proposed by Schoen and colleagues (Schoen *et al.* 2014). Hyper-mutable global regulators, or phasevarions, of metabolic loci could have evolved as an alternative to increasing mutation rates of essential metabolic genes, as necessary for rapid adaptation to changing environments, and as an alternative to evolution of environmental sensing mechanisms that are dependent on specific and predictable environmental changes.

Particular lineages may therefore require niche-specific phasevarions. Regulation of metabolic genes has been found to differ in hyperinvasive and non-hyperinvasive lineages (Pagliarulo *et al.* 2004); here, there was a statistical association between certain *mod* variants and hyperinvasive lineages (e.g. *modA12* and *modB2*) and between absence of *modB* and non-hyperinvasive lineages (Table 4). Further, certain variants were associated with invasive disease and this effect was dependent on whether variants occurred in hyperinvasive lineages or otherwise; one can speculate that the phenotypic effects of Mods are dependent on the genomic content (e.g. linkage to the genes they regulate) or to the niche inhabited by its associated lineage. Although dissociation of the DRD from linked genes through HGT is not lethal, it is unlikely to be beneficial, which is consistent with the non-absolute associations of variants and lineages observed (Table 3; Fig. 5). That lineages 2 and 5

had consistently different variants from other hyperinvasive lineages, that lineage 2 and lineage 3 sublineages had specific variants, and that lineage 3 sublineages were associated with different rates of disease was particularly interesting: genome-wide association studies (GWAS) will be of great interest in untangling the relationships between meningococcal genetic factors and phenotypes.

The role of PV Type III RMS systems in PV uptake of genetic material is uncertain. For PV RMS to cause sexual isolation of lineages and maintenance of meningococcal population structure, one would expect distinct lineages to possess distinct *mod* variants. This was only observed to an extent, such as between lineage 2 sub-lineages and between lineages 2 and 3. It is possible that PV RMS permit uptake of exogenous DNA for DNA repair. In either case, however, up to three quarters of *mod*-associated *res* genes have been found to be inactive (Fox *et al.* 2007, Srikhanta *et al.* 2011). Alternatively, for exogenous DNA to act as a source of genetic variation in rapidly changing environments, DNA would need to be donated by less related lineages possessing the same Mod variant as the recipient; this idea is not wholly consistent with meningococcal populations, which are structured with unhybridised lineages (Fig. 6). A thorough examination of the role of PV Type III restriction-modification systems in meningococci will require a survey of the extent of silencing *res* mutations, investigation of the Mod regulatory mechanism, further exploration of the particular gene expression changes in a variety of conditions, and GWAS to test associations of *mod* variants from different lineages and species with common phenotypes.

Chapter 5. Diversity of meningococcal vaccine antigens among meningococcal disease isolates

Abstract

Serogroup B meningococci remain the biggest cause of meningococcal disease in most endemic settings and up to 95% of cases occur in infants. The lack of an anti-capsular vaccine has necessitated exploration of alternative vaccine candidates, mostly proteins exhibiting varying degrees of antigenic variation. Geotemporal fluctuations in meningococcal lineages and their associated antigenic variants accentuates the need for molecular epidemiology in vaccine design, deployment, and post-implementation monitoring. Whole-genome sequencing of all meningococcal isolates from cases of meningococcal disease in England and Wales from July 2010 provides an opportunity to survey the current diversity of peptide variants of antigen encoding genes. Isolates from 1,355 cases of meningococcal disease were analysed for the prevalence of components of six meningococcal vaccines; their distribution by genogroup, lineage, and among patients of different ages was assessed. Most vaccine antigen peptides were highly diverse, yet certain variants were particularly prevalent and associated with certain lineages and age-groups. Sub-lineages of hyperinvasive lineages were found to be immunologically distinct and associated with patients of different ages. These data highlighted the importance of comprehensive surveillance of meningococcal disease with routine whole-genome sequencing, and indicated that vaccine formulations containing a judicious choice of antigen variants have potential for broad-spectrum protection against meningococcal disease.

Introduction

Meningococcal disease, caused by *Neisseria meningitidis*, accounts for significant worldwide morbidity and mortality due to meningitis and severe sepsis. The geotemporal distribution of asymptomatic meningococcal carriage, meningococcal disease incidence, and of meningococcal genotypes is varied and dynamic (Halperin *et al.* 2012). Meningococcal lineages associated with polysaccharide capsules corresponding to serogroups B and C are responsible for most disease in Western Europe, North and South America, Australia, and New Zealand but in the past two decades serogroup Y disease has increased in the USA and certain European countries (Harrison *et al.* 2009a, Broker *et al.* 2012b, Halperin *et al.* 2012, Ibarz-Pavon *et al.* 2012, Krauland *et al.* 2012, Ladhani *et al.* 2012b, Bettinger *et al.* 2013, Toros *et al.* 2014).

As such, effective meningococcal vaccine design and deployment is reliant on contemporary molecular epidemiological surveys. Conjugate polysaccharide vaccines have been extremely effective, not least in targeting the serogroup C disease outbreaks that increased across Europe in the 1990s (Gray *et al.* 2006, Kriz *et al.* 2011). Currently, however, there is no comprehensive vaccine against meningococcal disease. The structural similarity of serogroup B polysaccharide to polysaccharides associated with the human Neural Cell Adhesion Molecule (NCAM) renders it poorly immunogenic and raises concerns over autoimmune reactions if included in vaccines (Finne *et al.* 1983, Jodar *et al.* 2002). In the recent past, outer-membrane vesicle (OMV) vaccines were developed for the control of particular serogroup B epidemics, such as those in Cuba (VA-MENGOC-BC® vaccine), Norway (MenBvac® vaccine), and New Zealand (MeNZB™ vaccine) (Bjune *et al.* 1991, Sierra *et al.* 1991, O'Hallahan *et al.* 2004). These were formulated with the PorA antigen variants of the epidemic strain as the major immunogen and as such their effectiveness is limited in endemic scenarios where serogroup B meningococci are antigenically diverse and fluctuating in geotemporal prevalence.

Sub-capsular antigens, especially proteins, have therefore been investigated for inclusion in 'serogroup B-substitute' vaccines. Several such antigens have been incorporated into vaccine formulations that are in various stages of development. In order to achieve the widest possible strain coverage in endemic periods, vaccines should contain a breadth of antigenic diversity: the NonaMen vaccine (RIVM, the Netherlands) has been proposed for such purposes, as it is constituted by nine PorA subtypes corresponding to the most prevalent disease-associated meningococcal strains (Kaaijk *et al.* 2013). Bexsero® (GSK) is a supplemented OMV vaccine, which includes four components (PorA subtype P1.7-2,4, fHbp sub-variant 1.1, NHBA variant 2, and NadA-3.8 sub-variant) from diverse serogroup B meningococcal strains (Bambini *et al.* 2009); Bexsero® was licensed in Europe in 2013 and in the USA in 2015. The Trumenba® vaccine, developed by Pfizer, is a bivalent recombinant vaccine that includes two fHbp antigens from subfamily A and B; the U.S. Food and Drug Administration licensed this vaccine for use in adolescents in 2014.

Bexsero® is to be included in the UK infant immunization schedule from September 2015. The incidence of meningococcal disease has been declining in England and Wales for a decade, and at present serogroup B incidence is relatively low, as in other high-income countries such as the USA and Sweden (Ladhani *et al.* 2012a, Castelblanco *et al.* 2014). Meningococcal disease incidence is also age-associated: in recent years the UK has seen annual rates of 38.6 cases per 100,000 population in children under one year of age, of which more than 93% is caused by serogroup B clonal complex (cc) 41/44 and cc269 meningococci (Trotter *et al.* 2006, Ladhani *et al.* 2012a, Brehony *et al.* 2014). Recent increases in serogroup Y cc23 meningococci indicate that additional cases may occur in the older UK population, given the association of this serogroup with older age-groups (Ladhani *et al.* 2012b).

The association of meningococcal clonal complexes with certain antigens, including capsular polysaccharide and vaccine candidates, emphasizes that comprehensive surveillance is indispensable for vaccine design, deployment, and post-implementation monitoring (Urwin *et al.* 2004, Buckee *et*

al. 2011, Bambini *et al.* 2013, Hoiseth *et al.* 2013, Brehony *et al.* 2014). Over the past two decades, sequence based molecular typing has become integral to surveillance of fluctuations in meningococcal disease. The Public Health England Meningococcal Reference Unit (PHE-MRU) obtains capsular type and PorA subtype data from culture- and non-culture-confirmed cases in routine surveillance; following sustained increases in serogroup C cases in the mid-1990s, meningococcal C conjugate vaccines were introduced to UK infant immunization schedules with a 'catch-up' campaign targeted to the highest risk age-groups (Gray *et al.* 2006).

The PHE-MRU is to routinely obtain whole-genome sequence (WGS) data from culture-confirmed cases (CCCs) in England, Wales, and Northern Ireland from the 2015/16 epidemiological year (July to June) onwards. This means that comprehensive investigations of the likely and actual impact of vaccine interventions may be made more directly. Since July 2010/11, WGS data from all meningococcal isolates from CCCs in England and Wales have been acquired and made publically available in the MRF Meningococcal Genome Library (MRF-MGL). This resource contains molecular typing information for thousands of isolates, allowing investigation of temporal fluctuations in antigens and of the age-associated distribution of vaccine antigen variation. In this study, the diversity of vaccine antigen peptides and the prevalence of Bexsero®, Trumenba®, NonaMen, MenBvac®, MeNZB™, and VA-MENCOG-BC® components were identified in meningococcal lineages among 1,344 CCCs from patients of different ages in the MRF-MGL.

Methods

Isolates

This study made use of all isolates from culture-confirmed cases (CCCs) comprising the Meningitis Research Foundation Meningococcus Genome Library (MRF-MGL) in August 2014, with the exception of isolates from Northern Ireland which were not a comprehensive sample of CCCs from hospitals in Northern Ireland. In total, there were 1,344 isolates from England and Wales from the 2010/11, 2011/12, and 2012/13 epidemiological years (July 1st – June 30th). Patient age data were obtained from the Public Health England Meningococcal Reference Unit (PHE-MRU): this information is collected by PHE-MRU as the number of whole days or whole months of life for children aged less than one year of age and as the number of whole years for patients aged one year or greater. Isolates were assigned to ribosomal MLST (rMLST) and core-genome MLST (cgMLST) derived meningococcal lineages, sub-lineages, and to genogroups as in Chapter 2 (Harrison *et al.* 2013b).

Vaccine antigens in whole-genome sequence data

Genomic loci and deduced peptide sequences of the meningococcal vaccine antigens fHbp, NadA, NHBA, and PorA were annotated within isolate whole-genome sequence (WGS) data using the PubMLST 'autotagger' and 'autodefiner' tools as described in Chapter 2. Previously unreported nucleotide alleles or peptide variants were confirmed in MEGA6 (Tamura *et al.* 2013) and assigned a unique numerical identifier in the PubMLST database using 'web-based sequence tagging' (Jolley and Maiden 2013). Functions of genomic loci were assigned using Fam18 genome annotations (Bentley *et al.* 2007) or were bioinformatically predicted based on homology to proteins in the Pfam database (Finn *et al.* 2014).

PorA subtypes, NHBA peptide alleles, fHbp sub-variants, and NadA peptide alleles were categorized according to established nomenclature (Russell *et al.* 2004, Jolley *et al.* 2007, Brehony *et al.* 2009,

Bambini *et al.* 2014). For fHbp and NadA, peptide sub-variants were designated to the variant/family groupings of each antigen. For example, the fHbp sub-variant included in Bexsero® is sub-variant 1 and this belongs to variant family 1, thus, it is referred to as fHbp-1.1. Six NadA variants have been identified (NadA-1 to NadA-6), four of which have been recently grouped into two variant groups (NadA-2/3 and NadA-4/5) (Bambini *et al.* 2014); the NadA peptide allele included in Bexsero® is allele 8, which belongs to variant 3, and thus it is referred to as NadA-3.8.

Components of the following vaccines were investigated:

(1) Trumenba® (Pfizer), also called rLP2086, which contains fHbp-3.45 and fHbp-1.55 (Nissen *et al.* 2013).
(2) Bexsero® (Novartis), which contains: PorA P1.7-2,4; fHbp-1.1; NadA-3.8; and NHBA peptide 2 (Bambini *et al.* 2009).
(3) NonaMen (RIVM, the Netherlands), which contains nine PorA subtypes: P1.7,16; P1.5-1,2-2; P1.19,15-1; P1.5-2,10; P1.12-1,13; P1.7-2,4; P1.22,14; P1.7-1,1; P1.18-1,3 (van der Ley *et al.* 1995).
(4) MenBvac®, which contains PorA subtype P1.7,16 (Bjune *et al.* 1991).
(5) MeNZB™, which contains PorA subtype P1.7-2,4 (O'Hallahan *et al.* 2004).
(6) VA-MENGOC-BC®, which contains PorA subtype P1.19,15 (Sierra *et al.* 1991).

The potential baseline vaccine coverage of meningococcal lineages and patient age-groups was based on exact sequence matches of at least one vaccine component to deduced antigen peptide sequences within isolates. PorA variable regions were analysed as separate vaccine components. Analysis was carried out on isolates of all genogroups, and on genogroup B isolates separately. Antibody cross-reactivity among NadA-1, NadA-2, and NadA-3 peptides, and among fHbp family 1 sub-variants, has been identified in pre-clinical and clinical studies; therefore, the potential baseline coverage of Bexsero® was also calculated with cross-reactivity from these components taken into account (Comanducci *et al.* 2004, Brunelli *et al.* 2011).

Statistical analysis

Simpson's diversity index (*D*) and 95% confidence intervals were calculated in excel to assess the diversity of vaccine antigen peptide variants. Logistic regression models and Chi-square tests were carried out in R version 3.2.2 (R Development Core Team 2012) as described in Chapter 2. For the multinomial logistic regression of meningococcal lineage on patient age, lineage was modeled as a categorical outcome variable with six levels (lineages 3, 2, 11, 23, 13, and other lineages) with lineage 3 the baseline level for relative risk ratios. For the binomial logistic regression of lineage 2 sub-lineages on patient age, the two sub-lineages were modeled as a binomial outcome variable with lineage 2.2 relative to lineage 2.1. In both, patient age was modeled as a categorical predictor variable with four levels (<5 years, 5-14 years, 15-24 years, and >25 years). The map of UK regions and the age demographic of each region was obtained from the Office for National Statistics (www.ons.gov.uk).

Results

Geotemporal distribution of meningococcal lineages

The 1,344 isolates from culture-confirmed cases (CCCs) of meningococcal disease in England and Wales between 2010/11 and 2012/13 were assigned to 22 clonal complexes on the basis of seven-locus MLST and to 22 lineages on the basis of ribosomal MLST (rMLST) and core-genome MLST (cgMLST) (Fig. 1). There were 98 isolates not assigned to a known MLST clonal complex (ccND isolates), however, 69 (70.41%) of these were assigned to an rMLST/cgMLST lineage: for example, the majority (n=43) of ccND isolates were members of lineage 2 and nine ccND isolates were members of lineage 3.

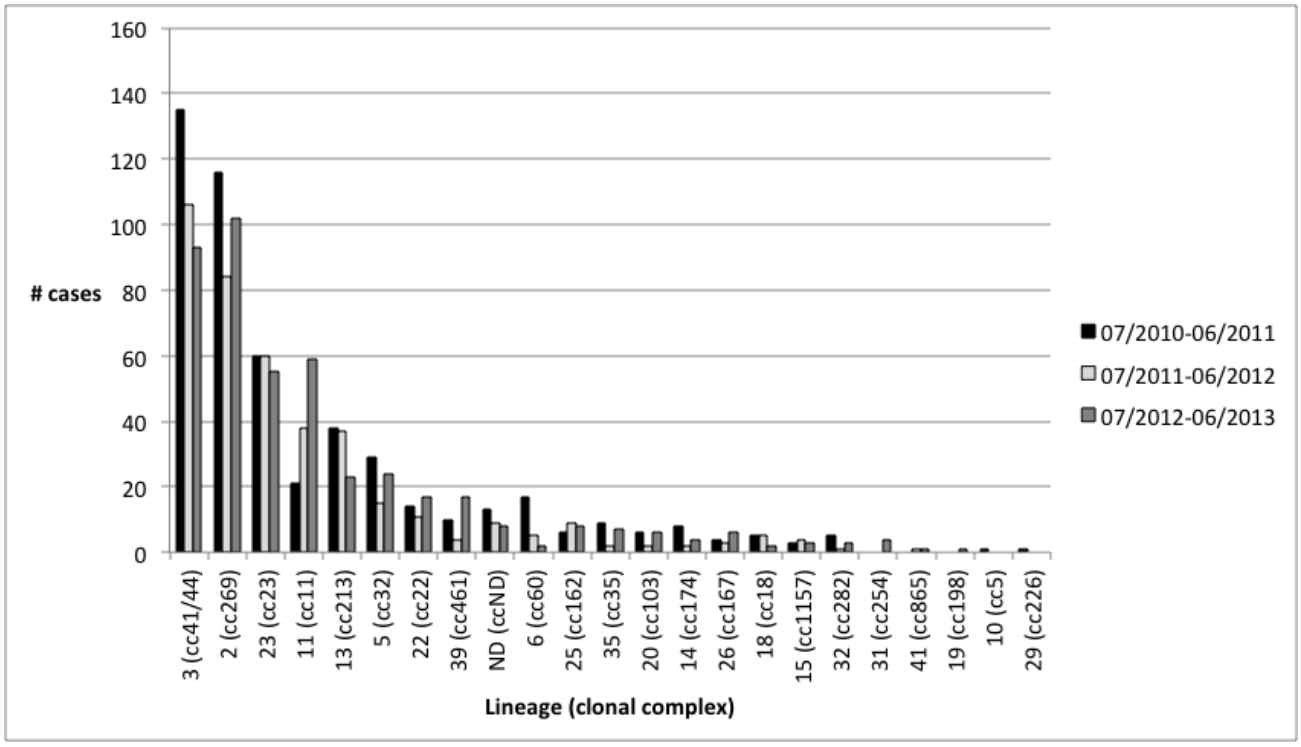


Figure 1. Lineages of *N. meningitidis* causing meningococcal disease in England and Wales. Isolates collected in national surveillance from all culture-confirmed cases in the 2010/11, 2011/12 and 2012/13 epidemiological years (July 1st – June 30th) were obtained from the MRF Meningococcus Genome Library. They were assigned to whole-genome sequence derived lineages using ribosomal MLST (49 loci) and core-genome (cg) MLST (1,605 loci from the PubMLST 'cgMLST scheme version 1.0').

The number of CCCs increased between 2011/12 (n=398) and 2012/13 (n=445), although not to the number observed in 2010/11 (n=501). Nine lineages were prevalent (each responsible for more than 30 cases across the three years) and together, the hyperinvasive lineages 3 (cc41/44), 2 (cc269), 23 (cc23), and 11 (cc11) caused 929 (69.12%) cases (Fig. 1). Lineage 3 appeared to follow a declining trend since it caused 42 fewer cases in 2012/13 (31.11% decrease) than in 2010/11, the greatest absolute change in any lineage. Lineage 23 cases also decreased slightly in 2012/13 (Fig. 1). In contrast, lineage 11 increased across the three years, with 38 additional cases in 2012/13 compared to 2010/11 (108.95% increase). Lineages 2 and 5 (cc32), which have found be more genetically similar than other clonal complexes (Budroni *et al.* 2011), followed similar trends across the three years: following reduced prevalence in 2011/12, they were together responsible for an additional 27 cases in 2012/13 (Fig. 1). Finally, lineages 39 (cc461) and 31 (cc254) also increased in prevalence: lineage 39 caused 13 additional cases in 2012/13 compared to the previous year, and lineage 31 caused four cases in 2012/13, having been absent in the previous two years (Fig. 1).

The geographical distribution of lineages was heterogeneous in England and Wales (Fig. 2). The most prevalent lineages in the majority of regions were lineages 3 (cc41/44), 2 (cc269), 23 (cc23), and 11 (cc11), however, a large proportion of CCCs in the North East, Yorkshire and Humberside, and the West Midlands was caused by lineage 13 (cc213) and a large proportion of CCCs in Wales was caused by lineage 5 (cc32) (Fig. 2). Notably, lineage 11 disease was particularly prevalent in London and lineage 2 was more prevalent than lineage 3 in the North West, Yorkshire and Humberside, and the West Midlands (Fig. 2). A large proportion of CCCs in the South East and London was caused by rare lineages compared to other UK regions.

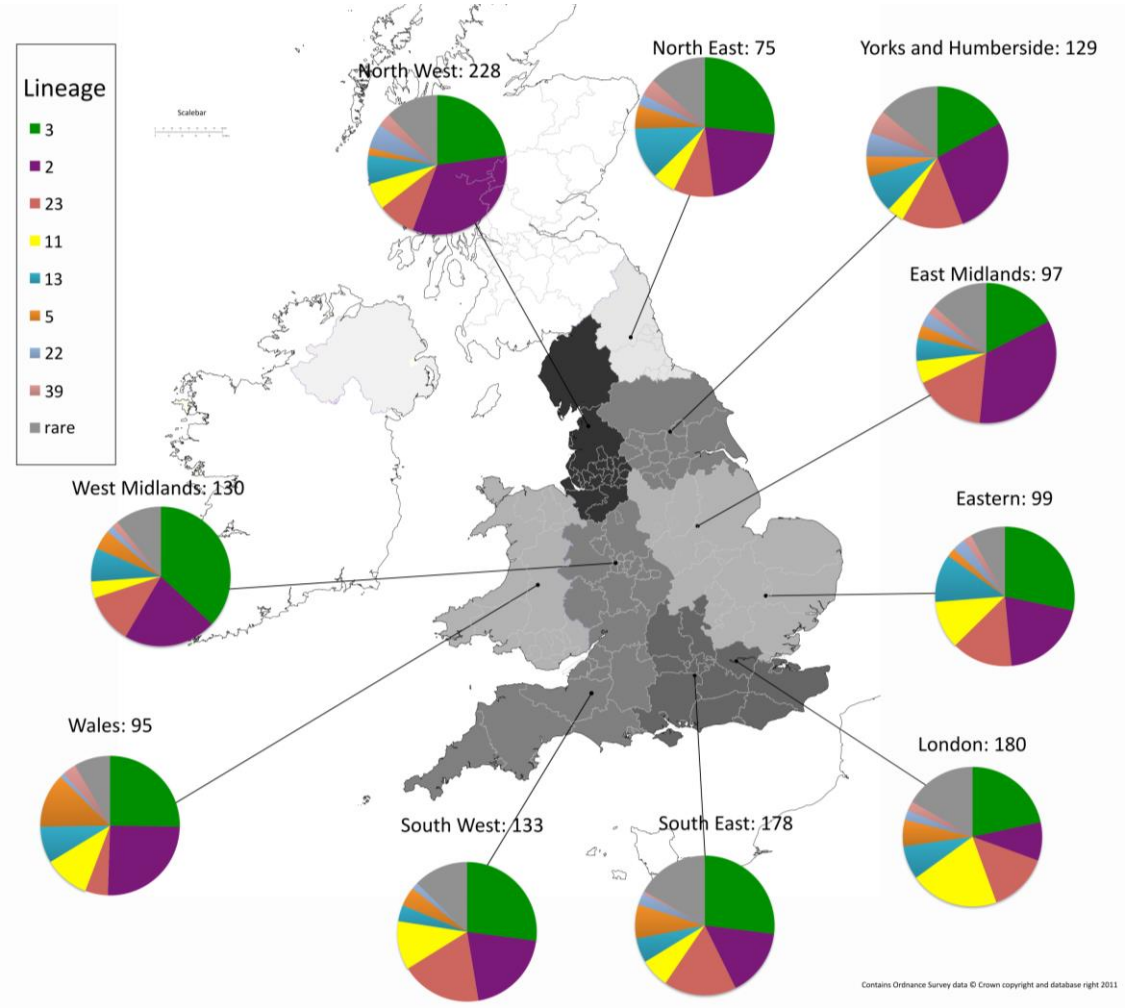


Figure 2. UK distribution of meningococcal lineages from culture-confirmed cases of meningococcal disease in the 2010/11, 2011/12, and 2012/13 epidemiological years. Regions of England and Wales are shaded from dark to light in accordance with high to low numbers of culture-confirmed cases received by PHE-MRU in the three years. The numbers of culture-confirmed cases are provided next to region names.

Antigenic diversity of meningococcal lineages

Simpson's Diversity Index (*D*) measures diversity from 0 to 1, where 1 is the greatest diversity, and NHBA, PorA VR2, and fHbp sub-variants each had a value of *D* greater than 0.9 (Table 1). PorA VR2 was the most diverse vaccine antigen peptide (*D*=0.92 [0.92-0.93]); non-overlapping confidence intervals of *D* indicate statistically significant differences in diversity and VR2 was significantly more diverse than VR1 (*D*=0.87 [0.86-0.88]). The *nadA* gene was absent in 1,025 (76.3%) CCCs and NadA

peptide sub-variants were therefore significantly less diverse (*D*=0.39 [0.36-0.42]) than those of other antigens.

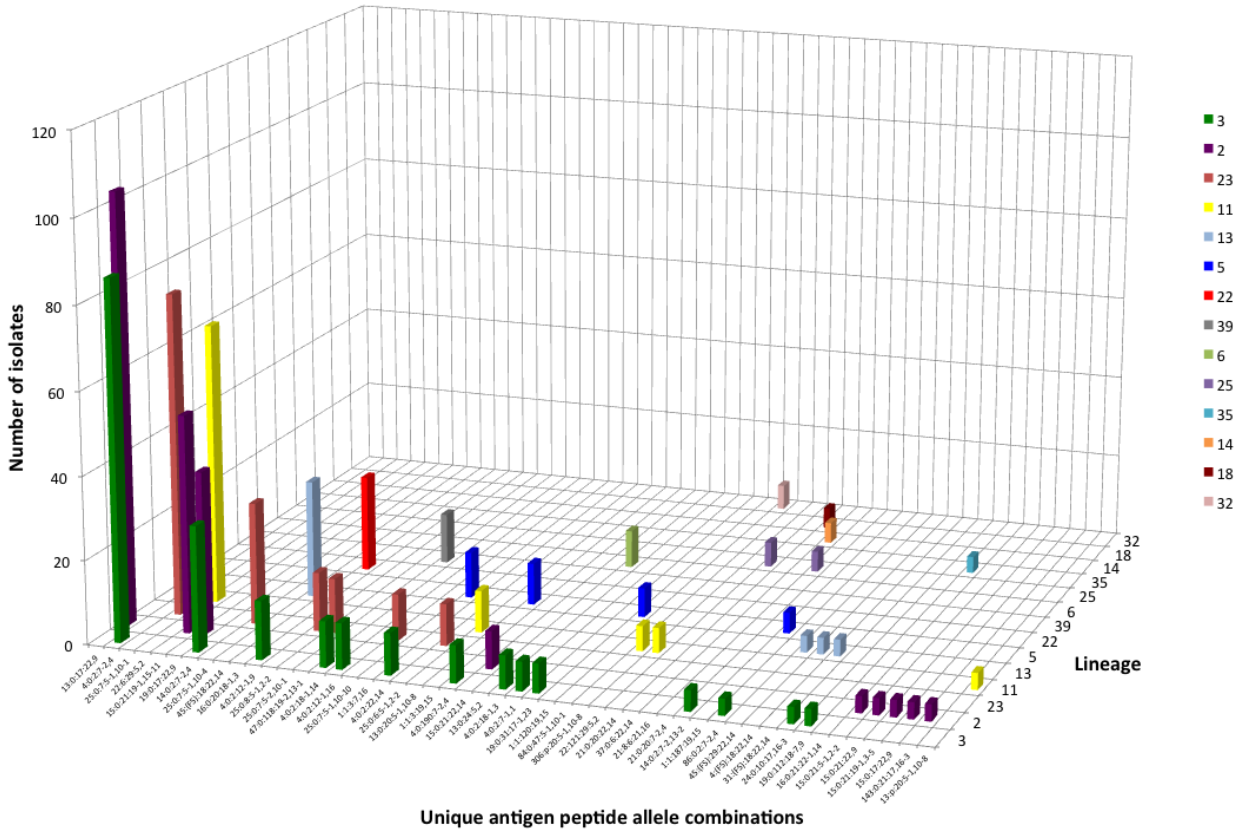
Table 1. Vaccine antigen peptide variant diversity among patient age-groups.

Age (years)	NadA *D* (95% CI)	NHBA *D* (95% CI)	fHbp *D* (95% CI)	VR2 *D* (95% CI)	VR1 *D* (95% CI)
<1 (n=300)	0.31 (0.24, 0.38)	0.88 (0.86, 0.91)	0.89 (0.87, 0.91)	0.88 (0.86, 0.9)	0.82 (0.78, 0.85)
1-4 (n=315)	0.39 (0.33, 0.46)	0.88 (0.86, 0.9)	0.9 (0.88, 0.92)	0.92 (0.91, 0.94)	0.84 (0.81, 0.87)
5-14 (n=99)	0.33 (0.21, 0.44)	0.91 (0.88, 0.94)	0.91 (0.88, 0.94)	0.91 (0.89, 0.94)	0.88 (0.85, 0.92)
15-24 (n=168)	0.42 (0.33, 0.52)	0.9 (0.88, 0.92)	0.9 (0.88, 0.92)	0.93 (0.92, 0.95)	0.88 (0.85, 0.9)
25-39 (n=73)	0.5 (0.36, 0.64)	0.92 (0.9, 0.95)	0.93 (0.91, 0.95)	0.93 (0.91, 0.96)	0.84 (0.79, 0.9)
40-64 (n=173)	0.42 (0.33, 0.51)	0.91 (0.89, 0.93)	0.9 (0.88, 0.92)	0.92 (0.91, 0.94)	0.87 (0.84, 0.9)
65+ (n=207)	0.41 (0.32, 0.49)	0.87 (0.84, 0.9)	0.85 (0.81, 0.89)	0.91 (0.89, 0.92)	0.82 (0.78, 0.86)
Total (n=1344)	**0.39 (0.36, 0.42)**	**0.91 (0.91, 0.92)**	**0.92 (0.91, 0.92)**	**0.92 (0.92, 0.93)**	**0.87 (0.86, 0.88)**

Simpson's Diversity Index (*D*) and 95% confidence intervals were measured in vaccine antigen peptides among age-groups in the MRF-MGL, 2010/11-2012/13. VR: PorA variable region.

In total there were 14 NadA peptide alleles, 109 NHBA peptide alleles, 131 fHbp peptide alleles, 35 PorA VR1 peptide alleles, and 72 VR2 peptide alleles among the 1,344 isolates. Each isolate had one of 475 unique combinations of vaccine antigen peptide alleles and each combination was specific to a single meningococcal lineage (Fig. 3a). Twelve unique combinations were prevalent (each identified in more than 13 isolates) and the four most prevalent combinations were each specific to lineages 3, 2, 23, and 11 (Fig. 1; Fig. 3a): no antigen peptide allele was identified in more than one of these four combinations. Where the same antigen peptide allele was identified in more than one prevalent combination, it was in isolates belonging to the same lineage, and overall, the average number of shared vaccine antigen peptide alleles between two isolates was always greater within lineages than among lineages (Fig. 3b). The most diverse lineage with respect to its vaccine antigen alleles was lineage 3 (cc41/44), and the least diverse lineage (constituted by more than two isolates) was lineage 32 (cc282) (Fig. 3b).

a)



b)

Lineage	10	11	13	14	15	18	19	2	20	22	23	25	26	29	3	31	32	35	39	41	5	6	ND
10	NA																						
11	0.0	2.0																					
13	0.0	0.1	2.4																				
14	0.3	0.0	0.4	1.8																			
15	0.0	0.2	0.2	0.4	2.6																		
18	0.3	0.2	0.4	0.9	0.3	1.2																	
19	0.0	0.0	0.1	0.0	0.0	0.0	NA																
2	0.4	0.1	0.5	0.3	0.3	0.3	0.0	0.9															
20	0.0	0.2	0.1	0.0	0.2	0.0	0.0	0.3	1.5														
22	0.0	1.1	0.0	0.1	0.2	0.1	0.0	0.2	0.8	2.0													
23	0.0	0.6	0.0	0.1	0.2	0.2	0.1	0.1	0.5	0.3	1.7												
25	0.2	1.0	1.5	0.5	0.1	0.4	0.0	0.5	0.0	1.1	0.0	2.6											
26	0.0	1.0	0.0	0.0	0.3	0.1	0.0	0.1	0.4	0.4	0.8	0.0	2.7										
29	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.2	1.4	1.2	0.0	0.0	0.0	NA									
3	0.0	0.1	0.2	0.2	0.1	0.1	0.5	0.1	0.2	0.1	0.2	0.2	0.1	0.2	0.8								
31	0.0	0.8	0.8	0.2	0.9	0.2	0.0	0.2	0.1	0.8	0.0	1.6	0.0	0.0	0.1	2.1							
32	0.0	1.0	0.0	0.0	0.3	0.2	0.0	0.1	0.3	0.5	1.1	0.0	1.5	0.0	0.1	0.0	3.3						
35	0.0	0.2	0.4	0.0	0.0	0.1	0.3	0.4	0.1	0.1	0.1	0.3	0.0	0.0	0.3	0.4	0.0	0.9					
39	0.0	0.0	0.0	0.1	0.2	0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.2	0.0	0.1	2.0				
41	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.1	0.3	0.0	0.4	0.0	0.3	0.0	0.2	0.0	0.5	0.0	0.1	0.0			
5	0.0	0.1	0.1	0.3	0.1	0.2	0.0	0.1	0.1	0.1	0.1	0.2	0.1	0.0	0.2	0.1	0.0	0.2	0.1	0.3	1.1		
6	0.0	0.3	0.1	1.0	1.2	0.4	0.0	0.3	0.2	0.1	0.2	0.1	0.2	0.0	0.1	0.8	0.2	0.1	0.2	0.0	0.2	1.5	
ND	0.0	0.2	0.3	0.4	0.1	0.2	0.2	0.3	0.3	0.4	0.1	0.4	0.2	0.2	0.3	0.2	0.2	0.3	0.1	0.1	0.2	0.3	0.3

Figure 3. Vaccine antigen peptide alleles in meningococcal lineages. a) Each unique combination of fHBP, NadA, NHBA, PorA VR1, and PorA VR2 alleles was identified in only one lineage. Combinations found in more than three isolates (n=51) are displayed. **b)** Average number of shared antigen peptide alleles between isolates, calculated within and between lineages, and coloured green to red from 0 alleles to 3.3 alleles. NA: lineages with only one isolate member.

Prevalence of vaccine components among lineages

In total, 1,102 (82.0%) isolates contained exact peptide sequence matches to one or more component of any of the vaccines investigated (potential baseline coverage). The proportion of isolates with exact matches differed among lineages: the hyperinvasive lineages 3 (cc41/44) and 2 (cc269) had the most exact matches overall (302 and 181 isolates, respectively). The NonaMen vaccine, which includes the greatest number of different antigen peptide alleles (nine PorA subtypes), had the greatest potential baseline coverage: 1,077 (80.1%) isolates possessed a VR1 or a VR2 allele included in the vaccine formulation and these belonged to the greatest number of lineages (n=20) of any vaccine (Fig. 4). A large proportion (302, 28.04%) were lineage 3 isolates due to the association of lineage 3 with PorA subtype P1.7-2,4, and unlike other vaccines, the prevalent hyperinvasive lineages 2 and 23 (cc23) were well represented due to strong associations with P1.22,9 and P1.5-1,10-1, respectively (Fig. 4). Exact matches to the components of Trumenba® were identified in isolates of a single lineage, lineage 13 (cc213) (Fig. 4).

Exact matches to one or more of the four Bexsero® antigen peptide variants were found in 378 (28.1%) isolates from 13 lineages (Fig. 4). The most prevalent among these was lineage 3 (290 (76.72%) isolates) since it was associated with NHBA peptide 2 and PorA VR2 P1.4, which are part of the Bexsero® formulation. Only two lineage 2 isolates possessed an exact match to a Bexsero® antigen variant. NadA sub-variant 8, the variant included in Bexsero®, was not found in any hyperinvasive lineages except for lineage 5 (cc5) (one isolate); however, this variant was identified in 13 lineage 14 (cc174) isolates. Bexsero® fHbp variant family 1 peptide 1 was identified in 47 (69.1%) lineage 5 (cc32) isolates, consistent with the derivation of this component from the lineage 5 MC58 isolate (Tettelin *et al.* 2000). A total of 356 (36.9%) genogroup B isolates possessed any of the peptide variants included in Bexsero®. When assuming cross-protection among peptide alleles of the NadA-1 and NadA-2/3 variants, and among peptide alleles of fHbp family 1, the overall number of

isolates containing a match more than doubled to 850 (63.2%) isolates, of which 708 were genogroup B (73.4% of genogroup B isolates).

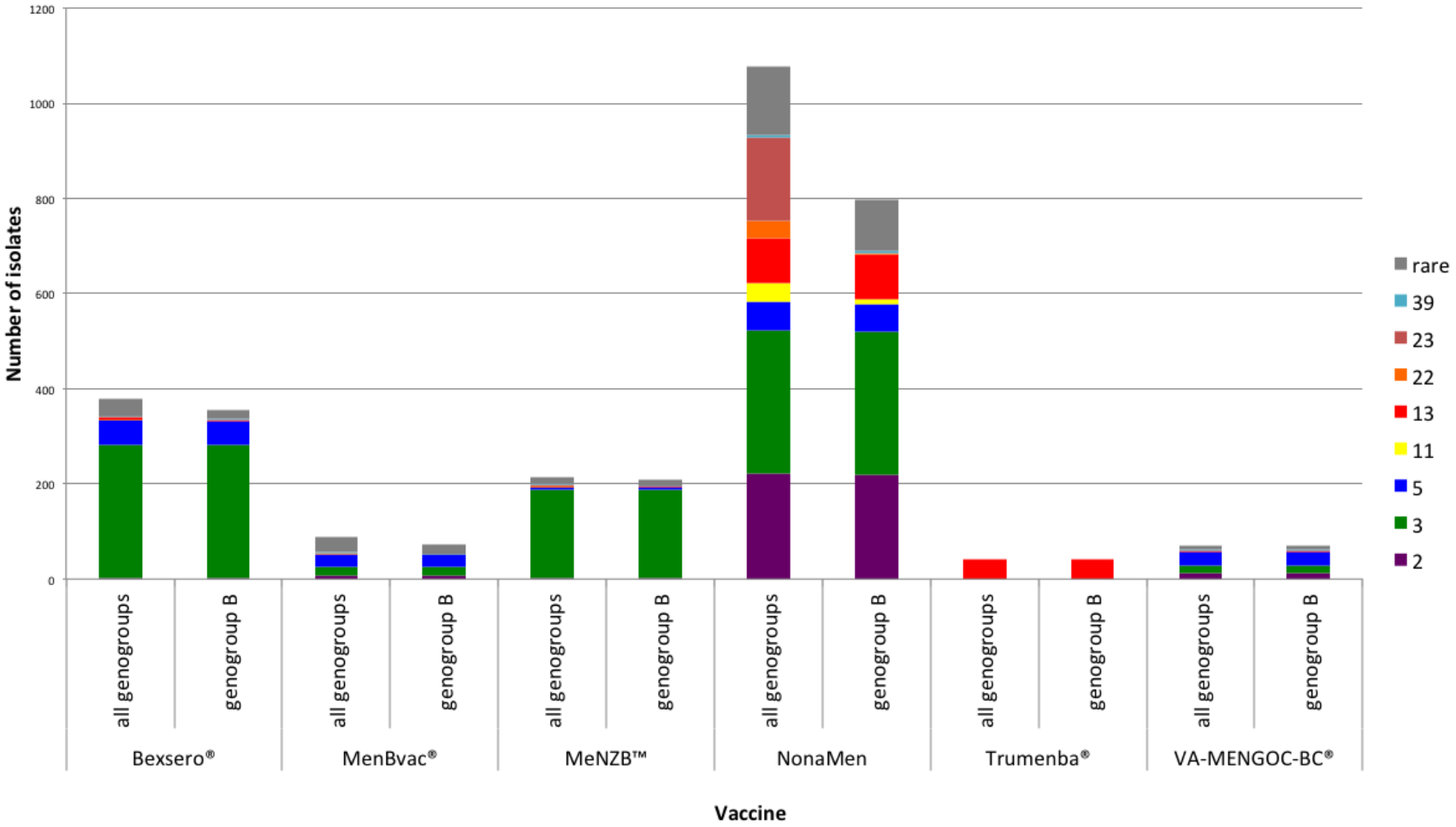


Figure 4. Number of isolates of each meningococcal lineage with exact sequence matches to peptide alleles included in meningococcal vaccines. Among the 1,344 isolates from cases of meningococcal disease in England and Wales between 2010/11-2012/13, there were 964 genogroup B isolates.

Association of meningococcal lineages with patient age

The prevalence of meningococcal disease in patients of different ages was consistent with previous descriptions: the majority of CCCs were from infants under a year of age and there was a secondary disease peak in young adults aged between 16 and 21 years (Fig. 5). The number of cases in certain age-groups altered throughout the study period. Meningococcal disease in children less than four years of age declined by an average of 16.75 cases per two-year age-group and cases in patients aged 13 to 21 years increased by an average of 7.33 per age-group (Fig. 5).

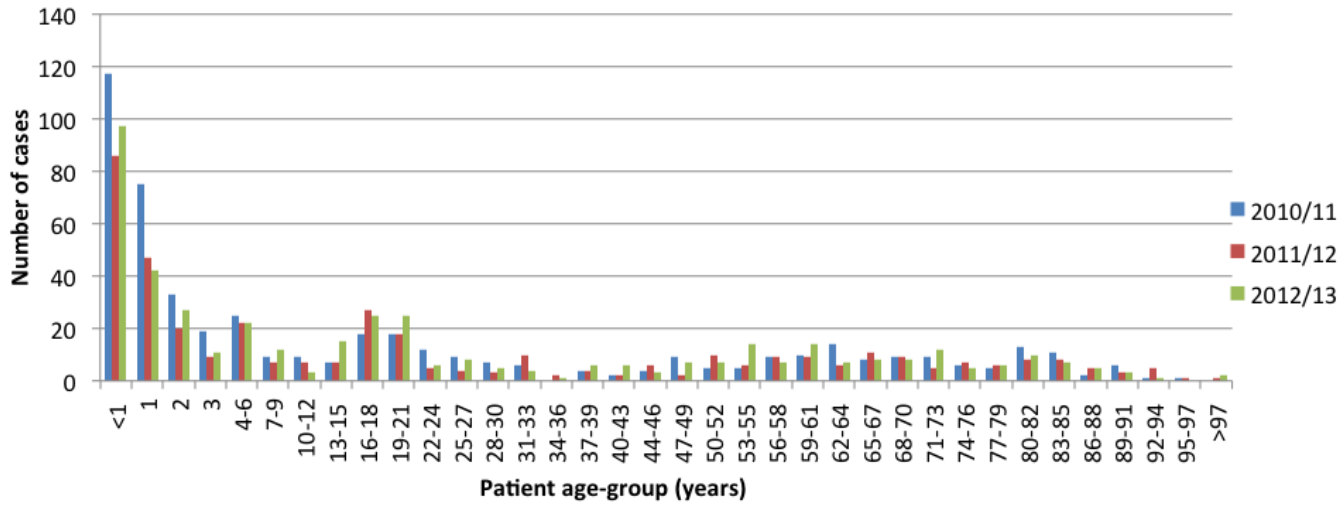


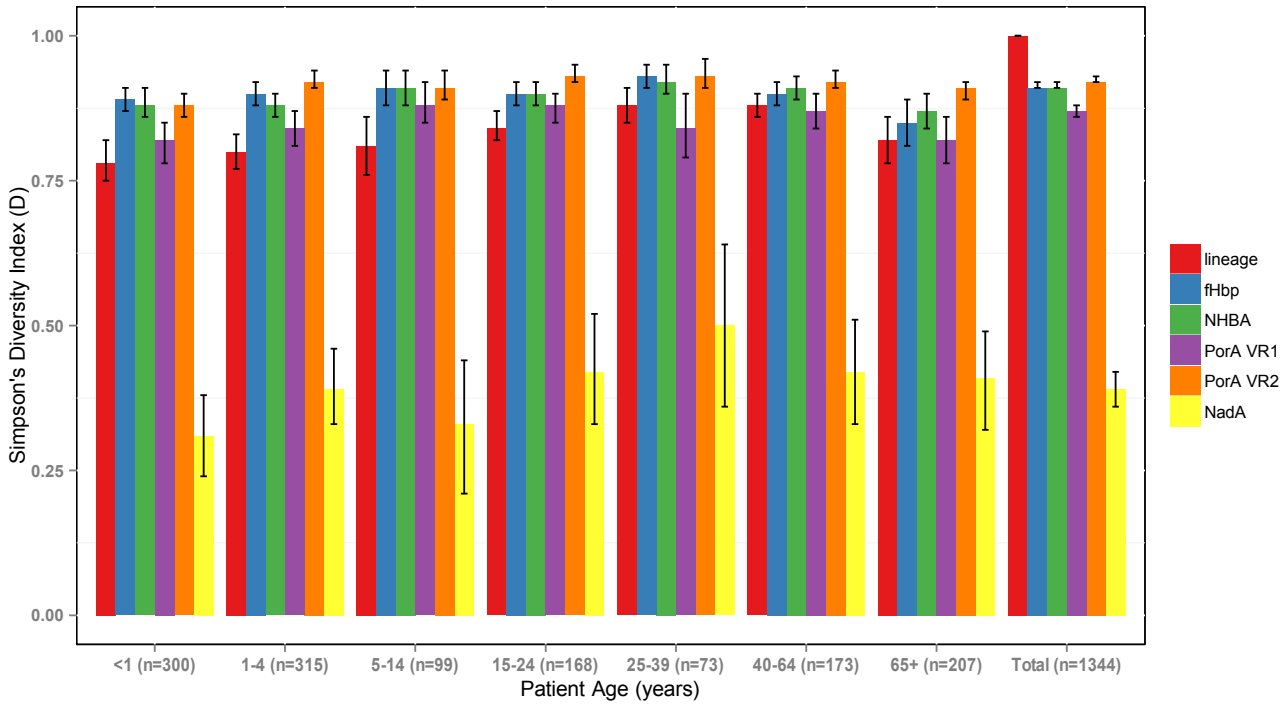
Figure 5. Distribution of meningococcal disease cases by patient age in England and Wales, 2010/11-2012/13.

The distribution of meningococcal lineages among isolates from patients of different ages was heterogeneous (Fig. 6). Although 18 different lineages caused disease in infants and in children under four, the diversity of lineages in these age-groups (*D*=0.78 [0.75, 0.82] and *D*=0.80 [0.77, 0.83], respectively) was significantly less than that in adults aged 25-39 and 40-64 years (*D*=0.88 [0.85, 0.91] and *D*=0.88 [0.86, 0.90], respectively) (Fig. 6a). This was partly the result of strong associations between lineages 2 and 3 (cc41/44) and disease young children (Fig. 6b).

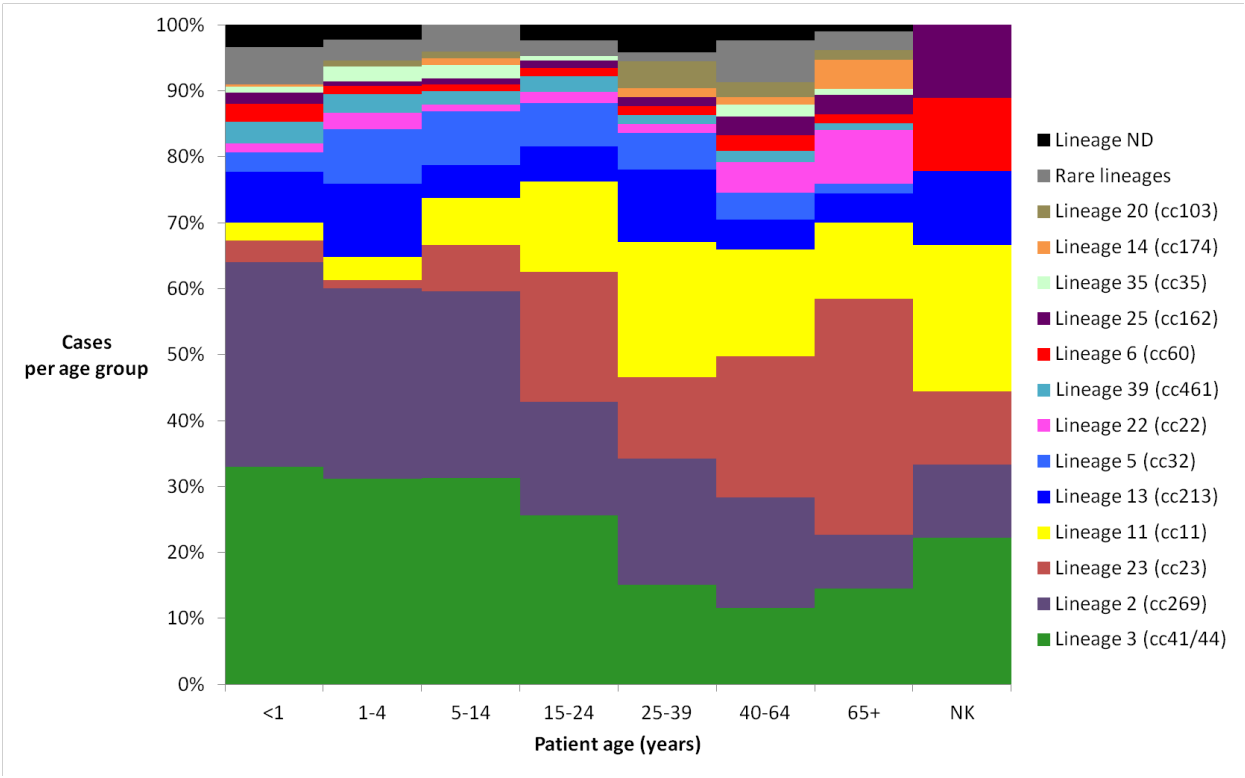
Associations between meningococcal lineages and patient age was modeled as a multinomial logistic regression model, with patient age as the predictor of the lineage of meningococci responsible for

disease (Fig. 6c). The baseline outcome was lineage 3 disease: the relationship of lineage 2 with patient age was never significantly different from that of lineage 3 (Fig. 6c). Lineage 3 was, however, significantly more likely to cause disease in children under four years of age than lineages 11 (cc11) and 13 (cc213); lineage 23 (cc23) was highly unlikely to cause disease in this age-group (relative risk ratio 0.08 [0.04-0.13]). Lineages 11 and 23 were more likely to cause meningococcal disease in individuals older than four years than lineage 3 and this effect increased with age (Fig. 6c). There was a slight association between lineage 13 and disease in individuals over 25 relative to lineage 3, however, this was not statistically significant. Lineages other than the aforementioned were strongly, and significantly, more likely to be the cause of meningococcal disease in patients over the age of 25 years and were also more likely to cause disease in this age-group than lineage 3.

a)



b)



c)

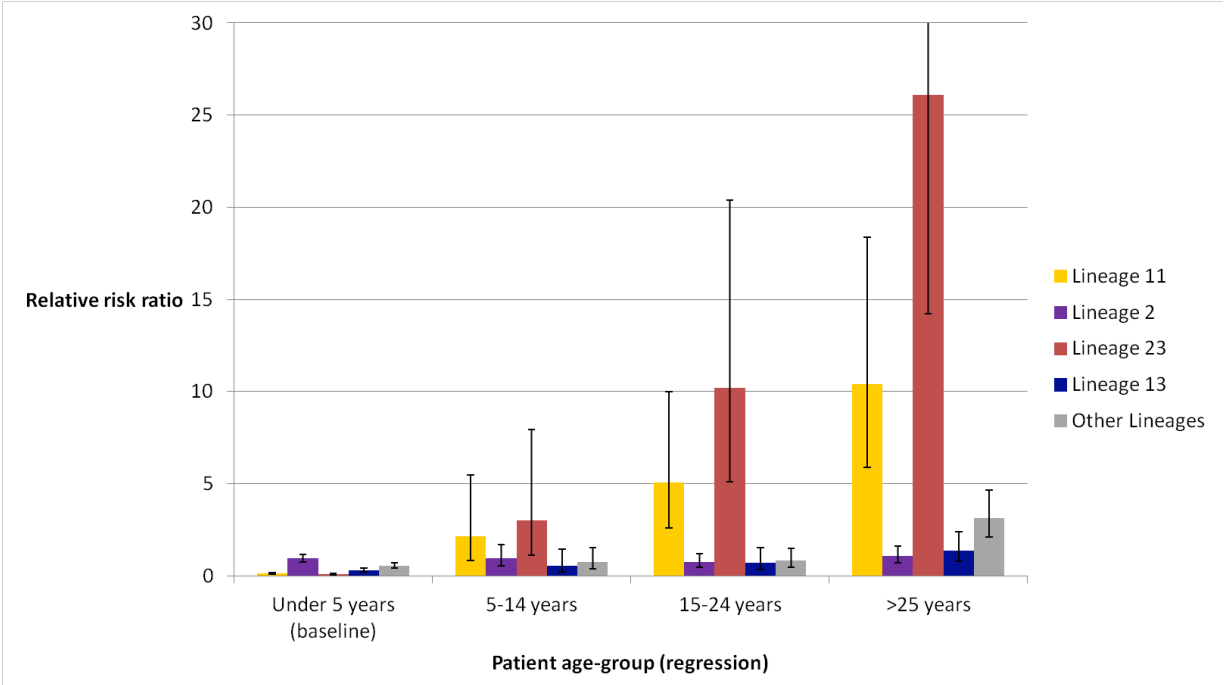


Figure 6. Meningococcal lineages causing disease in different ages. a) Diversity of lineages and vaccine antigens among disease in each age-group measured by Simpson's Diversity Index (*D*). **b)** Prevalence of lineages causing disease in each age-group. **c)** Risk of disease from meningococcal lineages relative to lineage 3 (cc41/44). All error bars represent 95% confidence intervals; the upper confidence interval for the relative risk ratio of lineage 23 in c) was 47.99.

Characteristics of hyperinvasive lineage sub-lineages

The four most prevalent lineages causing meningococcal disease in England and Wales, hyperinvasive lineages 3 (cc41/44), 2 (cc269), 23 (cc23), and 11 (cc11), each comprised distinct sub-lineages at core-genome resolution (1,592 meningococcal core-genome (cg) loci from the *N. meningitidis* cgMLST version 2.0 scheme, available at http://PubMLST.org/neisseria).

Lineage 3 (cc41/44)

The 344 lineage 3 isolates from CCCs were associated with three sub-lineages, lineages 3.1-3.3 (Fig. 7a). Each sub-lineage arose from a discrete cluster of internal nodes and isolates within them were more similar to each other than to isolates of other sub-lineages. Lineage 3.1 was the most prevalent, causing 289 cases from 2010/11-2012/13, and was responsible for the overall decline of the lineage across the three years (Fig. 7b); it was predominantly associated with seven-locus MLST ST-41. Sub-lineage 3.2 caused 41 cases and was predominantly composed of ST-1097 isolates, and sub-lineage 3.3 caused only four cases, each of a different ST (Fig. 7b). All lineage 3 isolates were genogroup B except for two lineage 3.2 genogroup C isolates. Lineages 3.1 and 3.2 had similar relationships with patient age: both caused most disease in infants and young children and there was evidence of smaller peaks of disease in young adults and the elderly (Fig. 7c).

Lineage 23 (cc23)

A Neighbor-Net graph of all lineage 23 isolates with whole-genome sequence (WGS) data in PubMLST (n=424), including the MRF-MGL CCCs (n=175), identified three sub-lineages (Fig. 8a) (Toros *et al.* 2015). The three sub-lineages were well supported: the level of support for splits in network graphs is relative to their weight and the edges separating sub-lineages 23.1, 23.2, and 23.3 were among the highest weighted with few contradictory splits (Appendix F1a). A Neighbor-Net graph (Appendix F1b) generated from variable nucleotide sites was congruent with that generated from allelic data (Fig. 8a) and edges among the sub-lineages had 100% bootstrap support following 1,000 replicates. Although isolates from a number of countries, including Sweden, comprised

lineages 23.1 and 23.2, lineage 23.3 was entirely formed of English and Welsh isolates (Fig. 8a). The predominant sub-lineage, lineage 23.1 (n=145), was composed mainly of ST-1655 isolates, however, the central genotype of the associated MLST clonal complex, ST-23, was associated with both lineages 23.1 and 23.2 (Fig. 8b). All lineage 23 isolates were genogroup Y except for single genogroup W/Y isolates in lineages 23.1 and 23.2. Lineage 23.1 declined from 2010/11-2012/13 in contrast to lineages 23.2 and 23.3, which fluctuated across the three epidemiological years (Fig. 8b). In common with the association of lineage 23 with older patients, lineage 23.1 was particularly prevalent in isolates from patients over the age of 50 years, although it also peaked in adolescents aged 15-19 years and was responsible for a small peak in infants under a year of age (Fig. 8c). The prevalence of sub-lineage 23.2 (n=25 cases) was insufficient for exploration of its association with age, however, its absence in infants under a year of age was unusual when compared to the general age distribution of *N. meningitidis* (Fig. 8c; Fig. 5).

Lineage 11 (cc11)

Lineage 11 isolates from CCCs (n=118) were associated with one of the two lineage 11 sub-lineages: lineage 11.2 (MLEE defined ET-15) and lineage 11.1 (Fig. 9a; Ch. 2; Ch. 3). A single isolate did not cluster with either sub-lineage, indicating that additional lineage 11 WGS data may reveal a third sub-lineage (Fig. 9a). Lineage 11 cases increased year on year: most lineage 11.1 meningococci were genogroup W (72 of 83 isolates) and these caused 23 additional cases by 2012/13; lineage 11.1 genogroup C meningococci caused five additional cases by 2012/13; most lineage 11.2 meningococci were genogroup C (26 of 35) and these caused an 11 additional cases by 2012/13 (Fig. 9b). The lineage 11 sub-lineages appeared to have distinct relationships with patient age (Fig. 9c). Lineage 11.1 meningococci were particularly prevalent in disease in patients over 50 years of age, although they were responsible for a smaller peak in 15-19 year olds and a minor peak in children under four years of age, similar to lineage 23.1 genogroup Y meningococci (Fig. 8c). Lineage 11.2 was more evenly distributed among age-groups, although it caused few cases in infants under a year of age and more cases in adults aged 30-39 years in comparison to lineage 11.1.

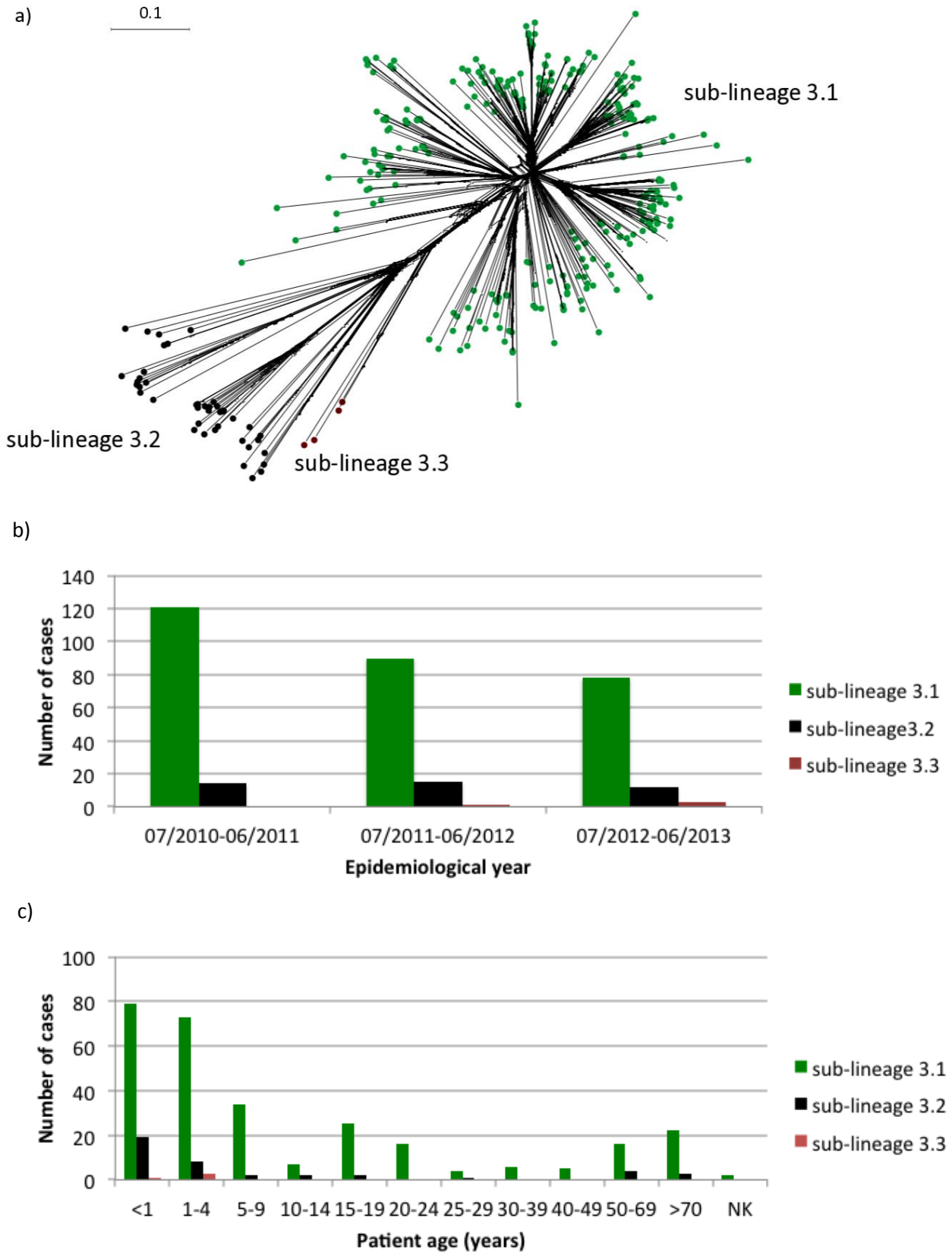


Figure 7. Characteristics of lineage 3 (cc41/44) sub-lineages. a) Neighbor-Net of allelic distances among 1,592 core-genome loci in lineage 3 isolates. **b)** Number of lineage 3 cases by epidemiological year. **c)** Number of lineage 3 cases among patients of different ages**.**

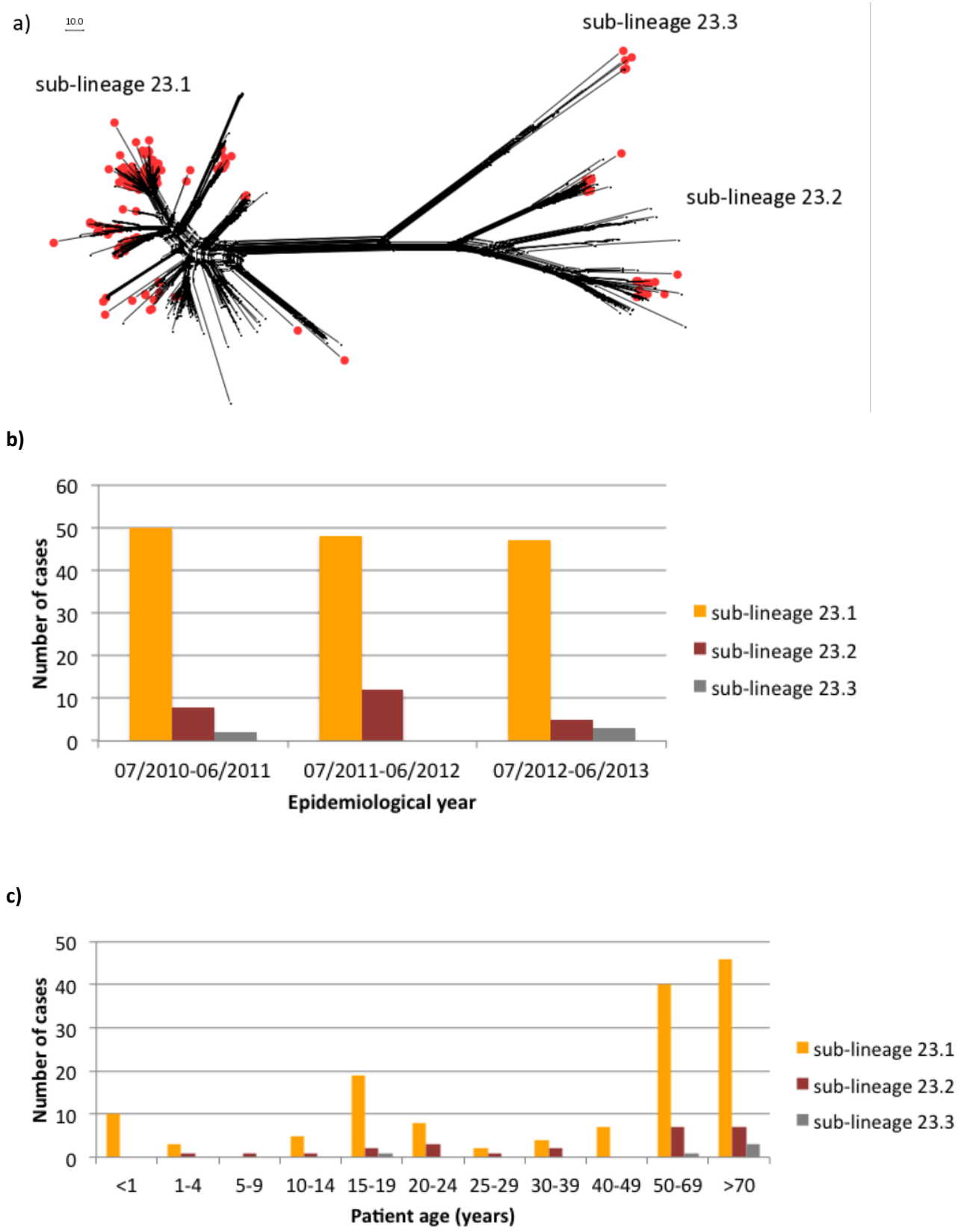


Figure 8. Characteristics of lineage 23 (cc23) sub-lineages. a) Neighbor-Net of allelic distances among 1,592 core-genome loci in 424 lineage 23 isolates in PubMLST: MRF-MGL isolates are indicated with red nodes. **b)** Number of MRF-MGL lineage 23 cases in epidemiological years. **c)** Number of MRF-MGL lineage 23 cases among patients of different ages.

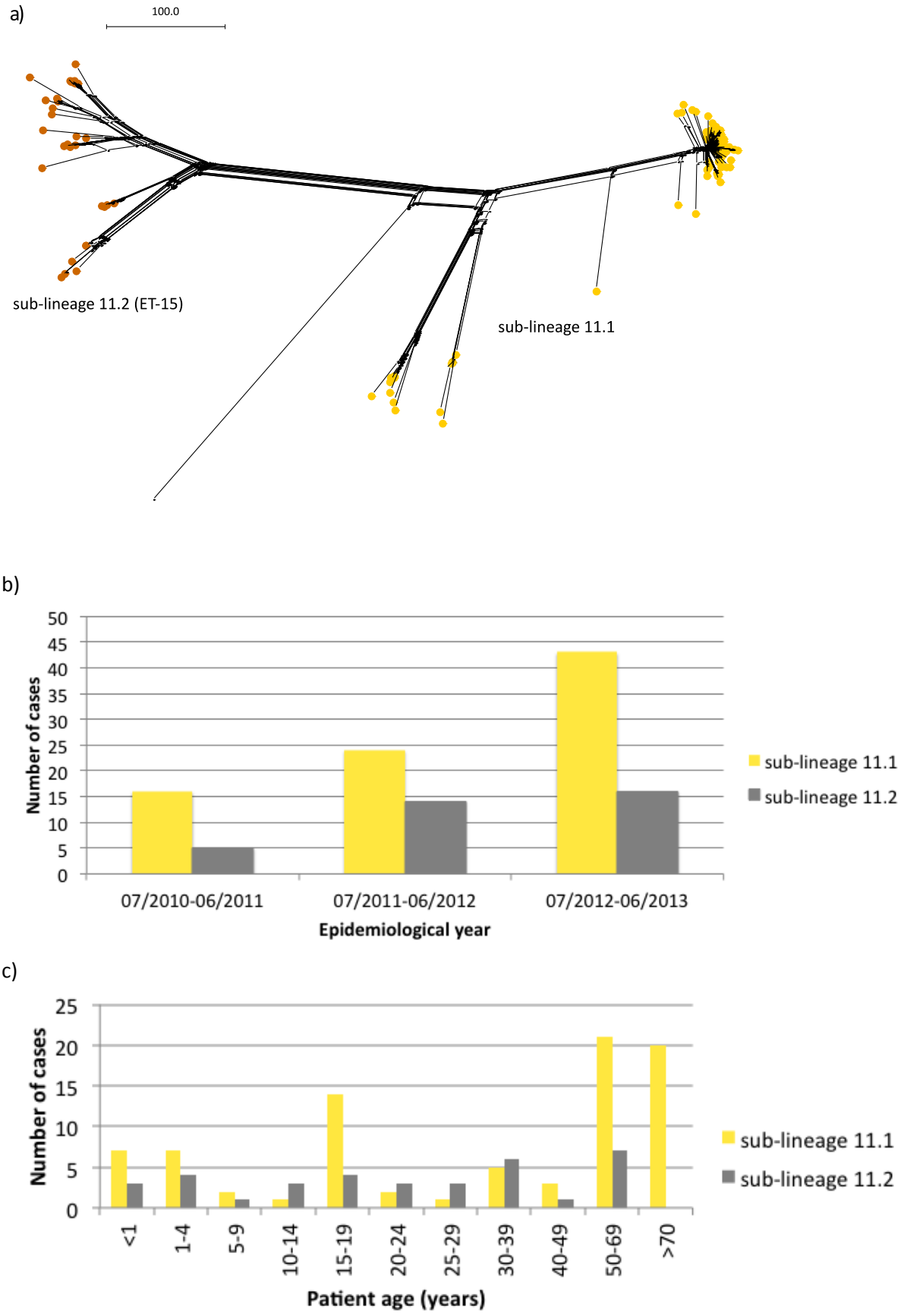


Figure 9. Characteristics of lineage 11 (cc11) sub-lineages. a) Neighbor-Net of allelic distances among 1,592 core-genome loci in lineage 11 isolates. **b)** Number of lineage 11 cases by epidemiological year. **c)** Number of lineage 11 cases among patients of different ages**.**

Lineage 2 (cc269)

Lineage 2 (cc269) isolates from CCCs (n=302) were associated with one of the two lineage 2 sub-lineages (lineages 2.1 and 2.2) and a putative third sub-lineage (lineage 2.3) (Fig. 10a; Ch. 2). As in 2010/11 and 2011/12, lineage 2.2 was the most prevalent, causing 60 cases compared to 40 lineage 2.1 cases. Sub-lineages followed the same general yearly prevalence fluctuations with reduced cases in 2011/12 and an increase in 2012/13 (Fig. 10b). There was, however, a different association between each sub-lineage and patient age: lineage 2.2 meningococcal disease was prevalent in infants and young children and caused a small secondary peak in patients aged over 50 years, whereas lineage 2.1 disease was prevalent in young children and adolescents but was rare in those >50 years of age (Fig. 10c). Binomial logistic regression found that lineage 2.2 disease was twice as likely as lineage 2.1 disease in children under five years (RRR 2.10 [1.54-2.88]), whereas lineage 2.1 disease was around five times as likely in patients 15-24 years of age (RRR 0.19 [0.08–0.46]).

Lineage 2 sub-lineages were unevenly distributed among UK regions in common with lineage 2 generally (Fig. 2; Fig. 11). Whilst lineage 2.2 disease was more prevalent than lineage 2.1 in Wales, the northern regions, South East, and South West of England, lineage 2.1 disease was more prevalent in the Midlands, Eastern England, and London (Fig. 11). This disparity was marked in the North West where there was three times as much, and in the North East where there was seven times as much, lineage 2.2 disease (Fig. 11). The greatest number of CCCs submitted to PHE-MRU was from the North West (n=228) (Fig. 2). Of these 56 (24.56%) were lineage 2.2 meningococci, more than 10% above the regional average. Only six (3.33%) of 180 isolates from London were lineage 2.2. Most patients with lineage 2.2 meningococcal disease in the North West were under five years of age (44 (78.57%) isolates); although the association between young children and lineage 2.2 was evident in all regions it was most extreme in the North West. The age demographic of all regions was similar (Office for National Statistics). The prevalence of lineage 2.2 disease in children under five years in the North West was apparent in each epidemiological year and was reported from an even range of hospitals.

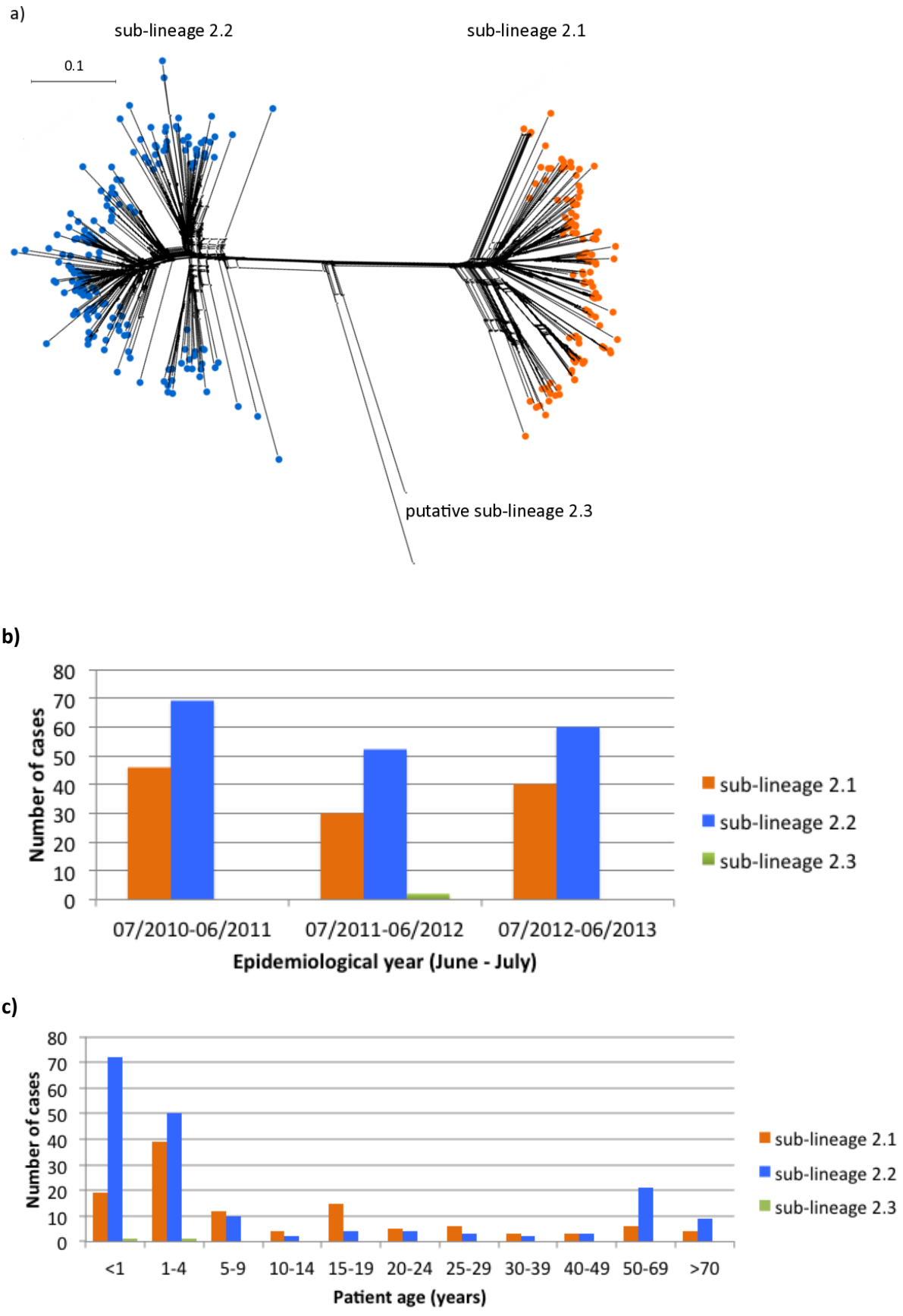


Figure 10. Characteristics of lineage 2 (cc269) sub-lineages. a) Neighbor-Net of allelic distances among 1,592 core-genome loci in 302 lineage 2 isolates. **b)** Number of lineage 2 cases by epidemiological year. **c)** Number of lineage 2 cases among patients of different ages.

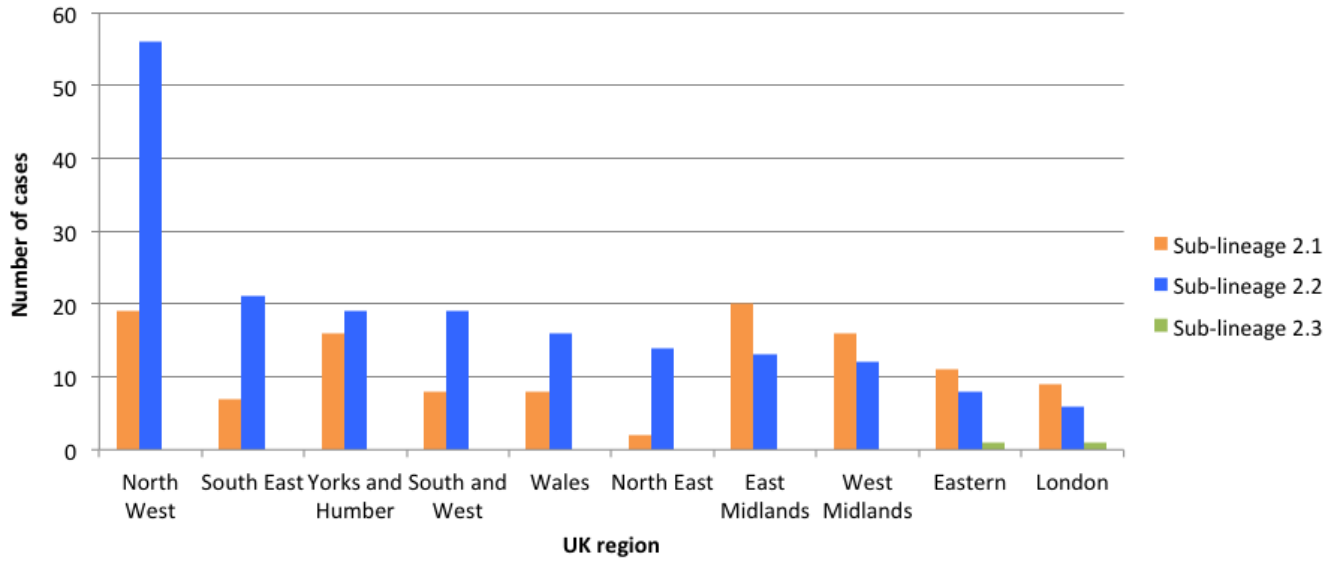


Figure 11. Culture-confirmed cases of lineage 2 meningococcal disease, England and Wales 2010/11-2012/13.

Genetic characteristics of lineage 2 sub-lineages

Alleles at 261 (16.3%) cgMLST loci completely distinguished lineage 2.1 from lineage 2.2 (Fig. 12). The majority did not have functional predictions, however, genes involved in amino acid-, carbohydrate-, and co-factor/vitamin-metabolism, and in replication and repair, each constituted around 7% of distinguishing loci; the latter two categories were particularly prevalent when compared to the lineage 11 Fam18 genome (Fig. 12). Six distinguishing loci were involved in host-meningococcal interactions: NEIS1108, which encodes the iron-uptake protein Bacterioferritin A (Perkins-Balding *et al.* 2004); NEIS1952, which encodes RegF, a regulator of PilE expression (De Reuse and Taha 1997); *tspA* (NEIS1829), the T-cell stimulating protein (Robinson *et al.* 2005); *gna93* (NEIS0010), an immunogenic outer-membrane lipoprotein (Adu-Bobie *et al.* 2004); and the meningococcal antigens *opcB* (NEIS1877) and *nhba* (NEIS2109) (Zhu *et al.* 1999, Serruto *et al.* 2010).

In addition, five accessory-genome loci were associated with lineage 2.1. The NEIS0973, NEIS0974, and NEIS0975 loci were identified in all lineage 2.1 meningococci but in only 50% of 2.2 meningococci. These loci did not have functional predictions in Fam18 but were found to be homologous to prophage ORFs from a DNA island (Klee *et al.* 2000): since they were adjacent on the

chromosome and were either all present or all absent in an isolate, they were likely to have been gained and lost in single recombination events. The NEIS1795 and NEIS1796 loci were identified in all lineage 2.1 isolates but in less than 1% of lineage 2.2 isolates. These were also chromosomally adjacent and both contained the 'immunity protein 22' domain, which is present in genes that are part of bacterial polymorphic toxin systems (Zhang *et al.* 2012). Three accessory-genome loci were identified in all lineage 2.2 isolates but not in lineage 2.1 isolates. These were the haemoglobin/haptoglobin receptors HpuA and HpuB (Perkins-Balding *et al.* 2004) and an immunogenic autotransporter protease, NalP (Turner *et al.* 2002).

Lineage 2 sub-lineages were antigenically distinct: NHBA variant 17 prevalent in lineage 2.2 (174/181 isolates) whereas variant 21 was prevalent in lineage 2.1 (110/116); fHbp variant family 1 peptide 13 was prevalent in lineage 2.2 (111/181 isolates) whereas family 1 peptide 15 was prevalent in lineage 2.1 (99/116); and PorA sub-type P1.22,9 was prevalent in lineage 2.2 (167/181 isolates), whereas sub-lineage 2.1 was associated with P1.9-1,15-11 (58/116). The *nadA* gene was identified only in two lineage 2.2 isolates where it contained frameshift mutations. Overall, 70 unique antigen peptide allele combinations were identified in lineage 2: each was specific to one sub-lineage. The dominant lineage 2.1 and sub-lineage 2.2 combinations did not share peptide variants; the average number of shared variants between two combinations was always greater within sub-lineages than between sub-lineages.

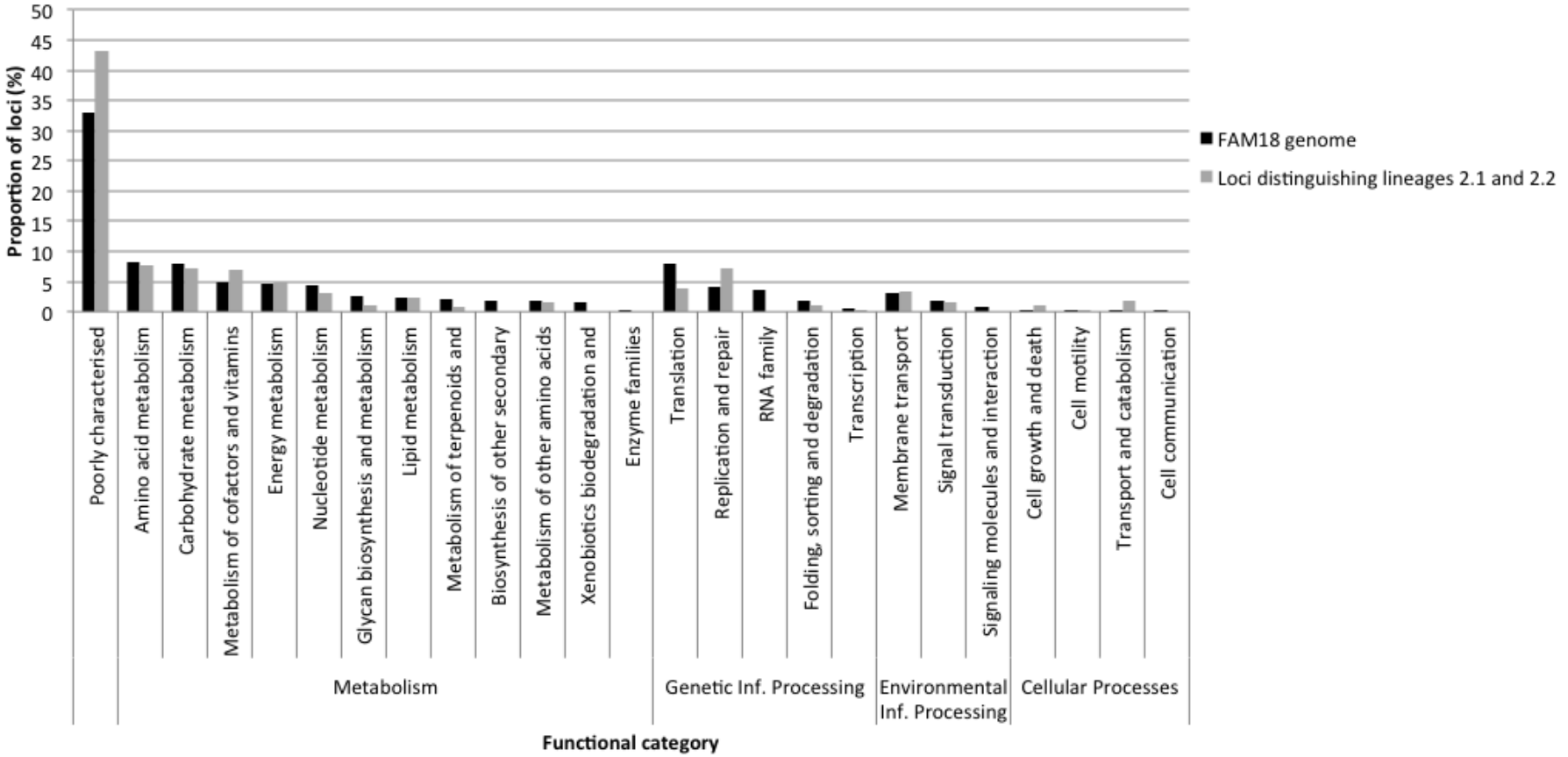


Figure 12. Functional categories of core-genome loci distinguishing lineage 2.1 from lineage 2.2. The sub-lineages of lineage 2 possessed 261 loci that had sub-lineage specific alleles. The functional categories of FAM18 genomic loci are provided for comparison.

Antigenic diversity of meningococci from patients of different ages

Antigen peptide diversity was lowest among isolates from patients under one year of age and from patients over 65 years of age, in accordance with lineage diversity in these age-groups (Table 1; Fig. 6a). Significant differences in peptide diversity were observed among antigens within and among age-groups. For example, fHbp diversity in patients aged 25-29 years was significantly greater (*D*=0.93 [0.91-0.95]) than in patients aged over 65 years (*D*=0.85 [0.81-0.89]) and PorA VR2 was significantly less diverse in patients under one year of age (*D*=0.88 [0.86-0.90]) than it was in patients aged 1-4 years and 15-64 years (Table 1, Fig. 4a).

Distinct vaccine antigen peptide variants were associated with each age-group. Patients aged less than 15 years were associated with fHbp sub-variants 1.13 or 1.4, whereas patients aged 15 years and older were associated with fHbp sub-variant 2.25 (Table 2). No age-groups were associated with Bexsero® fHbp sub-variant 1.1; this peptide was found at the lowest proportions in patients aged under one year and over 65 years. Most isolates from each age-group was missing *nadA*. Where *nadA* was present in patients under four years of age, it predominantly belonged to NadA-4/5 and contained frameshift mutations (p-value<0.05, χ^2). This was in contrast to isolates from patients aged 15-24 and over 40 years where the predominant peptide variant was NadA-2/3 allele 6 (p<0.05, χ^2 for each age category). Bexsero® variant NadA-3.8 was most frequently identified in patients aged over 65 years (8 isolates, 3.86% of age-group) and was identified in <2% of isolates from all other age-groups, except for the 1-4 and 15-24 year groups where it was absent.

Apart from among isolates from patients over 65 years of age, in which PorA P1.7-2 predominated, and from patients 40-64 years of age, in which NHBA-2 and PorA P1.7-2 were equally prevalent, NHBA-2 was the most frequently identified of the four Bexsero® antigen variants in any age category. NHBA-2 was the predominant variant in patients aged between one and 24 years (Table 2). The Bexsero® variant P1.7-2 was the second most prevalent PorA VR2 variant in age-groups 25-39 years and younger.

Table 2. Predominant antigen peptide variants among patients age-groups.

		Predominant antigen peptide variant (% isolates)				
Age (years)	**Number of isolates**	**fHbp**	**NadA**	**NHBA**	**PorA VR1**	**PorA VR2**
<1	300	1.13 (21.00)	FS (50.00)	17 (24.00)	22 (37.00)	9 (25.67)
1-4	315	1.4 (23.81)	FS (47.95)	**2 (24.13)**	22 (30.48)	14 (16.83)
5-14	99	1.4 (20.20)	1 (38.89)	**2 (20.20)**	22 (21.21)	9 (18.18)
15-24	168	2.25 (19.05)	6 (35.71)	**2 (20.83)**	5-1 (23.81)	4 & 10-1 (13.10)
25-39	73	2.25 & 1.39 (13.70)	FS (27.27)	20 (17.81)	5-1 (30.14)	14 (16.44)
40-64	173	2.25 (21.39)	6 (43.18)	7 (17.92)	5-1 (26.01)	2 (13.87)
65+	207	2.25 (34.30)	6 (38.78)	7 (28.99)	5-1 (36.23)	10-1 (20.29)
NK	9	1.4 & 1.13 (22.22)	FS & 6 (50.00)[+]	20 (22.22)	22 (33.33)	14 & 2 (22.22)
All	**1344**	**1.4 (16.29)**	**FS (34.77)**	**2 (17.41)**	**22 (24.48)**	**22 (24.48)**

Variants included in Bexsero® are shown in bold. The Novartis nomenclature for fHbp peptide variant is used here. [+]Only two isolates in the NK age category were found to possess *nadA*.

Overall, isolates possessing at least one of the four Bexsero® antigen peptide variants were found at the highest proportions in patients aged 1-4 and 5-14 years (114/315 (36.2%) and 34/99 (34.3%) isolates, respectively) and at the lowest proportions in patients aged over 65 years (37/207 (17.9%) isolates) (Fig. 13). The greatest proportion of genogroup B isolates with at least one exact amino-acid sequence match was identified in patients aged 15-24 years (48/107 (44.9%) genogroup B isolates) and the least was in patients aged <1 year (88/277 (31.8%) genogroup B isolates) (Fig. 13). The association of lineage 3 and rare lineages with patients under one year old meant that 30.3% of isolates contained at least one exact match to a Bexsero® peptide variant. When cross-protection was considered, isolates possessing any NadA-1, NadA-2/3, or fHbp family 1 variants, or PorA P1.4 or NHBA-2, were found at the highest proportion in patients aged 5 to 14 years (72/99 (72.7%) isolates). Of the vaccine formulations assessed, NonaMen components were the most frequently identified in all age-groups (Fig. 13), ranging from 76.3% of patients aged 40-64 years to 82.1% of patients over 65 years of age. A total of 81.7% of all isolates, and 82.3% of genogroup B isolates from patients aged less than one year, possessed at least one NonaMen PorA variant. Trumenba® (rLP2086) fHbp 45 and 55 were not found in isolates from patients aged 5-14 years (n=99).

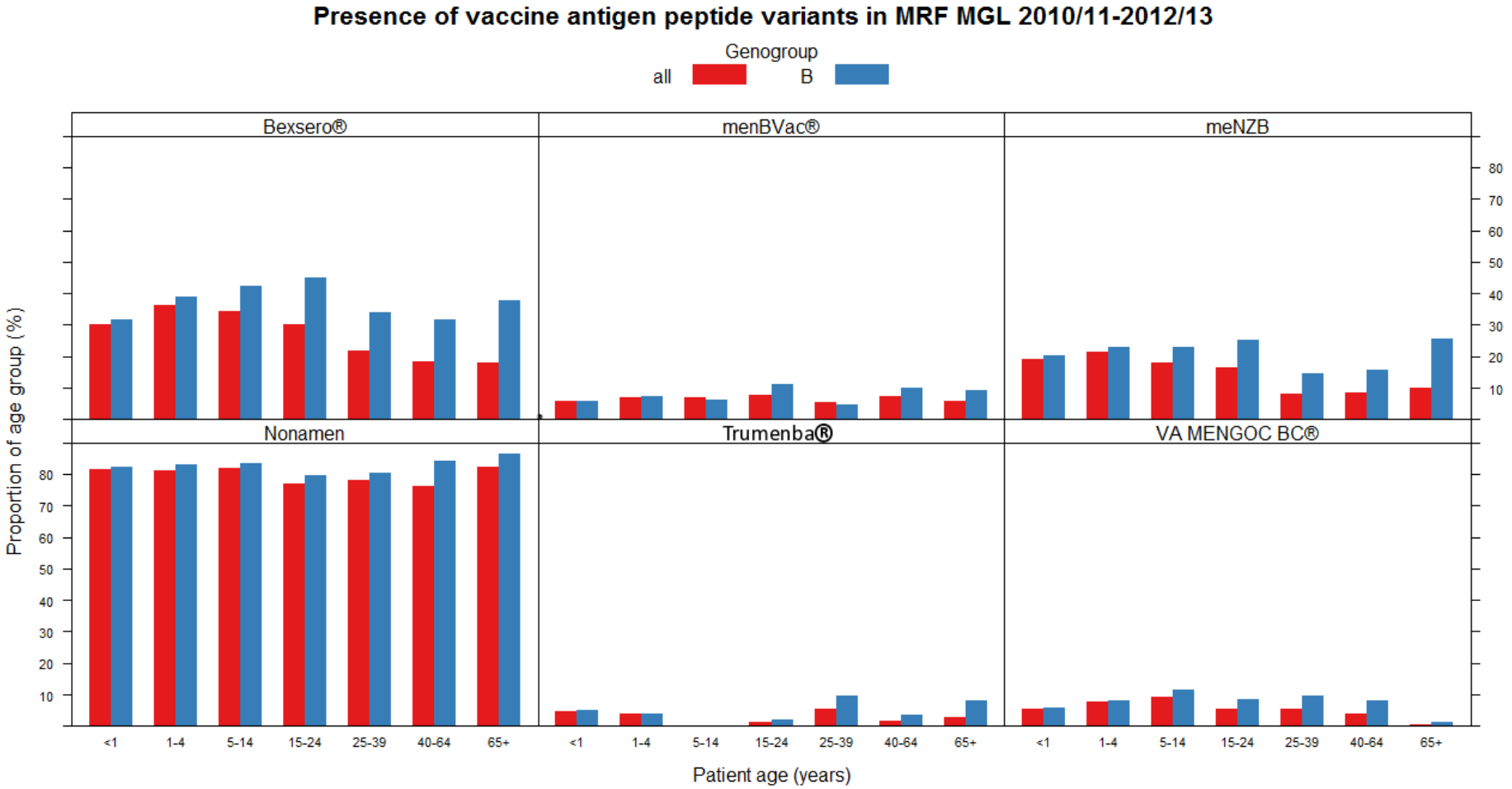


Figure 13. Proportion of isolates from patient age-groups with exact sequence matches to peptide sub-variants included in meningococcal vaccines.

Discussion

Highly variable pathogens require detailed characterisation for designing clinical and public health interventions. This is particularly true for meningococci that express the serogroup B polysaccharide, which are associated with a variety of meningococcal lineages and for which there is no anti-capsular vaccine. Molecular epidemiology has been successfully utilised in vaccine design and for targeted vaccine deployment: the meningococcal C polysaccharide vaccine was administered to students in one location during the 1993 Czech Republic epidemic (Kriz *et al.* 1995); meningococcal C conjugate vaccines were administered to those aged under 19 years in the UK in 1999 (Miller *et al.* 2001, Maiden *et al.* 2008); tailor-made outer-membrane vesicle vaccines were designed to contain PorA variants specific to serogroup B outbreak strains (Bjune *et al.* 1991, Sierra *et al.* 1991); and the broad spectrum NonaMen formulation was based on the most prevalent PorA serosubtypes documented in the Netherlands at the time of development (van der Ley *et al.* 1995, Kaaijk *et al.* 2013).

Whole-genome sequence (WGS) data permits bioinformatic identification of the complete repertoire of meningococcal antigens; indeed, a second generation 'serogroup B-substitute' vaccine, Bexsero®, includes novel antigens discovered through 'reverse vaccinology' within a single genome (Seib *et al.* 2012, Bambini *et al.* 2013). The inherent diversity of such antigens necessitates collection of antigen peptide sequence data from comprehensive epidemiological collections, for combination with immunological data, in order to estimate vaccine impact (Lucidarme *et al.* 2010, Lucidarme *et al.* 2013b, Holst *et al.* 2014). WGS data from surveillance is optimally efficient for such surveys. In combination with detailed descriptions of meningococcal population structure, geotemporal distributions, and patient information, these estimates could inform second generation serogroup B vaccine design and implementation strategies. For example, the UK is in a period of low meningococcal disease incidence caused by a diversity of meningococcal genotypes and particular age-groups are associated with particular antigenic variants (Fig. 1; Table 2); broad-spectrum

vaccines or phased immunization of at-risk groups may be appropriate (Fig. 13). Post-implementation, routine surveillance may detect temporal trends in lineage and antigen variant prevalence (Perez-Trallero *et al.* 2014), with particular forecasts requiring vaccine reformulation.

The MRF-MGL is an exemplar of a representative and contemporary database of WGS data from clinical meningococcal isolates. In June 2015, it was the most comprehensive epidemiological sample of meningococcal genome sequences available, and the PubMLST database, which supports the MRF-MGL, permitted assessment of vaccine antigen diversity and distribution among meningococcal lineages and patient age-groups. Although lineage 3 (cc41/44) has caused a third of cases in England and Wales in the past decade, this lineage declined year on year and caused only 21% of culture-confirmed cases (CCCs) in 2012/13 (Fig. 1; Fig. 7b) (Ladhani *et al.* 2012a). This was particularly encouraging given that: (i) the greatest decline was in lineage 3.1, which is associated with ST-41 and is therefore likely to be particularly invasive (Fig. 7) (Yazdankhah *et al.* 2004); (ii) lineage 3.1 is particularly associated with serogroup B disease in young children, who are at greatest risk of meningococcal disease (Fig. 5; Fig. 6; Fig. 7c); and (iii) Bexsero® antigen variants were predominantly identified among lineage 3 meningococci (Fig. 4). Should this trend continue, it is likely that immunization of infants with Bexsero® from September 2015 will combine to substantially reduce disease in children under five. Indeed, notwithstanding cross protection among variants, exact matches to Bexsero® components were identified in around 30% of isolates from patients aged under one year (Fig. 13).

Like lineage 3, lineage 2 has caused a third of meningococcal disease in England and Wales in recent years, and is particularly associated with young children (Fig. 1; Fig. 6b) (Ladhani *et al.* 2012a, Brehony *et al.* 2014). Comprehensive WGS and provenance data from CCCs permitted identification of two antigenically distinct lineage 2 sub-lineages that had significantly different distributions among patient age-groups: 20% of meningococcal disease cases in children under five was attributable to lineage 2.2, but neither sub-lineage was associated with the antigen peptide variants

included in Bexsero® (Fig. 10). A subset of genotypes comprising lineage 2.2 expanded in England and Wales from 2000 to 2008 (Lucidarme *et al.* 2009) although this trend was not apparent in these data with both sub-lineages following similar year-on-year trends (Fig. 10b). Notably, a large proportion of lineage 2.2 cases in patients under five occurred in the North West of England (Fig. 11); it is unclear whether this was a geographically restricted outbreak or due to particularly favorable lineage 2.2 transmission in this region.

The particular advantage of the gene-by-gene approach to genomic epidemiology is rapid identification of genetic and antigenic variants characteristic of meningococcal genotypes (Maiden *et al.* 2013). The locus variants particular to lineages 2.1 and 2.2 are an interesting precursor to genome-wide association studies (GWAS), which may find statistical links between genetic factors and environmental niches, such as hosts of certain age, exclusive of linkage disequilibrium (Fig. 12). For example, it was unlikely that the age-distributions of lineage 2 sub-lineages resulted from acquired immunity: trends were not linear, with both sub-lineages causing secondary peaks in distinct age-groups (Fig. 10c). Instead, it was plausible that lineage 2.1 is well adapted to transient carriage and rapid transmission among young children and among teenagers, who are likely to participate in high host density social scenarios and who experience the highest meningococcal carriage rates (MacLennan *et al.* 2006, Christensen *et al.* 2010, van Hoek *et al.* 2013). Analysis of a USA university outbreak identified intense social mixing as the primary risk factor for lineage 2.1 transmission in students (Mandal *et al.* 2013).

Of interest given this hypothesis was the absence of the haptoglobin-haemoglobin receptors HpuAB from lineage 2.1 (more prevalent in adolescents than 2.2) but the universal presence of these genes in lineage 2.2. Both sub-lineages possessed the haemoglobin receptor *hmbR*; presence of both receptors has been associated with hyperinvasive lineages that are rarely isolated from carriage compared to disease (Caugant and Maiden 2009, Harrison *et al.* 2009b, Harrison *et al.* 2013a). Second, *nalP* was specific to lineage 2.2. NalP is a phase-variable autotransporter whose expression

has been associated with a reduction in stable biofilm formation and transient carriage; this was seemingly contrary to its association with lineage 2.2 disease in infants (Turner *et al.* 2006, Arenas *et al.* 2013). Many of the loci characteristic of lineage 2 sub-lineages were involved in metabolic processes, which have been suggested to influence transmission efficiency (Buckee *et al.* 2008, Schoen *et al.* 2014). Future carriage studies will be of interest in identifying whether carriage of lineage 2.2 is unusually high in the North West and in young children, and GWAS of these loci with age-groups will be essential.

A relatively small proportion of isolates in age-groups above 40 years (group average 18%) contained exact sequence matches to Bexsero® components (Fig. 13). The major lineages responsible for meningococcal disease in these age-groups, lineages 11.1 and 23.1, also contained very few exact sequence matches to the variants included in Bexsero® (Fig. 4). These sub-lineages were associated with the serogroups W and Y, respectively, and therefore, although the proportion of genogroup B isolates possessing exact matches in the aforementioned age-groups appeared relatively high (average 35%) they numbered only 28 in each age-group (Fig. 13). The increasing prevalence of lineages 11.1 and 23.1 disease in England and Wales (Fig. 9) provides further evidence for decisions on alterations to vaccination schedules, since levels of natural immunity to serogroups W and Y are low in England (Fig. 8b; Fig. 9b) (Trotter *et al.* 2012). It is likely that the quadrivalent ACWY conjugate vaccine, to be introduced to adolescents and students starting university from September 2015, will be protective against the majority of lineage 11 and lineage 23 disease, subject to levels of capsule expression. If carriage studies indicate that transmission of lineage 11.1 and 23.1 meningococci is high in those aged 15-19 years, in line with disease peaks (Fig. 8c; Fig. 9c), addition of the quadrivalent vaccine to the MCC vaccine school program may also incur herd immunity that is protective against further increases in disease in patients over the age of 50 (Fig. 5) (Read *et al.* 2014). The unusual distribution of lineage 11.2 meningococci (ET-15) among patients of different ages (Fig. 9c) was likely to have been influenced by waning immunity to serogroup C meningococci among individuals immunized with the MCC vaccine as infants prior to 2006 (Pollard *et al.* 2013). The

proportionally greater burden of lineage 11 disease in London compared to other regions of the UK was plausibly the result of increased international travel in this region (Fig. 2).

Strong associations among lineages and antigen variants, and lineages and age-groups, demonstrated the potential for rational vaccine design and the importance of post-implementation monitoring (Fig. 3; Fig. 6). Immunologically distinct lineages and sub-lineages and the great diversity of antigen variants was indicative of diversifying selection resulting from immune system recognition of different antigen variants (Table 1). The presence of major, non-overlapping, combinations of vaccine antigen variants in the major hyperinvasive lineages and sub-lineages was consistent with models built on immune selection theory (Fig. 3). These models predict stable antigen-lineage relationships and therefore longevity of rationally designed vaccines (Callaghan *et al.* 2008, Buckee *et al.* 2010). Temporal and geographic stability of PorA and FetA variants, and of variant combinations, has been demonstrated in large isolate collections (Russell *et al.* 2008, Watkins and Maiden 2012). It is therefore likely that sub-lineage structuring by age may result in age-specific vaccine impacts (Fig. 9; Fig. 10).

The slightly fewer exact sequence matches to Bexsero® antigen variants in the MRF-MGL compared to isolate collections from other European countries (Holst *et al.* 2014, Perez-Trallero *et al.* 2014, Tzanakaki *et al.* 2014) was likely to be the result of particular meningococcal lineages circulating at the time of analysis. Europe and North America are currently experiencing low rates of meningococcal disease (Castelblanco *et al.* 2014, European Centre for Disease Prevention and Control 2014) and since a dominant epidemic clone is absent, multicomponent vaccine formulations would be required for broad coverage in these settings (Holst *et al.* 2014, Law *et al.* 2014). Here, multicomponent vaccines exhibited greater potential to protect against meningococci in England and Wales than vaccine formulations containing one or a few components (Fig. 13; Fig. 14). The majority of *nadA* genes in isolates from young children contained frameshift mutations and were likely to be nonfunctional (Table 2). However, while genotype data can provide valuable information

on the potential utility of vaccines, the evaluation of antigen expression and potential cross-reactivity is fundamental to gauging actual success of a given formulation (Martin and Snape 2013). Assays have been developed and expression studies carried out that attempt to predict the coverage of various meningococcal vaccine antigens in the population (Fletcher *et al.* 2004, Donnelly *et al.* 2010). One assay, the ELISA-based meningococcal antigen typing system (MATS), was developed to predict post-vaccination responses to the Bexsero® vaccine and the potential protection it offers (Donnelly *et al.* 2010). Estimates of Bexsero® coverage have reached almost 100% of serogroup B cases using European isolate panels (Lucidarme *et al.* 2010, Vogel *et al.* 2013). Coverage of lineage 2 cases was estimated to be almost 100% in Greece whereas Bexsero® antigen variants were rare in lineage 2 MRF-MGL isolates (Fig. 4) (Tzanakaki *et al.* 2014). One feature of the MRF-MGL is that phenotypic and genotypic information may be associated in isolate and allele records. Combined with monitoring of vaccine antigen peptide variants among culture-confirmed and non-culture-confirmed cases of meningococcal disease (Gray *et al.* 2006, Clark *et al.* 2014), the addition of MATS data to the MRF-MGL will permit essential future research and offer the best hope of producing vaccines with the widest possible coverage.

Chapter 6: Conclusions and future directions

The advances of the genomics era are particularly exciting for meningococcal molecular epidemiology. Now that thousands of bacterial whole-genome sequences are easily obtainable, the utility of these data in meningococcal disease reduction must be demonstrated to support their incorporation by national public health bodies. To this end, we set up the Meningitis Research Foundation Meningococcus Genome Library (MRF-MGL). This remains the largest collection of curated meningococcal genome sequences from an epidemiologically relevant population, and being openly available on the internet, it is a high quality resource for studies requiring meningococcal genetic data. In this thesis I aimed to show how differentiation of meningococci at the level of populations to individual antigenic variants could provide insight into meningococcal population biology relevant to public health applications. This thesis introduces the library and presents some of the first findings of meningococcal population genomics.

Population structure

The utility of meningococcal molecular epidemiology lies in identification of meningococcal lineages and strains. Throughout this thesis it becomes clear that MLEE and MLST defined complexes are also identified using rMLST and cgMLST loci in whole-genome sequence (WGS) data. A large number of isolates unassigned to MLST clonal complexes (cc) clustered into cc-level lineages. Lineage 2 (cc269) appeared to be a hybrid lineage: type III restriction-modification systems, metabolic loci, and vaccine antigens differed among strains, one of which had undergone recombination of its ribosomal loci (Ch. 2, 4-5). A lineage nomenclature, compatible with MLEE and MLST, was devised to permit assignment of meningococcal genomes to lineages in coming years (Ch. 2).

Immunologically distinct sub-lineages of prevalent hyperinvasive lineages were also visible. The novel nomenclature was extended to encompass sub-lineages since their identification will permit surveillance at a level of relatedness greater than the MLST cc whilst remaining practically useful. For

example, a sub-lineage of lineage 11 (cc11) not identified using MLST was named lineage 11.2: incidence of lineage 11.2 disease is increasing in the UK (Ch. 2, 3, 5) (Lucidarme *et al.* 2015). As WGS data from additional disease and carriage populations becomes available, research is required to develop methods for quantitative assignment of isolates to sub-lineages. For example, it may be possible to: devise an rMLST allele cut-off between isolates and central rSTs of sub-lineages, as with MLST assignments; to attempt a tree-based height cut-off with new WGS-specific phylogenetic methods such as ClonalFrame ML (Didelot and Wilson 2015); and to devise a statistical measure of within *versus* among sub-lineage diversity that delineates sub-lineages.

Maintenance of lineages

There was evidence that lineages and sub-lineages persist due to selective pressure. For example, each of the prevalent hyperinvasive lineages was associated with a different patient age-group (Ch. 2, 5). The extent to which this observation reflects adaptation to ecological niches and/or increased propensity for invasion in certain hosts may be clarified by future studies of age- and lineage-specific carriage rates and by genome-wide association studies to identify meningococcal genes associated with human age or behaviours (as with the association of the MDA phage with adolescents) (Bille *et al.* 2005, Falush and Bowden 2006). Population annotation (Bratcher *et al.* 2014) based on collections such as the MRF-MGL within the PubMLST database provides a starting point, however, the existence of two lineage 2 sub-lineages associated with distinct age-groups was suggestive of two fit phenotypes that reflect the characteristics of transmission during carriage (Ch. 5).

The genes that mediate differences in transmission fitness remain undefined; however, in this thesis the majority of loci with fixed lineage and sub-lineage specific alleles were those with metabolic functions (Ch. 3-5). It has recently been suggested that metabolic characteristics of lineages permit niche specific survival and transmission (Schoen *et al.* 2014). In Chapter 4, gene regulation by Mod phasevarions were examined. It was found that combinations of phasevarions are lineage and sub-lineage specific and that certain variants are associated with invasive disease. The modification sights of ModA12, the most pervasive variant, were mostly within metabolic loci. In addition, two

strains of the ET-15 clone responsible for outbreaks with contrasting epidemiological characteristics had different methylation profiles mediated by ModA12 (Ch. 3-4). Whether this reflects true strain characteristics could not be explored with these data, but research on within-host phase switching rates, for example using PacBio closed genomes, will increase our understanding of whether phasevarions lend reproductive fitness or short-sighted, within-host fitness (Levin and Bull 1994, Brown *et al.* 2012). The recognition sequence of ModA11 has recently been defined and here the recognition site of ModA12 was characterised (Seib *et al.* 2015a): functional studies of the genes regulated by these enzymes in different lineages, and under a variety of conditions, are now possible.

The relationship between transmission fitness and immune selection was first illustrated by the dynamic mathematical model of Buckee and colleagues (Buckee *et al.* 2008). There was evidence of immune selection activity in the non-overlapping alleles of vaccine components specific to prevalent hyperinvasive lineages and sub-lineages (Ch. 5). It will be particularly interesting to observe any fluctuations in these antigenic associations in the UK in coming years as the predominant lineage 3 (cc41/44) declines in prevalence and the Bexsero® vaccine is introduced.

Public health reporting and surveillance

One practical advantage of whole-genome sequencing is that most isolate genetic information is obtained in a single process; this increases the speed and decreases the human effort required for extraction of data pertinent to public health efforts. Here, variation in single genes, MLST loci, core-genome and accessory-genome loci were extracted to investigated the utility of WGS data in case characterisation, transmission investigation, population surveillance, and baseline vaccine coverage estimation (Ch. 2-3, 5). These findings demonstrated that routine collection of WGS data on a case-by-case basis would be of great advantage to public health efforts against meningococcal disease: Public Health England has agreed to extend the MRF-MGL in the coming years using MiSeq Illumina technology in real time.

Conventional, routinely derived, typing data could be accurately extracted from WGS data using the gene-by-gene approach in PubMLST (Gray *et al.* 2006, Maiden *et al.* 2013). Additional detail such as the phase-status of capsular genes and existence of hybrid W/Y capsule types was available, which could not be conclusively deduced using current methods (Ch. 2). MLEE ET assignments were also confirmed (Ch. 3). Some MLST loci were incompletely assembled, the result of repetitive sequence in the *fumC* locus: ongoing development of Illumina chemistry towards longer sequence reads is likely to resolve this in the near future. As technology becomes more affordable and skill sets alter, it is hoped that these approaches can be employed by NHS hospital laboratories. In particular, work is needed to link alleles at genes such as *rpoB* and *penA* to antibiotic susceptibility profiles: this will allow predictions from WGS data that can inform treatment (Taha *et al.* 2007).

Once WGS data is obtained from cases in real time it may be possible to identify outbreaks as they are developing and activate intervention measures at a previously unfeasible speed. Measurement of genome-wide genetic and epigenetic diversity at the lineage level (ch. 2), sub-lineage level (ch. 2-3, 5), and clone level (ch. 3) begins the process of defining the diversity cut-offs necessary to identify short-transmission chains for identification of outbreak sources. Additional work is needed to quantify within-host meningococcal diversity at single and longitudinal time points. Given that all isolates were unique at the core-genome level, it is likely that statistical approaches will be necessary (Didelot *et al.* 2014) and that WGS data will not replace comprehensive epidemiological data.

Spatiotemporal patterns of meningococcal diversity in the UK and the Czech Republic illustrated the importance of comprehensive WGS surveillance in understanding meningococcal strain dynamics following vaccine deployment. As in previous work (Buckee *et al.* 2007, Bharti *et al.* 2012) there was a role for host demographics in driving transmission: two disease hotspots in the Czech Republic epidemic were centered on major cities and were characterised by different ET-15 variants (Ch. 3); a particularly large proportion of disease in London was caused by lineage 11 (Ch. 5); and most disease in young children caused by lineage 2.2 occurred in the north west of England (Ch. 5). The increase in

lineage 11 (cc11) disease in young adults (Ch. 2, 5) should be halted by introduction of the quadrivalent vaccine in September, in part prompted by MRF-MGL data (Ladhani *et al.* 2015). Future surveillance with the MRF-MGL may improve understanding of the forces driving strain fluctuations and make detailed forecasting feasible.

Finally, an investigation of vaccine antigen diversity illustrated that representative WGS data from population samples is required to monitor effects of serogroup-B substitute vaccines in endemic periods (Ch. 5). The association of age-groups with immunologically distinct lineages and sub-lineages suggests that multivariant vaccines will provide broadest age-specific coverage. Future research linking MATS assay data to antigen variant profiles within the MRF-MGL will increase the power of both approaches, and annotated vaccine antigen variants in the MRF-MGL provide a comparison for data collected subsequent to the imminent introduction of Bexsero®.

To conclude, this thesis presents the Meningitis Research Foundation Meningococcus Genome Library and highlights aspects of meningococcal population biology that can be investigated with whole-genome sequence data for public health applications. The MRF-MGL and this D.Phil. were funded by a charitable donation intended to provide a lasting legacy for meningococcal disease research. To date, the MRF-MGL has been part of 11 published studies (Hubert *et al.* 2013, Hung *et al.* 2013, Lucidarme *et al.* 2013a, Oldfield *et al.* 2013, Bennett *et al.* 2014, Brehony *et al.* 2014, Chatt *et al.* 2014, Clark *et al.* 2014, van Tonder *et al.* 2014, Wormann *et al.* 2014, Ladhani *et al.* 2015, Lucidarme *et al.* 2015). With the planned extension of the MRF-MGL to coming years, it is clear that much insight into meningococcal biology will be achieved, not least in the wake of nationwide vaccine introduction from September 2015.

References

Abad, R., E. L. Lopez, R. Debbag and J. A. Vazquez (2014) "Serogroup W meningococcal disease: global spread and current affect on the Southern Cone in Latin America" Epidemiol Infect 142(12): 2461-2470

Abbott, J. D., D. M. Jones, M. J. Painter and S. E. Young (1985) "The epidemiology of meningococcal infections in England and Wales, 1912-1983" J Infect 11(3): 241-257

Achtman, M. (1994) "Clonal spread of serogroup A meningococci: a paradigm for the analysis of microevolution in bacteria" Mol Microbiol 11(1): 15-22

Achtman, M. (1995) "Epidemic spread and antigenic variability of *Neisseria meningitidis*" Trends Microbiol 3(5): 186-192

Achtman, M. (1997) "Microevolution and epidemic spread of serogroup A *Neisseria meningitidis*--a review" Gene 192(1): 135-140

Achtman, M., A. van der Ende, P. Zhu, I. S. Koroleva, B. Kusecek, G. Morelli, I. G. Schuurman, N. Brieske, et al. (2001) "Molecular epidemiology of serogroup a meningitis in Moscow, 1969 to 1997" Emerg Infect Dis 7(3): 420-427

Adamczyk-Poplawska, M., M. Lower and A. Piekarowicz (2009) "Characterization of the NgoAXP: phase-variable type III restriction-modification system in *Neisseria gonorrhoeae*" FEMS Microbiol Lett 300(1): 25-35

Adu-Bobie, J., P. Lupetti, B. Brunelli, D. Granoff, N. Norais, G. Ferrari, G. Grandi, R. Rappuoli, et al. (2004) "GNA33 of *Neisseria meningitidis* is a lipoprotein required for cell separation, membrane architecture, and virulence" Infect Immun 72(4): 1914-1919

Aho, E. L., J. A. Dempsey, M. M. Hobbs, D. G. Klapper and J. G. Cannon (1991) "Characterization of the opa (class 5) gene family of *Neisseria meningitidis*" Mol Microbiol 5(6): 1429-1437

Ala'Aldeen, D. A. (1996) "Transferrin receptors of *Neisseria meningitidis*: promising candidates for a broadly cross-protective vaccine" J Med Microbiol 44(4): 237-243

Ala'Aldeen, D. A., K. R. Neal, K. Ait-Tahar, J. S. Nguyen-Van-Tam, A. English, T. J. Falla, P. M. Hawkey and R. C. Slack (2000) "Dynamics of meningococcal long-term carriage among university students and their implications for mass vaccination" J Clin Microbiol 38(6): 2311-2316

Alexander, C. E., W. R. Sanborn, G. Cherriere, W. H. Crocker, Jr., P. E. Ewald and C. R. Kay (1968) "Sulfadiazine-resistant group A *Neisseria meningitidis*" Science 161(3845): 1019

Anderson, E. S. and A. Felix (1952) "Variation in Vi-phage II of *Salmonella typhi*" Nature 170(4325): 492-494

Anderson, R. M. and R. M. May (1979) "Population biology of infectious diseases: Part I" Nature 280(5721): 361-367

Ando, T., Q. Xu, M. Torres, K. Kusugami, D. A. Israel and M. J. Blaser (2000) "Restriction-modification system differences in *Helicobacter pylori* are a barrier to interstrain plasmid transfer" Mol Microbiol 37(5): 1052-1065

Andrews, S. M. and A. J. Pollard (2014) "A vaccine against serogroup B *Neisseria meningitidis*: dealing with uncertainty" Lancet Infect Dis 14(5): 426-434

Antignac, A., P. Kriz, G. Tzanakaki, J. M. Alonso and M. K. Taha (2001) "Polymorphism of *Neisseria meningitidis* penA gene associated with reduced susceptibility to penicillin" J Antimicrob Chemother 47(3): 285-296

Aragon, T. J. (2012). epitools: Epidemiology Tools. **R package version 0.5-7.**

Arber, W. (2000) "Genetic variation: molecular mechanisms and impact on microbial evolution" FEMS Microbiol Rev 24(1): 1-7

Archibald, F. S. and I. W. DeVoe (1978) "Iron in *Neisseria meningitidis*: minimum requirements, effects of limitation, and characteristics of uptake" J Bacteriol 136(1): 35-48

Arenas, J., R. Nijland, F. J. Rodriguez, T. N. Bosma and J. Tommassen (2013) "Involvement of three meningococcal surface-exposed proteins, the heparin-binding protein NhbA, the alpha-peptide of IgA protease and the autotransporter protease NalP, in initiation of biofilm formation" Mol Microbiol 87(2): 254-268

Arkwright, J. A. (1907) "On Variations of the Meningococcus and its Differentiation from other Cocci occurring in the Cerebro-Spinal Fluid" J Hyg (London) 7(2): 193-204

Arnold, R., Y. Galloway, A. McNicholas and J. O'Hallahan (2011) "Effectiveness of a vaccination programme for an epidemic of meningococcal B in New Zealand" Vaccine 29(40): 7100-7106

Artenstein, M. S., R. Gold, J. G. Zimmerly, F. A. Wyle, H. Schneider and C. Harkins (1970) "Prevention of meningococcal disease by group C polysaccharide vaccine" N Engl J Med 282(8): 417-420

Ashton, F. E., J. A. Ryan, A. Borczyk, D. A. Caugant, L. Mancino and D. Huang (1991) "Emergence of a Virulent Clone of *Neisseria meningitidis* Serotype-2a That Is Associated with Meningococcal Group-C Disease in Canada" J Clin Microbiol 29(11): 2489-2493

Baccarini, C., A. Ternouth, H. Wieffer and A. Vyse (2013) "The changing epidemiology of meningococcal disease in North America 1945-2010" Hum Vaccin Immunother 9(1): 162-171

Balmer, P., R. Borrow and E. Miller (2002) "Impact of meningococcal C conjugate vaccine in the UK" J Med Microbiol 51(9): 717-722

Bambini, S., M. De Chiara, A. Muzzi, M. Mora, J. Lucidarme, C. Brehony, R. Borrow, V. Masignani, et al. (2014) "Neisseria adhesin A variation and revised nomenclature scheme" Clin Vaccine Immunol 21(7): 966-971

Bambini, S., A. Muzzi, P. Olcen, R. Rappuoli, M. Pizza and M. Comanducci (2009) "Distribution and genetic variability of three vaccine components in a panel of strains representative of the diversity of serogroup B meningococcus" Vaccine 27(21): 2794-2803

Bambini, S., J. Piet, A. Muzzi, W. Keijzers, S. Comandi, L. De Tora, M. Pizza, R. Rappuoli, et al. (2013) "An analysis of the sequence variability of meningococcal fHbp, NadA and NHBA over a 50-year period in the Netherlands" PLoS One 8(5): e65043

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, et al. (2012) "SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing" J Comput Biol 19(5): 455-477

Barlow, A. K., J. E. Heckels and I. N. Clarke (1989) "The class 1 outer membrane protein of *Neisseria meningitidis*: gene sequence and structural and immunological similarities to gonococcal porins" Mol Microbiol 3(2): 131-139

Barrett, S. J. and P. H. Sneath (1994) "A numerical phenotypic taxonomic study of the genus *Neisseria*" Microbiology 140 (Pt 10): 2867-2891

Bayliss, C. D. (2009) "Determinants of phase variation rate and the fitness implications of differing rates for bacterial pathogens and commensals" FEMS Microbiol Rev 33(3): 504-520

Bayliss, C. D., M. J. Callaghan and E. R. Moxon (2006) "High allelic diversity in the methyltransferase gene of a phase variable type III restriction-modification system has implications for the fitness of *Haemophilus influenzae*" Nucleic Acids Res 34(14): 4046-4059

Bayliss, C. D., J. C. Hoe, K. Makepeace, P. Martin, D. W. Hood and E. R. Moxon (2008) "*Neisseria meningitidis* escape from the bactericidal activity of a monoclonal antibody is mediated by phase variation of lgtG and enhanced by a mutator phenotype" Infect Immun 76(11): 5038-5048

Bennett, J. S., S. D. Bentley, G. S. Vernikos, M. A. Quail, I. Cherevach, B. White, J. Parkhill and M. C. J. Maiden (2010) "Independent evolution of the core and accessory gene sets in the genus *Neisseria*: insights gained from the genome of *Neisseria lactamica* isolate 020-06" BMC Genomics 11

Bennett, J. S., D. T. Griffiths, N. D. McCarthy, K. L. Sleeman, K. A. Jolley, D. W. Crook and M. C. Maiden (2005) "Genetic diversity and carriage dynamics of *Neisseria lactamica* in infants" Infect Immun 73(4): 2424-2432

Bennett, J. S., K. A. Jolley, S. G. Earle, C. Corton, S. D. Bentley, J. Parkhill and M. C. Maiden (2012) "A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*" Microbiology 158(Pt 6): 1570-1580

Bennett, J. S., K. A. Jolley, P. F. Sparling, N. J. Saunders, C. A. Hart, I. M. Feavers and M. C. Maiden (2007) "Species status of *Neisseria gonorrhoeae*: evolutionary and epidemiological inferences from multilocus sequence typing" BMC Biol 5: 35

Bennett, J. S., E. R. Watkins, K. A. Jolley, O. B. Harrison and M. C. Maiden (2014) "Identifying Neisseria species by use of the 50S ribosomal protein L6 (rplF) gene" J Clin Microbiol 52(5): 1375-1381

Bentley, D. R., S. Balasubramanian, H. P. Swerdlow, G. P. Smith, J. Milton, C. G. Brown, K. P. Hall, D. J. Evers, et al. (2008) "Accurate whole human genome sequencing using reversible terminator chemistry" Nature 456(7218): 53-59

Bentley, S. D., G. S. Vernikos, L. A. S. Snyder, C. Churcher, C. Arrowsmith, T. Chillingworth, A. Cronin, P. H. Davis, et al. (2007) "Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18" PLoS Genet 3(2): 230-240

Bettinger, J. A., D. W. Scheifele, N. Le Saux, S. A. Halperin, W. Vaudry, R. Tsang and A. Members of Canadian Immunization Monitoring Program (2013) "The disease burden of invasive meningococcal serogroup B disease in Canada" Pediatr Infect Dis J 32(1): e20-25

Bharti, N., H. Broutin, R. F. Grais, M. J. Ferrari, A. Djibo, A. J. Tatem and B. T. Grenfell (2012) "Spatial dynamics of meningococcal meningitis in Niger: observed patterns in comparison with measles" Epidemiol Infect 140(8): 1356-1365

Bidmos, F. A., K. R. Neal, N. J. Oldfield, D. P. Turner, D. A. Ala'Aldeen and C. D. Bayliss (2011) "Persistence, replacement, and rapid clonal expansion of meningococcal carriage isolates in a 2008 university student cohort" J Clin Microbiol 49(2): 506-512

Bijlsma, M. W., V. Bekker, M. C. Brouwer, L. Spanjaard, D. van de Beek and A. van der Ende (2014) "Epidemiology of invasive meningococcal disease in the Netherlands, 1960-2012: an analysis of national surveillance data" Lancet Infect Dis 14(9): 805-812

Bille, E., R. Ure, S. J. Gray, E. B. Kaczmarski, N. D. McCarthy, X. Nassif, M. C. Maiden and C. R. Tinsley (2008) "Association of a bacteriophage with meningococcal disease in young adults" PLoS One 3(12): e3885

Bille, E., J. R. Zahar, A. Perrin, S. Morelle, P. Kriz, K. A. Jolley, M. C. Maiden, C. Dervin, et al. (2005) "A chromosomally integrated bacteriophage in invasive meningococci" J Exp Med 201(12): 1905-1913

Bjune, G., E. A. Hoiby, J. K. Gronnesby, O. Arnesen, J. H. Fredriksen, A. Halstensen, E. Holten, A. K. Lindbak, et al. (1991) "Effect of outer membrane vesicle vaccine against group B meningococcal disease in Norway" Lancet 338(8775): 1093-1096

Blackwell, C. C., G. Tzanakaki, J. Kremastinou, D. M. Weir, N. Vakalis, R. A. Elton, A. Mentis and N. Fatouros (1992) "Factors affecting carriage of *Neisseria meningitidis* among Greek military recruits" Epidemiol Infect 108(3): 441-448

Borrow, R., R. Abad, C. Trotter, F. R. van der Klis and J. A. Vazquez (2013) "Effectiveness of meningococcal serogroup C vaccine programmes" Vaccine 31(41): 4477-4486

Borrow, R., N. Andrews, D. Goldblatt and E. Miller (2001) "Serological basis for use of meningococcal serogroup C conjugate vaccines in the United Kingdom: reevaluation of correlates of protection" Infect Immun 69(3): 1568-1573

Boulton, I. C., A. R. Gorringe, B. Gorinsky, M. D. Retzer, A. B. Schryvers, C. L. Joannou and R. W. Evans (1999a) "Purified meningococcal transferrin-binding protein B interacts with a secondary, strain-specific, binding site in the N-terminal lobe of human transferrin" Biochem J 339 (Pt 1): 143-149

Boulton, I. C., A. R. Gorringe, J. K. Shergill, C. L. Joannou and R. W. Evans (1999b) "A dynamic model of the meningococcal transferrin receptor" J Theor Biol 198(4): 497-505

Boyer, H. W. (1971) "DNA restriction and modification mechanisms in bacteria" Annu Rev Microbiol 25: 153-176

Branham, S. E. (1953) "Serological relationships among meningococci" Bacteriol Rev 17(3): 175-188

Bratcher, H. B., C. Corton, K. A. Jolley, J. Parkhill and M. C. Maiden (2014) "A gene-by-gene population genomics platform: de novo assembly, annotation and genealogical analysis of 108 representative *Neisseria meningitidis* genomes" BMC Genomics 15: 1138

Brehony, C., C. L. Trotter, M. E. Ramsay, M. Chandra, K. A. Jolley, A. van der Ende, F. Carion, L. Berthelsen, et al. (2014) "Implications of differential age distribution of disease-associated meningococcal lineages for vaccine development" Clin Vaccine Immunol 21(6): 847-853

Brehony, C., D. J. Wilson and M. C. Maiden (2009) "Variation of the factor H-binding protein of *Neisseria meningitidis*" Microbiology 155(Pt 12): 4155-4169

Broker, M., S. Emonet, C. Fazio, S. Jacobsson, M. Koliou, M. Kuusi, D. Pace, M. Paragi, et al. (2015) "Meningococcal serogroup Y disease in Europe: Continuation of high importance in some European regions in 2013" Hum Vaccin Immunother 11(9): 2281-2286

Broker, M., S. Jacobsson, L. DeTora, D. Pace and M. K. Taha (2012a) "Increase of meningococcal serogroup Y cases in Europe: a reason for concern?" Hum Vaccin Immunother 8(5): 685-688

Broker, M., S. Jacobsson, M. Kuusi, D. Pace, M. J. Simoes, A. Skoczynska, M. K. Taha, M. Toropainen, et al. (2012b) "Meningococcal serogroup Y emergence in Europe: update 2011" Hum Vaccin Immunother 8(12): 1907-1911

Broome, C. V. (1986) "The Carrier State - *Neisseria Meningitidis*" J Antimicrob Chemother 18: 25-34

Brown, S. P., D. M. Cornforth and N. Mideo (2012) "Evolution of virulence in opportunistic pathogens: generalism, plasticity, and control" Trends Microbiol 20(7): 336-342

Bruen, T. C., H. Philippe and D. Bryant (2006) "A simple and robust statistical test for detecting the presence of recombination" Genetics 172(4): 2665-2681

Brundage, J. F., M. A. Ryan, B. H. Feighner and F. J. Erdtmann (2002) "Meningococcal disease among United States military service members in relation to routine uses of vaccines with different serogroup-specific components, 1964-1998" Clin Infect Dis 35(11): 1376-1381

Brunelli, B., E. Del Tordello, E. Palumbo, A. Biolchi, S. Bambini, M. Comanducci, A. Muzzi, M. Pizza, et al. (2011) "Influence of sequence variability on bactericidal activity sera induced by Factor H binding protein variant 1.1" Vaccine 29(5): 1072-1081

Bryant, J. M., D. M. Grogono, D. Greaves, J. Foweraker, I. Roddick, T. Inns, M. Reacher, C. S. Haworth, et al. (2013) "Whole-genome sequencing to identify transmission of Mycobacterium abscessus between patients with cystic fibrosis: a retrospective cohort study" Lancet 381(9877): 1551-1560

Bryant, P. A., H. Y. Li, A. Zaia, J. Griffith, G. Hogg, N. Curtis and J. R. Carapetis (2004) "Prospective study of a real-time PCR that is highly sensitive, specific, and clinically useful for diagnosis of meningococcal disease in children" J Clin Microbiol 42(7): 2919-2925

Buckee, C., L. Danon and S. Gupta (2007) "Host community structure and the maintenance of pathogen diversity" Proc Biol Sci 274(1619): 1715-1721

Buckee, C. O., S. Gupta, P. Kriz, M. C. Maiden and K. A. Jolley (2010) "Long-term evolution of antigen repertoires among carried meningococci" Proc Biol Sci 277(1688): 1635-1641

Buckee, C. O., K. A. Jolley, M. Recker, B. Penman, P. Kriz, S. Gupta and M. C. Maiden (2008) "Role of selection in the emergence of lineages and the evolution of virulence in *Neisseria meningitidis*" Proc Natl Acad Sci U S A 105(39): 15082-15087

Buckee, C. O., K. Koelle, M. J. Mustard and S. Gupta (2004) "The effects of host contact network structure on pathogen diversity and strain structure" Proc Natl Acad Sci U S A 101(29): 10839-10844

Buckee, C. O., M. Recker, E. R. Watkins and S. Gupta (2011) "Role of stochastic processes in maintaining discrete strain structure in antigenically diverse pathogen populations" Proc Natl Acad Sci U S A 108(37): 15504-15509

Budroni, S., E. Siena, J. C. D. Hotopp, K. L. Seib, D. Serruto, C. Nofroni, M. Comanducci, D. R. Riley, et al. (2011) "*Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination" Proc Natl Acad Sci U S A 108(11): 4494-4499

Callaghan, M. J., C. O. Buckee, K. A. Jolley, P. Kriz, M. C. Maiden and S. Gupta (2008) "The effect of immune selection on the structure of the meningococcal opa protein repertoire" PLoS Pathog 4(3): e1000020

Callaghan, M. J., S. Lewis, M. Sadarangani, S. E. Bailey, H. Chan, D. J. Ferguson, J. P. Derrick, I. Feavers, et al. (2011) "Potential of recombinant opa proteins as vaccine candidates against hyperinvasive meningococci" Infect Immun 79(7): 2810-2818

Caron, F., I. P. du Chatelet, J. P. Leroy, C. Ruckly, M. Blanchard, N. Bohic, N. Massy, I. Morer, et al. (2011) "From tailor-made to ready-to-wear meningococcal B vaccines: longitudinal study of a clonal meningococcal B outbreak" Lancet Infect Dis 11(6): 455-463

Cartwright, K. A., D. M. Jones, A. J. Smith, J. M. Stuart, E. B. Kaczmarski and S. R. Palmer (1991) "Influenza A and meningococcal disease" Lancet 338(8766): 554-557

Cartwright, K. A., J. M. Stuart and N. D. Noah (1986) "An outbreak of meningococcal disease in Gloucestershire" Lancet 2(8506): 558-561

Castelblanco, R. L., M. Lee and R. Hasbun (2014) "Epidemiology of bacterial meningitis in the USA from 1997 to 2010: a population-based observational study" Lancet Infect Dis 14(9): 813-819

Castresana, J. (2000) "Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis" Mol Biol Evol 17(4): 540-552

Caugant, D. A. (1998) "Population genetics and molecular epidemiology of *Neisseria meningitidis*" APMIS 106(5): 505-525

Caugant, D. A. (2008) "Genetics and evolution of *Neisseria meningitidis*: importance for the epidemiology of meningococcal disease" Infect Genet Evol 8(5): 558-565

Caugant, D. A., K. Bovre, P. Gaustad, K. Bryn, E. Holten, E. A. Hoiby and L. O. Froholm (1986a) "Multilocus genotypes determined by enzyme electrophoresis of *Neisseria Meningitidis* isolated from patients with systemic-disease and from healthy carriers" J Gen Microbiol 132: 641-652

Caugant, D. A., L. O. Froholm, K. Bovre, E. Holten, C. E. Frasch, L. F. Mocca, W. D. Zollinger and R. K. Selander (1986b) "Intercontinental spread of a genetically distinctive complex of clones of *Neisseria meningitidis* causing epidemic disease" Proc Natl Acad Sci U S A 83(13): 4927-4931

Caugant, D. A., E. A. Hoiby, P. Magnus, O. Scheel, T. Hoel, G. Bjune, E. Wedege, J. Eng, et al. (1994) "Asymptomatic carriage of *Neisseria meningitidis* in a randomly sampled population" J Clin Microbiol 32(2): 323-330

Caugant, D. A., E. A. Hoiby, E. Rosenqvist, L. O. Froholm and R. K. Selander (1992) "Transmission of *Neisseria meningitidis* among asymptomatic military recruits and antibody analysis" Epidemiol Infect 109(2): 241-253

Caugant, D. A. and M. C. Maiden (2009) "Meningococcal carriage and disease--population biology and evolution" Vaccine 27 Suppl 2: B64-70

Caugant, D. A., L. F. Mocca, C. E. Frasch, L. O. Froholm, W. D. Zollinger and R. K. Selander (1987a) "Genetic structure of *Neisseria meningitidis* populations in relation to serogroup, serotype, and outer membrane protein pattern" J Bacteriol 169(6): 2781-2792

Caugant, D. A. and P. Nicolas (2007) "Molecular surveillance of meningococcal meningitis in Africa" Vaccine 25 Suppl 1: A8-11

Caugant, D. A., G. Tzanakaki and P. Kriz (2007) "Lessons from meningococcal carriage studies" FEMS Microbiol Rev 31(1): 52-63

Caugant, D. A., W. D. Zollinger, L. F. Mocca, C. E. Frasch, T. S. Whittam, L. O. Froholm and R. K. Selander (1987b) "Genetic relationships and clonal population structure of serotype 2 strains of *Neisseria meningitidis*" Infect Immun 55(6): 1503-1512

Chatt, C., R. Gajraj, J. Hawker, K. Neal, M. Tahir, M. Lawrence, S. Gray, J. Lucidarme, et al. (2014) "Four-month outbreak of invasive meningococcal disease caused by a rare serogroup B strain, identified through the use of molecular PorA subtyping, England, 2013" Euro Surveill 19(44)

Chin, C. S., D. H. Alexander, P. Marks, A. A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, et al. (2013) "Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data" Nat Methods 10(6): 563-569

Christensen, H., M. May, L. Bowen, M. Hickman and C. L. Trotter (2010) "Meningococcal carriage by age: a systematic review and meta-analysis" Lancet Infect Dis 10(12): 853-861

Clark, S. A., J. Lucidarme, L. S. Newbold and R. Borrow (2014) "Genotypic analysis of meningococcal factor h-binding protein from non-culture clinical specimens" PLoS One 9(2): e89921

Clark, T. A., I. A. Murray, R. D. Morgan, A. O. Kislyuk, K. E. Spittle, M. Boitano, A. Fomenkov, R. J. Roberts, et al. (2012) "Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing" Nucleic Acids Res 40(4): e29

Claus, H., A. Friedrich, M. Frosch and U. Vogel (2000) "Differential distribution of novel restriction-modification systems in clonal lineages of *Neisseria meningitidis*" J Bacteriol 182(5): 1296-1303

Claus, H., M. C. Maiden, D. J. Wilson, N. D. McCarthy, K. A. Jolley, R. Urwin, F. Hessler, M. Frosch, et al. (2005) "Genetic analysis of meningococci carried by children and young adults" J Infect Dis 191(8): 1263-1271

Claus, H., M. C. J. Maiden, R. Maag, M. Frosch and U. Vogel (2002) "Many carried meningococci lack the genes required for capsule synthesis and transport" Microbiology-Sgm 148: 1813-1819

Cohen, C., E. Singh, H. M. Wu, S. Martin, L. de Gouveia, K. P. Klugman, S. Meiring, N. Govender, et al. (2010) "Increased incidence of meningococcal disease in HIV-infected individuals associated with higher case-fatality ratios in South Africa" AIDS 24(9): 1351-1360

Comanducci, M., S. Bambini, B. Brunelli, J. Adu-Bobie, B. Arico, B. Capecchi, M. M. Giuliani, V. Masignani, et al. (2002) "NadA, a novel vaccine candidate of *Neisseria meningitidis*" J Exp Med 195(11): 1445-1454

Comanducci, M., S. Bambini, D. A. Caugant, M. Mora, B. Brunelli, B. Capecchi, L. Ciucchi, R. Rappuoli, et al. (2004) "NadA diversity and carriage in *Neisseria meningitidis*" Infect Immun 72(7): 4217-4223

Croucher, N. J., S. R. Harris, C. Fraser, M. A. Quail, J. Burton, M. van der Linden, L. McGee, A. von Gottberg, et al. (2011) "Rapid pneumococcal evolution in response to clinical interventions" Science 331(6016): 430-434

Croucher, N. J., S. R. Harris, Y. H. Grad and W. P. Hanage (2013) "Bacterial genomes in epidemiology--present and future" Philos Trans R Soc Lond B Biol Sci 368(1614): 20120202

Crowe, B. A., T. Olyhoek, B. Neumann, B. Wall, M. Hassan-King, B. Greenwood and M. Achtman (1987) "A clonal analysis of *Neisseria meningitidis* serogroup A" Antonie Van Leeuwenhoek 53(6): 381-388

Crowe, B. A., R. A. Wall, B. Kusecek, B. Neumann, T. Olyhoek, H. Abdillahi, M. Hassan-King, B. M. Greenwood, et al. (1989) "Clonal and variable properties of *Neisseria meningitidis* isolated from cases and carriers during and after an epidemic in The Gambia, West Africa" J Infect Dis 159(4): 686-700

Darch, S. E., A. McNally, F. Harrison, J. Corander, H. L. Barr, K. Paszkiewicz, S. Holden, A. Fogarty, et al. (2015) "Recombination is a key driver of genomic and phenotypic diversity in a *Pseudomonas aeruginosa* population during cystic fibrosis infection" Sci Rep 5: 7649

Darling, A. E., B. Mau and N. T. Perna (2010) "progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement" PLoS One 5(6)

Daugla, D. M., J. P. Gami, K. Gamougam, N. Naibei, L. Mbainadji, M. Narbe, J. Toralta, B. Kodbesse, et al. (2014) "Effect of a serogroup A meningococcal conjugate vaccine (PsA-TT) on serogroup A meningococcal meningitis and carriage in Chad: a community study [corrected]" Lancet 383(9911): 40-47

Davidsen, T., E. A. Rodland, K. Lagesen, E. Seeberg, T. Rognes and T. Tonjum (2004) "Biased distribution of DNA uptake sequences towards genome maintenance genes" Nucleic Acids Res 32(3): 1050-1058

De Bolle, X., C. D. Bayliss, D. Field, T. van de Ven, N. J. Saunders, D. W. Hood and E. R. Moxon (2000) "The length of a tetranucleotide repeat tract in *Haemophilus influenzae* determines the phase variation rate of a gene with homology to type III DNA methyltransferases" Mol Microbiol 35(1): 211-222

De Reuse, H. and M. K. Taha (1997) "RegF, an SspA homologue, regulates the expression of the *Neisseria gonorrhoeae* pilE gene" Res Microbiol 148(4): 289-303

De Wals, P. and A. Bouckaert (1985) "Methods for estimating the duration of bacterial carriage" Int J Epidemiol 14(4): 628-634

Decosas, J. and J. B. Koama (2002) "Chronicle of an outbreak foretold: meningococcal meningitis W135 in Burkina Faso" Lancet Infect Dis 2(12): 763-765

Deghmane, A. E., D. Giorgini, M. Larribe, J. M. Alonso and M. K. Taha (2002) "Down-regulation of pili and capsule of *Neisseria meningitidis* upon contact with epithelial cells is mediated by CrgA regulatory protein" Mol Microbiol 43(6): 1555-1564

Delisle, E., S. Larrieu, J. Simoes, N. Laylle, M. De Pommerol, M. K. Taha, J. L. Termignon and I. Parent du Chatelet (2010) "Community outbreak of group B meningococcal disease in southwest France--December 2008 to September 2009" Euro Surveill 15(37)

Derrick, J. P., R. Urwin, J. Suker, I. M. Feavers and M. C. Maiden (1999) "Structural and evolutionary inference from molecular variation in *Neisseria* porins" Infect Immun 67(5): 2406-2413

Diaz Romero, J. and I. M. Outschoorn (1994) "Current status of meningococcal group B vaccine candidates: capsular or noncapsular?" Clin Microbiol Rev 7(4): 559-575

Didelot, X., R. Bowden, D. J. Wilson, T. E. Peto and D. W. Crook (2012) "Transforming clinical microbiology with bacterial genome sequencing" Nat Rev Genet 13(9): 601-612

Didelot, X. and D. Falush (2007) "Inference of bacterial microevolution using multilocus sequence data" Genetics 175(3): 1251-1266

Didelot, X., J. Gardy and C. Colijn (2014) "Bayesian inference of infectious disease transmission from whole-genome sequence data" Mol Biol Evol 31(7): 1869-1879

Didelot, X., R. Urwin, M. C. Maiden and D. Falush (2009) "Genealogical typing of *Neisseria meningitidis*" Microbiology 155(Pt 10): 3176-3186

Didelot, X. and D. J. Wilson (2015) "ClonalFrameML: efficient inference of recombination in whole bacterial genomes" PLoS Comput Biol 11(2): e1004041

Diermayer, M., K. Hedberg, F. Hoesly, M. Fischer, B. Perkins, M. Reeves and D. Fleming (1999) "Epidemic serogroup B meningococcal disease in Oregon: the evolving epidemiology of the ET-5 strain" JAMA 281(16): 1493-1497

Donnelly, J., D. Medini, G. Boccadifuoco, A. Biolchi, J. Ward, C. Frasch, E. R. Moxon, M. Stella, et al. (2010) "Qualitative and quantitative assessment of meningococcal antigens to evaluate the potential strain coverage of protein-based vaccines" Proc Natl Acad Sci U S A 107(45): 19490-19495

Drummond, A. J., M. A. Suchard, D. Xie and A. Rambaut (2012) "Bayesian phylogenetics with BEAUti and the BEAST 1.7" Mol Biol Evol 29(8): 1969-1973

Dryden, D. T., N. E. Murray and D. N. Rao (2001) "Nucleoside triphosphate-dependent restriction enzymes" Nucleic Acids Res 29(18): 3728-3741

Dupuis, M. E., M. Villion, A. H. Magadan and S. Moineau (2013) "CRISPR-Cas and restriction-modification systems are compatible and increase phage resistance" Nat Commun 4: 2087

Dybvig, K., R. Sitaraman and C. T. French (1998) "A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements" Proc Natl Acad Sci U S A 95(23): 13923-13928

Edgar, R. C. (2004) "MUSCLE: multiple sequence alignment with high accuracy and high throughput" Nucleic Acids Res 32(5): 1792-1797

Edmond, K., A. Clark, V. S. Korczak, C. Sanderson, U. K. Griffiths and I. Rudan (2010) "Global and regional risk of disabling sequelae from bacterial meningitis: a systematic review and meta-analysis" Lancet Infect Dis 10(5): 317-328

Edmond, M. B., R. J. Hollis, A. K. Houston and R. P. Wenzel (1995) "Molecular epidemiology of an outbreak of meningococcal disease in a university community" J Clin Microbiol 33(8): 2209-2211

Edwards, E. A., L. F. Devine, G. H. Sengbusch and H. W. Ward (1977) "Immunological investigations of meningococcal disease. III. Brevity of group C acquisition prior to disease occurrence" Scand J Infect Dis 9(2): 105-110

Eid, J., A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, et al. (2009) "Real-time DNA sequencing from single polymerase molecules" Science 323(5910): 133-138

Eldridge, J., E. M. Sutcliffe, J. D. Abbott and D. M. Jones (1978) "Serological grouping of meningococci and detection of antigen in cerebrospinal fluid by coagglutination" Med Lab Sci 35(1): 63-66

Elias, J., D. Harmsen, H. Claus, W. Hellenbrand, M. Frosch and U. Vogel (2006) "Spatiotemporal analysis of invasive meningococcal disease, Germany" Emerg Infect Dis 12(11): 1689-1695

Elias, J. and U. Vogel (2007) "IS1301 fingerprint analysis of *Neisseria meningitidis* strains belonging to the ET-15 clone" J Clin Microbiol 45(1): 159-167

Erickson, L. and P. De Wals (1998) "Complications and sequelae of meningococcal disease in Quebec, Canada, 1990-1994" Clin Infect Dis 26(5): 1159-1164

Ersts, P. J. Geographic Distance Matrix Generator, American Museum of Natural History, Center for Biodiversity and Conservation.

European Centre for Disease Prevention and Control (2014). Annual epidemiological report 2014 – Vaccine preventable diseases - invasive bacterial infections.

Excoffier, L., G. Laval and S. Schneider (2005) "Arlequin (version 3.0): an integrated software package for population genetics data analysis" Evol Bioinform Online 1: 47-50

Excoffier, L., P. E. Smouse and J. M. Quattro (1992) "Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data" Genetics 131(2): 479-491

Eyre, D. W., M. L. Cule, D. Griffiths, D. W. Crook, T. E. Peto, A. S. Walker and D. J. Wilson (2013a) "Detection of mixed infection from bacterial whole genome sequence data allows assessment of its role in Clostridium difficile transmission" PLoS Comput Biol 9(5): e1003059

Eyre, D. W., M. L. Cule, D. J. Wilson, D. Griffiths, A. Vaughan, L. O'Connor, C. L. Ip, T. Golubchik, et al. (2013b) "Diverse sources of C. difficile infection identified on whole-genome sequencing" N Engl J Med 369(13): 1195-1205

Fagnocchi, L., A. Biolchi, F. Ferlicca, G. Boccadifuoco, B. Brunelli, S. Brier, N. Norais, E. Chiarot, et al. (2013) "Transcriptional regulation of the nadA gene in *Neisseria meningitidis* impacts the prediction of coverage of a multicomponent meningococcal serogroup B vaccine" Infect Immun 81(2): 560-569

Fagnocchi, L., E. Pigozzi, V. Scarlato and I. Delany (2012) "In the NadR regulon, adhesins and diverse meningococcal functions are regulated in response to signals in human saliva" J Bacteriol 194(2): 460-474

Falush, D. and R. Bowden (2006) "Genome-wide association mapping in bacteria?" Trends Microbiol 14(8): 353-355

Falush, D., M. Stephens and J. K. Pritchard (2003) "Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies" Genetics 164(4): 1567-1587

Fang, G., D. Munera, D. I. Friedman, A. Mandlik, M. C. Chao, O. Banerjee, Z. Feng, B. Losic, et al. (2012) "Genome-wide mapping of methylated adenine residues in pathogenic Escherichia coli using single-molecule real-time sequencing" Nat Biotechnol 30(12): 1232-1239

Feavers, I. M., S. J. Gray, R. Urwin, J. E. Russell, J. A. Bygraves, E. B. Kaczmarski and M. C. Maiden (1999) "Multilocus sequence typing and antigen gene sequencing in the investigation of a meningococcal disease outbreak" J Clin Microbiol 37(12): 3883-3887

Feavers, I. M., J. Suker, A. J. McKenna, A. B. Heath and M. C. Maiden (1992) "Molecular analysis of the serotyping antigens of *Neisseria meningitidis*" Infect Immun 60(9): 3620-3629

Feil, E., J. Zhou, J. Maynard Smith and B. G. Spratt (1996) "A comparison of the nucleotide sequences of the adk and recA genes of pathogenic and commensal *Neisseria* species: evidence for extensive interspecies recombination within adk" J Mol Evol 43(6): 631-640

Feil, E. J. (2004) "Small change: keeping pace with microevolution" Nat Rev Microbiol 2(6): 483-495

Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage and B. G. Spratt (2004) "eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data" J Bacteriol 186(5): 1518-1530

Feil, E. J., M. C. Maiden, M. Achtman and B. G. Spratt (1999) "The relative contributions of recombination and mutation to the divergence of clones of *Neisseria meningitidis*" Mol Biol Evol 16(11): 1496-1502

Fiebelkorn, K. R., S. A. Crawford and J. H. Jorgensen (2005) "Mutations in folP associated with elevated sulfonamide MICs for *Neisseria meningitidis* clinical isolates from five continents" Antimicrob Agents Chemother 49(2): 536-540

Finn, R. D., A. Bateman, J. Clements, P. Coggill, R. Y. Eberhardt, S. R. Eddy, A. Heger, K. Hetherington, et al. (2014) "Pfam: the protein families database" Nucleic Acids Res 42(Database issue): D222-230

Finne, J., M. Leinonen and P. H. Makela (1983) "Antigenic similarities between brain components and bacteria causing meningitis. Implications for vaccine development and pathogenesis" Lancet 2(8346): 355-357

Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, et al. (1995) "Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd" Science 269(5223): 496-512

Fletcher, L. D., L. Bernfield, V. Barniak, J. E. Farley, A. Howell, M. Knauf, P. Ooi, R. P. Smith, et al. (2004) "Vaccine potential of the *Neisseria meningitidis* 2086 lipoprotein" Infect Immun 72(4): 2088-2100

Flusberg, B. A., D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, J. Korlach and S. W. Turner (2010) "Direct detection of DNA methylation during single-molecule, real-time sequencing" Nat Methods 7(6): 461-465

Ford, C. B., R. R. Shah, M. K. Maeda, S. Gagneux, M. B. Murray, T. Cohen, J. C. Johnston, J. Gardy, et al. (2013) "*Mycobacterium tuberculosis* mutation rate estimates from different lineages

predict substantial differences in the emergence of drug-resistant tuberculosis" Nat Genet 45(7): 784-790

Fox, K. L., Y. N. Srikhanta and M. P. Jennings (2007) "Phase variable type III restriction-modification systems of host-adapted bacterial pathogens" Mol Microbiol 65(6): 1375-1379

Frasch, C. E. and S. S. Chapman (1972) "Classification of *Neisseria meningitidis* group B into distinct serotypes. I. Serological typing by a microbactericidal method" Infect Immun 5(1): 98-102

Frasch, C. E., L. van Alphen, J. Holst, J. T. Poolman and E. Rosenqvist (2001) "Outer membrane protein vesicle vaccines for meningococcal disease" Methods Mol Med 66: 81-107

Frasch, C. E., W. D. Zollinger and J. T. Poolman (1985) "Serotype antigens of *Neisseria meningitidis* and a proposed scheme for designation of serotypes" Rev Infect Dis 7(4): 504-510

Fraser, C., W. P. Hanage and B. G. Spratt (2005) "Neutral microepidemic evolution of bacterial pathogens" Proc Natl Acad Sci U S A 102(6): 1968-1973

Frosi, G., A. Biolchi, M. Lo Sapio, F. Rigat, S. Gilchrist, J. Lucidarme, J. Findlow, R. Borrow, et al. (2013) "Bactericidal antibody against a representative epidemiological meningococcal serogroup B panel confirms that MATS underestimates 4CMenB vaccine strain coverage" Vaccine 31(43): 4968-4974

Frye, S. A., M. Nilsen, T. Tonjum and O. H. Ambur (2013) "Dialects of the DNA uptake sequence in neisseriaceae" PLoS Genet 9(4): e1003458

Fukuda, E., K. H. Kaminska, J. M. Bujnicki and I. Kobayashi (2008) "Cell death upon epigenetic genome methylation: a novel function of methyl-specific deoxyribonucleases" Genome Biol 9(11): R163

Furuta, Y., K. Abe and I. Kobayashi (2010) "Genome comparison and context analysis reveals putative mobile forms of restriction-modification systems and related rearrangements" Nucleic Acids Res 38(7): 2428-2443

Gawthorne, J. A., S. A. Beatson, Y. N. Srikhanta, K. L. Fox and M. P. Jennings (2012) "Origin of the diversity in DNA recognition domains in phasevarion associated modA genes of pathogenic *Neisseria* and *Haemophilus influenzae*" PLoS One 7(3): e32337

Gilca, R., G. Deceuninck, B. Lefebvre, R. Tsang, R. Amini, V. Gilca, M. Douville-Fradet, F. Markowski, et al. (2012) "The changing epidemiology of meningococcal disease in Quebec, Canada, 1991-2011: potential implications of emergence of new strains" PLoS One 7(11): e50659

Gilmore, A., G. Jones, M. Barker, N. Soltanpoor and J. M. Stuart (1999) "Meningococcal disease at the University of Southampton: outbreak investigation" Epidemiol Infect 123(2): 185-192

Giuliani, M. M., J. Adu-Bobie, M. Comanducci, B. Arico, S. Savino, L. Santini, B. Brunelli, S. Bambini, et al. (2006) "A universal vaccine for serogroup B meningococcus" Proc Natl Acad Sci U S A 103(29): 10834-10839

Glitza, I. C., I. Ehrhard, B. Muller-Pebody, R. Reintjes, T. Breuer, A. Ammon and H. G. Sonntag (2008) "Longitudinal study of meningococcal carrier rates in teenagers" Int J Hyg Environ Health 211(3-4): 263-272

Goldschneider, I., E. C. Gotschlich and M. S. Artenstein (1969) "Human immunity to the meningococcus. I. The role of humoral antibodies" J Exp Med 129(6): 1307-1326

Golubchik, T., A. B. Brueggemann, T. Street, R. E. Gertz, Jr., C. C. Spencer, T. Ho, E. Giannoulatou, R. Link-Gelles, et al. (2012) "Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event" Nat Genet 44(3): 352-355

Gordon, M. H. (1918) "Identification of the meningococcus" J Hyg (Lond) 17(2-3): 290-315

Gorringe, A. R., K. M. Reddin, S. G. Funnell, L. Johansson, A. Rytkonen and A. B. Jonsson (2005) "Experimental disease models for the assessment of meningococcal vaccines" Vaccine 23(17-18): 2214-2217

Gorringe, A. R., S. Taylor, C. Brookes, M. Matheson, M. Finney, M. Kerr, M. Hudson, J. Findlow, et al. (2009) "Phase I safety and immunogenicity study of a candidate meningococcal disease vaccine based on *Neisseria lactamica* outer membrane vesicles" Clin Vaccine Immunol 16(8): 1113-1120

Gossger, N., M. D. Snape, L. M. Yu, A. Finn, G. Bona, S. Esposito, N. Principi, J. Diez-Domingo, et al. (2012) "Immunogenicity and tolerability of recombinant serogroup B meningococcal vaccine administered with or without routine infant vaccinations according to different immunization schedules: a randomized controlled trial" JAMA 307(6): 573-582

Gotschlich, E. C., I. Goldschneider and M. S. Artenstein (1969a) "Human immunity to the meningococcus. IV. Immunogenicity of group A and group C meningococcal polysaccharides in human volunteers" J Exp Med 129(6): 1367-1384

Gotschlich, E. C., I. Goldschneider and M. S. Artenstein (1969b) "Human immunity to the meningococcus. V. The effect of immunization with meningococcal group C polysaccharide on the carrier state" J Exp Med 129(6): 1385-1395

Gotschlich, E. C., T. Y. Liu and M. S. Artenstein (1969c) "Human immunity to the meningococcus. 3. Preparation and immunochemical properties of the group A, group B, and group C meningococcal polysaccharides" J Exp Med 129(6): 1349-1365

Gould, S. J. and R. C. Lewontin (1979) "The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme" Proc R Soc Lond B Biol Sci 205(1161): 581-598

Gray, S. J., C. L. Trotter, M. E. Ramsay, M. Guiver, A. J. Fox, R. Borrow, R. H. Mallard and E. B. Kaczmarski (2006) "Epidemiology of meningococcal disease in England and Wales 1993/94 to 2003/04: contribution and experiences of the Meningococcal Reference Unit" J Med Microbiol 55(Pt 7): 887-896

Greenwood, B. M. and S. S. Wali (1980) "Control of meningococcal infection in the African meningitis belt by selective vaccination" Lancet 1(8171): 729-732

Guibourdenche, M., E. A. Hoiby, J. Y. Riou, F. Varaine, C. Joguet and D. A. Caugant (1996) "Epidemics of serogroup A *Neisseria meningitidis* of subgroup III in Africa, 1989-94" Epidemiol Infect 116(2): 115-120

Gulcher, J. and K. Stefansson (1998) "Population genomics: laying the groundwork for genetic disease modeling and targeting" Clin Chem Lab Med 36(8): 523-527

Gupta, S., M. C. Maiden, I. M. Feavers, S. Nee, R. M. May and R. M. Anderson (1996) "The maintenance of strain structure in populations of recombining infectious agents" Nat Med 2(4): 437-442

Guttman, D. S. and J. Stavrinides (2010). Population Genomics of Bacteria. Bacterial Population Genetics In Infectious Disease. A. Robinson, D. Falush and E. Feil, John Wiley & Sons, Inc.**:** 121-151.

Halperin, S. A., J. A. Bettinger, B. Greenwood, L. H. Harrison, J. Jelfs, S. N. Ladhani, P. McIntyre, M. E. Ramsay, et al. (2012) "The changing and dynamic epidemiology of meningococcal disease" Vaccine 30 Suppl 2: B26-36

Hammerschmidt, S., R. Hilse, J. P. van Putten, R. Gerardy-Schahn, A. Unkmeir and M. Frosch (1996a) "Modulation of cell surface sialic acid expression in *Neisseria meningitidis* via a transposable genetic element" EMBO J 15(1): 192-198

Hammerschmidt, S., A. Muller, H. Sillmann, M. Muhlenhoff, R. Borrow, A. Fox, J. van Putten, W. D. Zollinger, et al. (1996b) "Capsule phase variation in *Neisseria meningitidis* serogroup B by slipped-strand mispairing in the polysialyltransferase gene (siaD): correlation with bacterial invasion and the outbreak of meningococcal disease" Mol Microbiol 20(6): 1211-1220

Hao, W. L., J. H. Ma, K. Warren, R. S. W. Tsang, D. E. Low, F. B. Jamieson and D. C. Alexander (2011) "Extensive genomic variation within clonal complexes of *Neisseria meningitidis*" Genome Biol Evol 3: 1406-1418

Harris, S. R., E. J. Feil, M. T. Holden, M. A. Quail, E. K. Nickerson, N. Chantratita, S. Gardete, A. Tavares, et al. (2010) "Evolution of MRSA during hospital transmission and intercontinental spread" Science 327(5964): 469-474

Harrison, L. H. (2015) "Vaccines for prevention of group B meningococcal disease: Not your father's vaccines" Vaccine 33 Suppl 4: D32-38

Harrison, L. H., K. A. Jolley, K. A. Shutt, J. W. Marsh, M. O'Leary, L. T. Sanza and M. C. Maiden (2006) "Antigenic shift and increased incidence of meningococcal disease" J Infect Dis 193(9): 1266-1274

Harrison, L. H., C. L. Trotter and M. E. Ramsay (2009a) "Global epidemiology of meningococcal disease" Vaccine 27 Suppl 2: B51-63

Harrison, O. B., J. S. Bennett, J. P. Derrick, M. C. Maiden and C. D. Bayliss (2013a) "Distribution and diversity of the haemoglobin-haptoglobin iron-acquisition systems in pathogenic and non-pathogenic *Neisseria*" Microbiology 159(Pt 9): 1920-1930

Harrison, O. B., J. E. Bray, M. C. Maiden and D. A. Caugant (2015) "Genomic analysis of the evolution and global spread of hyper-invasive meningococcal lineage 5" EBioMedicine 2(3): 234-243

Harrison, O. B., H. Claus, Y. Jiang, J. S. Bennett, H. B. Bratcher, K. A. Jolley, C. Corton, R. Care, et al. (2013b) "Description and nomenclature of *Neisseria meningitidis* capsule locus" Emerg Infect Dis 19(4): 566-573

Harrison, O. B., N. J. Evans, J. M. Blair, H. S. Grimes, C. R. Tinsley, X. Nassif, P. Kriz, R. Ure, et al. (2009b) "Epidemiological evidence for the role of the hemoglobin receptor, hmbR, in meningococcal virulence" J Infect Dis 200(1): 94-98

Harrison, O. B., M. C. J. Maiden and B. Rokbi (2008) "Distribution of transferrin binding protein B gene (tbpB) variants among *Neisseria* species" BMC Microbiol 8

Hedge, J. and D. J. Wilson (2014) "Bacterial phylogenetic reconstruction from whole genomes is robust to recombination but demographic inference is not" MBio 5(6): e02158

Hedman, A. K., M. S. Li, P. R. Langford and J. S. Kroll (2012) "Transcriptional profiling of serogroup B *Neisseria meningitidis* growing in human blood: An approach to vaccine antigen discovery" PLoS One 7(6): e39718

Heimann, G. A., D. E. Mikel, M. Naus, V. Goel, M. Tischler, J. A. Carlson, R. Schabas, G. Pasut, et al. (1989) "Meningococcal disease in Ontario during the winter of 1988-1989" Can Dis Wkly Rep 15(11): 59-62

Heinsbroek, E., S. Ladhani, S. Gray, M. Guiver, E. Kaczmarski, R. Borrow and M. Ramsay (2013) "Added value of PCR-testing for confirmation of invasive meningococcal disease in England" J Infect 67(5): 385-390

Hill, D. J., N. J. Griffiths, E. Borodina and M. Virji (2010) "Cellular and molecular biology of *Neisseria meningitidis* colonization and invasive disease" Clin Sci (Lond) 118(9): 547-564

Hilse, R., J. Stoevesandt, D. A. Caugant, H. Claus, M. Frosch and U. Vogel (2000) "Distribution of the meningococcal insertion sequence IS1301 in clonal lineages of *Neisseria meningitidis*" Epidemiol Infect 124(2): 337-340

Hoiseth, S. K., E. Murphy, L. Andrew, U. Vogel, M. Frosch, W. Hellenbrand, R. Abad, J. A. Vazquez, et al. (2013) "A multi-country evaluation of *Neisseria meningitidis* serogroup B factor H-binding proteins and implications for vaccine coverage in different age groups" Pediatr Infect Dis J 32(10): 1096-1101

Holmes, E. C., R. Urwin and M. C. Maiden (1999) "The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*" Mol Biol Evol 16(6): 741-749

Holst, J., M. Comanducci, S. Bambini, A. Muzzi, S. Comandi, J. Oksnes, L. DeTora, M. Pizza, et al. (2014) "Variability of genes encoding surface proteins used as vaccine antigens in meningococcal endemic and epidemic strain panels from Norway" Vaccine 32(23): 2722-2731

Hotopp, J. C. D., R. Grifantini, N. Kumar, Y. L. Tzeng, D. Fouts, E. Frigimelica, M. Draghi, M. M. Giuliani, et al. (2006) "Comparative genomics of *Neisseria meningitidis*: core genome, islands of horizontal transfer and pathogen-specific genes" Microbiology-Sgm 152: 3733-3749

Hubert, K., N. Devos, I. Mordhorst, C. Tans, G. Baudoux, C. Feron, K. Goraj, J. Tommassen, et al. (2013) "ZnuD, a potential candidate for a simple and universal *Neisseria meningitidis* vaccine" Infect Immun 81(6): 1915-1927

Humbelin, M., B. Suri, D. N. Rao, D. P. Hornby, H. Eberle, T. Pripfl, S. Kenel and T. A. Bickle (1988) "Type III DNA restriction and modification systems EcoP1 and EcoP15. Nucleotide sequence of the EcoP1 operon, the EcoP15 mod gene and some EcoP1 mod mutants" J Mol Biol 200(1): 23-29

Hung, M. C., J. E. Heckels and M. Christodoulides (2013) "The adhesin complex protein (ACP) of *Neisseria meningitidis* is a new adhesin with vaccine potential" MBio 4(2)

Huson, D. H. (1998) "SplitsTree: analyzing and visualizing evolutionary data" Bioinformatics 14(1): 68-73

Huson, D. H. and D. Bryant (2006) "Application of phylogenetic networks in evolutionary studies" Mol Biol Evol 23(2): 254-267

Ibarz-Pavon, A. B., A. P. Lemos, M. C. Gorla, M. Regueira, I. Sireva Working Group and J. M. Gabastou (2012) "Laboratory-based surveillance of *Neisseria meningitidis* isolates from disease cases in Latin American and Caribbean countries, SIREVA II 2006-2010" PLoS One 7(8): e44102

Imrey, P. B., L. A. Jackson, P. H. Ludwinski, A. C. England, 3rd, G. A. Fella, B. C. Fox, L. B. Isdale, M. W. Reeves, et al. (1996) "Outbreak of serogroup C meningococcal disease associated with campus bar patronage" Am J Epidemiol 143(6): 624-630

Jackson, L. A., A. Schuchat, M. W. Reeves and J. D. Wenger (1995) "Serogroup C meningococcal outbreaks in the United States. An emerging threat" JAMA 273(5): 383-389

Janda, W. M., M. Bohnoff, J. A. Morello and S. A. Lerner (1980) "Prevalence and site-pathogen studies of *Neisseria meningitidis* and *N gonorrhoeae* in homosexual men" JAMA 244(18): 2060-2064

Jarva, H., S. Ram, U. Vogel, A. M. Blom and S. Meri (2005) "Binding of the complement inhibitor C4bp to serogroup B *Neisseria meningitidis*" J Immunol 174(10): 6299-6307

Jelfs, J. (1997). ETs - clonal diversity of the Australian kind? Australian Society for Microbiology Annual Scientific Meeting and Exhibition. Adelaide, Australia. **Symposium 43.1**.

Jelfs, J., B. Jalaludin, R. Munro, M. Patel, M. Kerr, D. Daley, S. Neville and A. Capon (1998) "A cluster of meningococcal disease in western Sydney, Australia initially associated with a nightclub" Epidemiol Infect 120(3): 263-270

Jelfs, J., R. Munro, F. E. Ashton and D. A. Caugant (2000) "Genetic characterization of a new variant within the ET-37 complex of *Neisseria meningitidis* associated with outbreaks in various parts of the world" Epidemiol Infect 125(2): 285-298

Jen, F. E., K. L. Seib and M. P. Jennings (2014) "Phasevarions mediate epigenetic regulation of antimicrobial susceptibility in Neisseria meningitidis" Antimicrob Agents Chemother 58(7): 4219-4221

Jodar, L., I. M. Feavers, D. Salisbury and D. M. Granoff (2002) "Development of vaccines against meningococcal disease" Lancet 359(9316): 1499-1508

Jolley, K., S. J. Gray, J. Suker and R. Urwin (2006). Methods for Typing of Meningococci. Handbook of Meningococcal Disease. M. Frosch and M. Maiden, WILEY-WCH**:** 37-52.

Jolley, K. A., C. M. Bliss, J. S. Bennett, H. B. Bratcher, C. Brehony, F. M. Colles, H. Wimalarathna, O. B. Harrison, et al. (2012a) "Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain" Microbiology 158(Pt 4): 1005-1015

Jolley, K. A., C. Brehony and M. C. Maiden (2007) "Molecular typing of meningococci: recommendations for target choice and nomenclature" FEMS Microbiol Rev 31(1): 89-96

Jolley, K. A., D. M. Hill, H. B. Bratcher, O. B. Harrison, I. M. Feavers, J. Parkhill and M. C. Maiden (2012b) "Resolution of a meningococcal disease outbreak from whole-genome sequence data with rapid Web-based analysis methods" J Clin Microbiol 50(9): 3046-3053

Jolley, K. A., J. Kalmusova, E. J. Feil, S. Gupta, M. Musilek, P. Kriz and M. C. J. Maiden (2000) "Carried meningococci in the Czech Republic: A diverse recombining population" J Clin Microbiol 38(12): 4492-4498

Jolley, K. A. and M. C. Maiden (2010) "BIGSdb: Scalable analysis of bacterial genome variation at the population level" BMC Bioinformatics 11: 595

Jolley, K. A. and M. C. Maiden (2013) "Automated extraction of typing information for bacterial pathogens from whole genome sequence data: Neisseria meningitidis as an exemplar" Euro Surveill 18(4): 20379

Jolley, K. A., D. J. Wilson, P. Kriz, G. McVean and M. C. Maiden (2005) "The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in *Neisseria meningitidis*" Mol Biol Evol 22(3): 562-569

Jones, C., M. Virji and P. R. Crocker (2003) "Recognition of sialylated meningococcal lipopolysaccharide by siglecs expressed on myeloid cells leads to enhanced bacterial uptake" Mol Microbiol 49(5): 1213-1225

Jones, D. M. (1988) "Epidemiology of meningococcal infection in England and Wales" J Med Microbiol 26(3): 165-168

Jones, D. M., R. Borrow, A. J. Fox, S. Gray, K. A. Cartwright and J. T. Poolman (1992) "The lipooligosaccharide immunotype as a virulence determinant in *Neisseria meningitidis*" Microb Pathog 13(3): 219-224

Joseph, B., S. Schneiker-Bekel, A. Schramm-Gluck, J. Blom, H. Claus, B. Linke, R. F. Schwarz, A. Becker, et al. (2010) "Comparative genome biology of a serogroup B carriage and disease strain supports a polygenic nature of meningococcal virulence" J Bacteriol 192(20): 5363-5377

Joseph, B., R. F. Schwarz, B. Linke, J. Blom, A. Becker, H. Claus, A. Goesmann, M. Frosch, et al. (2011) "Virulence evolution of the human pathogen *Neisseria meningitidis* by recombination in the core and accessory genome" PLoS One 6(4)

Kaaijk, P., I. van Straaten, B. van de Waterbeemd, E. P. Boot, L. M. Levels, H. H. van Dijken and G. P. van den Dobbelsteen (2013) "Preclinical safety and immunogenicity evaluation of a

nonavalent PorA native outer membrane vesicle vaccine against serogroup B meningococcal disease" Vaccine 31(7): 1065-1071

Kaczmarski, E. B. (1997) "Meningococcal disease in England and Wales: 1995" Commun Dis Rep CDR Rev 7(4): R55-59

Katz, L. S., J. C. Humphrey, A. B. Conley, V. Nelakuditi, A. O. Kislyuk, S. Agrawal, P. Jayaraman, B. H. Harcourt, et al. (2011) "Neisseria Base: a comparative genomics database for *Neisseria meningitidis*" Database (Oxford) 2011: bar035

Klee, S. R., X. Nassif, B. Kusecek, P. Merker, J. L. Beretti, M. Achtman and C. R. Tinsley (2000) "Molecular and biological analysis of eight genetic islands that distinguish Neisseria meningitidis from the closely related pathogen *Neisseria gonorrhoeae*" Infect Immun 68(4): 2082-2095

Kobayashi, I. (2001) "Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution" Nucleic Acids Res 29(18): 3742-3756

Kohl, T. A., R. Diel, D. Harmsen, J. Rothganger, K. M. Walter, M. Merker, T. Weniger and S. Niemann (2014) "Whole-genome-based *Mycobacterium tuberculosis* surveillance: a standardized, portable, and expandable approach" J Clin Microbiol 52(7): 2479-2486

Koren, S., M. C. Schatz, B. P. Walenz, J. Martin, J. T. Howard, G. Ganapathy, Z. Wang, D. A. Rasko, et al. (2012) "Hybrid error correction and de novo assembly of single-molecule sequencing reads" Nat Biotechnol 30(7): 693-700

Korlach, J. and S. W. Turner (2012) "Going beyond five bases in DNA sequencing" Curr Opin Struct Biol 22(3): 251-261

Kortekaas, J., A. Pettersson, J. van der Biezen, V. E. Weynants, P. van der Ley, J. Poolman, M. P. Bos and J. Tommassen (2007) "Shielding of immunogenic domains in *Neisseria meningitidis* FrpB (FetA) by the major variable region" Vaccine 25(1): 72-84

Koser, C. U., J. M. Bryant, J. Becq, M. E. Torok, M. J. Ellington, M. A. Marti-Renom, A. J. Carmichael, J. Parkhill, et al. (2013) "Whole-genome sequencing for rapid susceptibility testing of M. tuberculosis" N Engl J Med 369(3): 290-292

Koser, C. U., L. J. Fraser, A. Ioannou, J. Becq, M. J. Ellington, M. T. Holden, S. Reuter, M. E. Torok, et al. (2014) "Rapid single-colony whole-genome sequencing of bacterial pathogens" J Antimicrob Chemother 69(5): 1275-1281

Krauland, M. G., J. C. D. Hotopp, D. R. Riley, S. C. Daugherty, J. W. Marsh, N. E. Messonnier, L. W. Mayer, H. Tettelin, et al. (2012) "Whole genome sequencing to investigate the emergence of clonal complex 23 *Neisseria meningitidis* serogroup Y disease in the United States" PLoS One 7(4)

Kremastinou, J., C. Blackwell, G. Tzanakaki, C. Kallergi, R. Elton and D. Weir (1994) "Parental smoking and carriage of *Neisseria meningitidis* among Greek schoolchildren" Scand J Infect Dis 26(6): 719-723

Kremastinou, J., G. Tzanakaki, A. Pagalis, M. Theodondou, D. M. Weir and C. C. Blackwell (1999) "Detection of IgG and IgM to meningococcal outer membrane proteins in relation to carriage of *Neisseria meningitidis* or *Neisseria lactamica*" FEMS Immunol Med Microbiol 24(1): 73-78

Kristiansen, P. A., A. K. Ba, A. S. Ouedraogo, I. Sanou, R. Ouedraogo, L. Sangare, F. Diomande, D. Kandolo, et al. (2014) "Persistent low carriage of serogroup A Neisseria meningitidis two years after mass vaccination with the meningococcal conjugate vaccine, MenAfriVac" BMC Infect Dis 14: 663

Kristiansen, P. A., A. K. Ba, I. Sanou, A. S. Ouedraogo, R. Ouedraogo, L. Sangare, F. Diomande, D. Kandolo, et al. (2013) "Phenotypic and genotypic characterization of meningococcal carriage and disease isolates in Burkina Faso after mass vaccination with a serogroup a conjugate vaccine" BMC Infect Dis 13: 363

Kriz, P. (2004) "Surveillance of invasive meningococcal disease in the Czech Republic" Euro Surveill 9(11): 37-39

Kriz, P., D. Giorgini, M. Musilek, M. Larribe and M. K. Taha (1999a) "Microevolution through DNA exchange among strains of *Neisseria meningitidis* isolated during an outbreak in the Czech Republic" Res Microbiol 150(4): 273-280

Kriz, P., B. Kriz, E. Svandova and M. Musilek (1999b) "Antimeningococcal herd immunity in the Czech Republic--influence of an emerging clone, *Neisseria meningitidis* ET-15/37" Epidemiol Infect 123(2): 193-200

Kriz, P. and M. Musilek (1996). Antigenic variations within the genetic clone ET-15/37 of *Neisseria meningitidis* occurring in the Czech Republic. 10th Internation Pathogenic Neisseria Conference. Baltimore**:** 482-483.

Kriz, P., J. Vlckova and M. Bobak (1995) "Targeted vaccination with meningococcal polysaccharide vaccine in one district of the Czech Republic" Epidemiol Infect 115(3): 411-418

Kriz, P., H. Wieffer, K. Holl, M. Rosenlund, S. Budhia and A. Vyse (2011) "Changing epidemiology of meningococcal disease in Europe from the mid-20th to the early 21st Century" Expert Rev Vaccines 10(10): 1477-1486

Krizova, P. and M. Musilek (1995) "Changing epidemiology of meningococcal invasive disease in the Czech republic caused by new clone *Neisseria meningitidis* C:2a:P1.2(P1.5), ET-15/37" Cent Eur J Public Health 3(4): 189-194

Krizova, P. and M. Musilek (1997) "Development of the epidemiological situation in invasive meningococcal disease in the Czech Republic cause by emerging *Neisseria meningitidis* clone ET-15/37" Centr eur J publ Hlth 5(4): 214-218

Krizova, P., M. Musilek and V. Lebedova (1994) "[The new epidemiologic situation in the Czech Republic caused by the C:2a:P1.2 (P1.5) strain of meningococci]" Epidemiol Mikrobiol Imunol 43(4): 183-187

Krizova-Kuzemenska, P., C. Benes, J. Dlhy, S. Kartusek, J. Matuska, J. Vlckova, A. Galetkova and L. Roznovsky (1993) "[Increased incidence of invasive meningococcal disease caused by C:2a:P1.2 meningococci in certain regions of the Czech Republic]" Cesk Epidemiol Mikrobiol Imunol 42(4): 165-171

Kroll, J. S., K. E. Wilks, J. L. Farrant and P. R. Langford (1998) "Natural genetic exchange between *Haemophilus* and *Neisseria*: intergeneric transfer of chromosomal genes between major human pathogens" Proc Natl Acad Sci U S A 95(21): 12381-12385

Kupferschmidt, K. (2013) "Infectious diseases. Bacterial meningitis finds new niche in gay communities" Science 341(6144): 328

Ladhani, S. N., K. Beebeejaun, J. Lucidarme, H. Campbell, S. Gray, E. Kaczmarski, M. E. Ramsay and R. Borrow (2015) "Increase in endemic Neisseria meningitidis capsular group W sequence type 11 complex associated with severe invasive disease in England and Wales" Clin Infect Dis 60(4): 578-585

Ladhani, S. N., J. S. Flood, M. E. Ramsay, H. Campbell, S. J. Gray, E. B. Kaczmarski, R. H. Mallard, M. Guiver, et al. (2012a) "Invasive meningococcal disease in England and Wales: Implications for the introduction of new vaccines" Vaccine 30(24): 3710-3716

Ladhani, S. N., J. Lucidarme, L. S. Newbold, S. J. Gray, A. D. Carr, J. Findlow, M. E. Ramsay, E. B. Kaczmarski, et al. (2012b) "Invasive meningococcal capsular group Y disease, England and Wales, 2007-2009" Emerg Infect Dis 18(1): 63-70

Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg (2009) "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome" Genome Biol 10(3)

Lapeyssonnie, L. (1963) "[Cerebrospinal Meningitis in Africa]" Bull World Health Organ 28 Suppl: 1-114

Lappann, M., H. Claus, T. van Alen, M. Harmsen, J. Elias, S. Molin and U. Vogel (2010) "A dual role of extracellular DNA during biofilm formation of *Neisseria meningitidis*" Mol Microbiol 75(6): 1355-1371

Law, D. K., M. Lorange, L. Ringuette, R. Dion, M. Giguere, A. M. Henderson, J. Stoltz, W. D. Zollinger, et al. (2006) "Invasive meningococcal disease in Quebec, Canada, due to an emerging clone of ST-269 serogroup B meningococci with serotype antigen 17 and serosubtype antigen P1.19 (B:17:P1.19)" J Clin Microbiol 44(8): 2743-2749

Law, D. K., J. Zhou, S. Deng, L. Hoang, G. Tyrrell, G. Horsman, J. Wylie and R. S. Tsang (2014) "Determination of serotyping antigens, clonal analysis and genetic characterization of the 4CMenB vaccine antigen genes in invasive *Neisseria meningitidis* from Western Canada, 2009 to 2013" J Med Microbiol 63(Pt 11): 1490-1499

Le Saux, N., F. Ashton, M. Rahman, A. Ryan, E. Ellis, S. Tamblyn, J. Morris, A. Borczyk, et al. (1992) "Carriage of *Neisseria* species in communities with different rates of meningococcal disease" Can J Infect Dis 3(2): 60-64

Levin, B. R. (1996) "The evolution and maintenance of virulence in microparasites" Emerg Infect Dis 2(2): 93-102

Levin, B. R. and J. J. Bull (1994) "Short-sighted evolution and the virulence of pathogenic microorganisms" Trends Microbiol 2(3): 76-81

Lewis, L. A., E. Gray, Y. P. Wang, B. A. Roe and D. W. Dyer (1997) "Molecular characterization of hpuAB, the haemoglobin-haptoglobin-utilization operon of *Neisseria meningitidis*" Mol Microbiol 23(4): 737-749

Lipsitch, M. and E. R. Moxon (1997) "Virulence and transmissibility of pathogens: what is the relationship?" Trends Microbiol 5(1): 31-37

Lluch-Senar, M., K. Luong, V. Llorens-Rico, J. Delgado, G. Fang, K. Spittle, T. A. Clark, E. Schadt, et al. (2013) "Comprehensive Methylome Characterization of Mycoplasma genitalium and Mycoplasma pneumoniae at Single-Base Resolution" PLoS Genet 9(1): e1003191

Loh, E., E. Kugelberg, A. Tracy, Q. Zhang, B. Gollan, H. Ewles, R. Chalmers, V. Pelicic, et al. (2013) "Temperature triggers immune evasion by Neisseria meningitidis" Nature 502(7470): 237-240

Low, D. A. and J. Casadesus (2008) "Clocks and switches: bacterial gene regulation by DNA adenine methylation" Curr Opin Microbiol 11(2): 106-112

Lucidarme, J., M. Comanducci, J. Findlow, S. J. Gray, E. B. Kaczmarski, M. Guiver, E. Kugelberg, P. J. Vallely, et al. (2009) "Characterization of fHbp, nhba (gna2132), nadA, porA, sequence type (ST), and genomic presence of IS1301 in group B meningococcal ST269 clonal complex isolates from England and Wales" J Clin Microbiol 47(11): 3577-3585

Lucidarme, J., M. Comanducci, J. Findlow, S. J. Gray, E. B. Kaczmarski, M. Guiver, P. J. Vallely, P. Oster, et al. (2010) "Characterization of fHbp, nhba (gna2132), nadA, porA, and Sequence Type in Group B Meningococcal Case Isolates Collected in England and Wales during January 2008 and Potential Coverage of an Investigational Group B Meningococcal Vaccine" Clin Vaccine Immunol 17(6): 919-929

Lucidarme, J., J. Findlow, H. Chan, I. M. Feavers, S. J. Gray, E. B. Kaczmarski, J. Parkhill, X. Bai, et al. (2013a) "The distribution and phase variation status of hemoglobin receptors in invasive meningococcal serogroup B disease: genotypic and phenotypic analysis" PLoS One 8(9): e76932

Lucidarme, J., S. Gilchrist, L. S. Newbold, S. J. Gray, E. B. Kaczmarski, L. Richardson, J. S. Bennett, M. C. Maiden, et al. (2013b) "Genetic distribution of noncapsular meningococcal group B vaccine antigens in *Neisseria lactamica*" Clin Vaccine Immunol 20(9): 1360-1369

Lucidarme, J., D. M. Hill, H. B. Bratcher, S. J. Gray, M. du Plessis, R. S. Tsang, J. A. Vazquez, M. K. Taha, et al. (2015) "Genomic resolution of an aggressive, widespread, diverse and expanding meningococcal serogroup B, C and W lineage" J Infect 71(5): 544-552

MacLennan, J., G. Kafatos, K. Neal, N. Andrews, J. C. Cameron, R. Roberts, M. R. Evans, K. Cann, et al. (2006) "Social behavior and meningococcal carriage in British teenagers" Emerg Infect Dis 12(6): 950-957

Macneil, J. R., A. C. Cohn, E. R. Zell, S. Schmink, E. Miller, T. Clark, N. E. Messonnier, T. Active Bacterial Core surveillance, et al. (2011) "Early estimate of the effectiveness of quadrivalent meningococcal conjugate vaccine" Pediatr Infect Dis J 30(6): 451-455

Madico, G., J. A. Welsch, L. A. Lewis, A. McNaughton, D. H. Perlman, C. E. Costello, J. Ngampasutadol, U. Vogel, et al. (2006) "The meningococcal vaccine candidate GNA1870 binds the complement regulatory protein factor H and enhances serum resistance" J Immunol 177(1): 501-510

Maiden, M. C. (2006) "Multilocus sequence typing of bacteria" Annu Rev Microbiol 60: 561-588

Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, et al. (1998) "Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms" Proc Natl Acad Sci U S A 95(6): 3140-3145

Maiden, M. C., A. B. Ibarz-Pavon, R. Urwin, S. J. Gray, N. J. Andrews, S. C. Clarke, A. M. Walker, M. R. Evans, et al. (2008) "Impact of meningococcal serogroup C conjugate vaccines on carriage and herd immunity" J Infect Dis 197(5): 737-743

Maiden, M. C. and B. G. Spratt (1999) "Meningococcal conjugate vaccines: new opportunities and new challenges" Lancet 354(9179): 615-616

Maiden, M. C., J. M. Stuart and U. K. M. C. Group (2002) "Carriage of serogroup C meningococci 1 year after meningococcal C conjugate polysaccharide vaccination" Lancet 359(9320): 1829-1831

Maiden, M. C., M. J. van Rensburg, J. E. Bray, S. G. Earle, S. A. Ford, K. A. Jolley and N. D. McCarthy (2013) "MLST revisited: the gene-by-gene approach to bacterial genomics" Nat Rev Microbiol 11(10): 728-736

Mairey, E., A. Genovesio, E. Donnadieu, C. Bernard, F. Jaubert, E. Pinard, J. Seylaz, J. C. Olivo-Marin, et al. (2006) "Cerebral microcirculation shear stress levels determine *Neisseria meningitidis* attachment sites along the blood-brain barrier" J Exp Med 203(8): 1939-1950

Makela, P. H., H. Peltola, H. Kayhty, H. Jousimies, O. Pettay, E. Ruoslahti, A. Sivonen and O. V. Renkonen (1977) "Polysaccharide vaccines of group A *Neisseria meningtitidis* and *Haemophilus influenzae* type b: a field trial in Finland" J Infect Dis 136 Suppl: S43-50

Mandal, S., H. M. Wu, J. R. MacNeil, K. Machesky, J. Garcia, B. D. Plikaytis, K. Quinn, L. King, et al. (2013) "Prolonged university outbreak of meningococcal disease associated with a serogroup B strain rarely seen in the United States" Clin Infect Dis 57(3): 344-348

Mandrell, R. E., J. M. Griffiss and B. A. Macher (1988) "Lipooligosaccharides (LOS) of *Neisseria gonorrhoeae* and *Neisseria meningitidis* have components that are immunochemically

similar to precursors of human blood group antigens. Carbohydrate sequence specificity of the mouse monoclonal antibodies that recognize crossreacting antigens on LOS and human erythrocytes" J Exp Med 168(1): 107-126

Marc LaForce, F., N. Ravenscroft, M. Djingarey and S. Viviani (2009) "Epidemic meningitis due to Group A *Neisseria meningitidis* in the African meningitis belt: a persistent problem with an imminent solution" Vaccine 27 Suppl 2: B13-19

Marchler-Bauer, A., M. K. Derbyshire, N. R. Gonzales, S. Lu, F. Chitsaz, L. Y. Geer, R. C. Geer, J. He, et al. (2015) "CDD: NCBI's conserved domain database" Nucleic Acids Res 43(Database issue): D222-226

Marri, P. R., M. Paniscus, N. J. Weyand, M. A. Rendon, C. M. Calton, D. R. Hernandez, D. L. Higashi, E. Sodergren, et al. (2010) "Genome sequencing reveals widespread virulence gene exchange among human *Neisseria* species" PLoS One 5(7): e11835

Marshall, H. S., P. C. Richmond, M. D. Nissen, A. Wouters, J. Baber, Q. Jiang, A. S. Anderson, T. R. Jones, et al. (2013) "A phase 2 open-label safety and immunogenicity study of a meningococcal B bivalent rLP2086 vaccine in healthy adults" Vaccine 31(12): 1569-1575

Martin, D. R., N. Ruijne, L. McCallum, J. O'Hallahan and P. Oster (2006) "The VR2 epitope on the PorA P1.7-2,4 protein is the major target for the immune response elicited by the strain-specific group B meningococcal vaccine MeNZB" Clin Vaccine Immunol 13(4): 486-491

Martin, N. G. and M. D. Snape (2013) "A multicomponent serogroup B meningococcal vaccine is licensed for use in Europe: what do we know, and what are we yet to learn?" Expert Rev Vaccines 12(8): 837-858

Martinon-Torres, F., F. Gimenez-Sanchez, E. Bernaola-Iturbe, J. Diez-Domingo, Q. Jiang and J. L. Perez (2014) "A randomized, phase 1/2 trial of the safety, tolerability, and immunogenicity of bivalent rLP2086 meningococcal B vaccine in healthy infants" Vaccine 32(40): 5206-5211

Masignani, V., M. Comanducci, M. M. Giuliani, S. Bambini, J. Adu-Bobie, B. Arico, B. Brunelli, A. Pieri, et al. (2003) "Vaccination against *Neisseria meningitidis* using three variants of the lipoprotein GNA1870" J Exp Med 197(6): 789-799

Mayer, L. W., M. W. Reeves, N. Al-Hamdan, C. T. Sacchi, M. K. Taha, G. W. Ajello, S. E. Schmink, C. A. Noble, et al. (2002) "Outbreak of W135 meningococcal disease in 2000: Not emergence of a new W135 strain but clonal expansion within the electrophoretic type-37 complex" J Infect Dis 185(11): 1596-1605

McGuinness, B., A. K. Barlow, I. N. Clarke, J. E. Farley, A. Anilionis, J. T. Poolman and J. E. Heckels (1990) "Deduced amino acid sequences of class 1 protein (PorA) from three strains of *Neisseria meningitidis*. Synthetic peptides define the epitopes responsible for serosubtype specificity" J Exp Med 171(6): 1871-1882

McIntosh, E. D., V. Carey, D. Toneatto, P. Dull and J. Wassil (2015) "Prevention of rare diseases: how revolutionary techniques can help vulnerable individuals-the example of serogroup B meningococcal infection" Ther Adv Vaccines 3(1): 13-23

McQuaid, F. and M. D. Snape (2014) "Will booster doses be required for serogroup B meningococcal vaccine?" Expert Rev Vaccines 13(3): 313-315

Medini, D., D. Serruto, J. Parkhill, D. A. Relman, C. Donati, R. Moxon, S. Falkow and R. Rappuoli (2008) "Microbiology in the post-genomic era" Nat Rev Microbiol 6(6): 419-430

Melican, K., P. Michea Veloso, T. Martin, P. Bruneval and G. Dumenil (2013) "Adhesion of *Neisseria meningitidis* to dermal vessels leads to local vascular damage and purpura in a humanized mouse model" PLoS Pathog 9(1): e1003139

MenAfriCar, c. (2015) "The diversity of meningococcal carriage across the African Meningitis Belt and the impact of vaccination with a group A meningococcal conjugate vaccine" J Infect Dis 212(8): 1298-1307

Merz, A. J. and M. So (2000) "Interactions of pathogenic neisseriae with epithelial cell membranes" Annu Rev Cell Dev Biol 16: 423-457

Meyers, L. A., B. R. Levin, A. R. Richardson and I. Stojiljkovic (2003) "Epidemiology, hypermutation, within-host evolution and the virulence of *Neisseria meningitidis*" Proc Biol Sci 270(1525): 1667-1677

Michaelsen, T. E., A. Aase, J. Kolberg, E. Wedge and E. Rosenqvist (2001) "PorB3 outer membrane protein on *Neisseria meningitidis* is poorly accessible for antibody binding on live bacteria" Vaccine 19(11-12): 1526-1533

Millar, J. W., E. E. Siess, H. A. Feldman, C. Silverman and P. Frank (1963) "In vivo and in vitro resistance to sulfadiazine in strains of *Neisseria meningitidis*" JAMA 186: 139-141

Miller, E., D. Salisbury and M. Ramsay (2001) "Planning, registration, and implementation of an immunisation campaign against meningococcal serogroup C disease in the UK: a success story" Vaccine 20 Suppl 1: S58-67

Milne, I., M. Bayer, L. Cardle, P. Shaw, G. Stephen, F. Wright and D. Marshall (2010) "Tablet-next generation sequence assembly visualization" Bioinformatics 26(3): 401-402

Montano, V., X. Didelot, M. Foll, B. Linz, R. Reinhardt, S. Suerbaum, Y. Moodley and J. D. Jensen (2015) "Worldwide population structure, long-term demography, and local adaptation of *Helicobacter pylori*" Genetics 200(3): 947-963

Moore, P. S., M. W. Reeves, B. Schwartz, B. G. Gellin and C. V. Broome (1989) "Intercontinental spread of an epidemic group A *Neisseria meningitidis* strain" Lancet 2(8657): 260-263

Morand, P. E. and T. Rudel (2006). Genetics, structure and function of pili. In Handbook of Meningococcal Disease. M. F. M. C. Maiden. Weinheim, Germany, Wiley-VCH**:** 235-254.

Morelli, G., B. Malorny, K. Muller, A. Seiler, J. F. Wang, J. del Valle and M. Achtman (1997) "Clonal descent and microevolution of *Neisseria meningitidis* during 30 years of epidemic spread" Mol Microbiol 25(6): 1047-1064

Morrison, D. A. (2011). Introduction to phylogenetic networks, RJR Productions.

Moxon, E. R. and V. A. Jansen (2005) "Phage variation: understanding the behaviour of an accidental pathogen" Trends Microbiol 13(12): 563-565

Moxon, R., C. Bayliss and D. Hood (2006) "Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation" Annu Rev Genet 40: 307-333

Mulks, M. H., A. G. Plaut, H. A. Feldman and B. Frangione (1980) "IgA proteases of two distinct specificities are released by *Neisseria meningitidis*" J Exp Med 152(5): 1442-1447

Murphy, J., J. Klumpp, J. Mahony, M. O'Connell-Motherway, A. Nauta and D. van Sinderen (2014) "Methyltransferases acquired by lactococcal 936-type phage provide protection against restriction endonuclease activity" BMC Genomics 15: 831

Musilek, M., D. Giorgini, N. Hamadouche, P. Kriz and M. K. Taha (1998) "Genetic heterogeneity of strains of *Neisseria meningitidis* belonging to serotype 22 isolated in the Czech Republic" J Clin Microbiol 36(2): 563-565

Naito, T., K. Kusano and I. Kobayashi (1995) "Selfish behavior of restriction-modification systems" Science 267(5199): 897-899

Nassif, X., S. Bourdoulous, E. Eugene and P. O. Couraud (2002) "How do extracellular pathogens cross the blood-brain barrier?" Trends Microbiol 10(5): 227-232

Nassif, X., C. Pujol, P. Morand and E. Eugene (1999) "Interactions of pathogenic *Neisseria* with host cells. Is it possible to assemble the puzzle?" Mol Microbiol 32(6): 1124-1132

Neal, K. R., J. S. Nguyen-Van-Tam, N. Jeffrey, R. C. Slack, R. J. Madeley, K. Ait-Tahar, K. Job, M. C. Wale, et al. (2000) "Changing carriage rate of *Neisseria meningitidis* among university students during the first week of term: cross sectional study" BMJ 320(7238): 846-849

Nissen, M. D., H. S. Marshall, P. C. Richmond, Q. Jiang, S. L. Harris, T. R. Jones, K. U. Jansen and J. L. Perez (2013) "A randomized, controlled, phase 1/2 trial of a *Neisseria meningitidis* serogroup B bivalent rLP2086 vaccine in healthy children and adolescents" Pediatr Infect Dis J 32(4): 364-371

Noinaj, N., S. K. Buchanan and C. N. Cornelissen (2012) "The transferrin-iron import system from pathogenic *Neisseria* species" Mol Microbiol 86(2): 246-257

North, E. (1980) "Classics in infectious diseases. Concerning the epidemic of spotted fever in New England: Elisha North, 1811" Rev Infect Dis 2(5): 811-816

O'Connor, B. D., B. Merriman and S. F. Nelson (2010) "SeqWare Query Engine: storing and searching sequence data in the cloud" BMC Bioinformatics 11 Suppl 12: S2

O'Hallahan, J., D. Lennon and P. Oster (2004) "The strategy to control New Zealand's epidemic of group B meningococcal disease" Pediatr Infect Dis J 23(12 Suppl): S293-298

Olcen, P., J. Kjellander, D. Danielsson and B. L. Lindquist (1981) "Epidemiology of *Neisseria meningitidis*; prevalence and symptoms from the upper respiratory tract in family members to patients with meningococcal disease" Scand J Infect Dis 13(2): 105-109

Oldfield, N. J., S. J. Bland, M. Taraktsoglou, F. J. Dos Ramos, K. Robinson, K. G. Wooldridge and D. A. Ala'Aldeen (2007) "T-cell stimulating protein A (TspA) of *Neisseria meningitidis* is required for optimal adhesion to human cells" Cell Microbiol 9(2): 463-478

Oldfield, N. J., S. Matar, F. A. Bidmos, M. Alamro, K. R. Neal, D. P. Turner, C. D. Bayliss and D. A. Ala'aldeen (2013) "Prevalence and Phase Variable Expression Status of Two Autotransporters, NalP and MspA, in Carriage and Disease Isolates of *Neisseria meningitidis*" PLoS One 8(7): e69746

Olyhoek, T., B. A. Crowe and M. Achtman (1987) "Clonal population structure of *Neisseria meningitidis* serogroup A isolated from epidemics and pandemics between 1915 and 1983" Rev Infect Dis 9(4): 665-692

Omer, H., G. Rose, K. A. Jolley, E. Frapy, J. R. Zahar, M. C. Maiden, S. D. Bentley, C. R. Tinsley, et al. (2011) "Genotypic and phenotypic modifications of *Neisseria meningitidis* after an accidental human passage" PLoS One 6(2): e17145

Oviedo-Orta, E., S. Ahmed, R. Rappuoli and S. Black (2015) "Prevention and control of meningococcal outbreaks: The emerging role of serogroup B meningococcal vaccines" Vaccine

Pace, D. and A. J. Pollard (2007) "Meningococcal A, C, Y and W-135 polysaccharide-protein conjugate vaccines" Arch Dis Child 92(10): 909-915

Pace, D. and A. J. Pollard (2012) "Meningococcal disease: clinical presentation and sequelae" Vaccine 30 Suppl 2: B3-9

Pagliarulo, C., P. Salvatore, L. R. De Vitis, R. Colicchio, C. Monaco, M. Tredici, A. Tala, M. Bardaro, et al. (2004) "Regulation and differential expression of gdhA encoding NADP-specific glutamate dehydrogenase in *Neisseria meningitidis* clinical isolates" Mol Microbiol 51(6): 1757-1772

Palmer, M. E., M. Lipsitch, E. R. Moxon and C. D. Bayliss (2013) "Broad conditions favor the evolution of phase-variable Loci" MBio 4(1)

Paradis, E., J. Claude and K. Strimmer (2004) "APE: Analyses of Phylogenetics and Evolution in R language" Bioinformatics 20(2): 289-290

Parkhill, J., M. Achtman, K. D. James, S. D. Bentley, C. Churcher, S. R. Klee, G. Morelli, D. Basham, et al. (2000) "Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491" Nature 404(6777): 502-506

Parkhill, J. and B. W. Wren (2011) "Bacterial epidemiology and biology--lessons from genome sequencing" Genome Biol 12(10): 230

Peak, I. R., Y. N. Srikhanta, V. E. Weynants, C. Feron, J. T. Poolman and M. P. Jennings (2013) "Evaluation of truncated NhhA protein as a candidate meningococcal vaccine antigen" PLoS One 8(9): e72003

Peng, J., L. Yang, F. Yang, J. Yang, Y. Yan, H. Nie, X. Zhang, Z. Xiong, et al. (2008) "Characterization of ST-4821 complex, a unique *Neisseria meningitidis* clone" Genomics 91(1): 78-87

Perez-Trallero, E., O. Esnal and J. M. Marimon (2014) "Progressive decrease in the potential usefulness of meningococcal serogroup B vaccine (4CMenB, Bexsero(R)) in Gipuzkoa, Northern Spain" PLoS One 9(12): e116024

Perkins-Balding, D., M. Ratliff-Griffin and I. Stojiljkovic (2004) "Iron transport systems in *Neisseria meningitidis*" Microbiol Mol Biol Rev 68(1): 154-171

Pettersson, A., A. Maas, D. van Wassenaar, P. van der Ley and J. Tommassen (1995) "Molecular characterization of FrpB, the 70-kilodalton iron-regulated outer membrane protein of *Neisseria meningitidis*" Infect Immun 63(10): 4181-4184

PHE (2014) "Meningococcal disease: laboratory confirmed cases in England and Wales." Meningococcal disease: guidance, data and analysis Retrieved 28/01/14, 2014, from https://www.gov.uk/government/publications/meningococcal-disease-laboratory-confirmed-cases-in-england-and-wales.

Piet, J. R., A. Zariri, F. Fransen, K. Schipper, P. van der Ley, D. van de Beek and A. van der Ende (2014) "Meningitis caused by a lipopolysaccharide deficient *Neisseria meningitidis*" J Infect 69(4): 352-357

Pinner, R. W., B. G. Gellin, W. F. Bibb, C. N. Baker, R. Weaver, S. B. Hunter, S. H. Waterman, L. F. Mocca, et al. (1991) "Meningococcal disease in the United States--1986. Meningococcal Disease Study Group" J Infect Dis 164(2): 368-374

Pizza, M., V. Scarlato, V. Masignani, M. M. Giuliani, B. Arico, M. Comanducci, G. T. Jennings, L. Baldi, et al. (2000) "Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing" Science 287(5459): 1816-1820

Pollard, A. J., C. Green, M. Sadarangani and M. D. Snape (2013) "Adolescents need a booster of serogroup C meningococcal vaccine to protect them and maintain population control of the disease" Arch Dis Child 98(4): 248-251

Price, J. R., T. Golubchik, K. Cole, D. J. Wilson, D. W. Crook, G. E. Thwaites, R. Bowden, A. S. Walker, et al. (2014) "Whole-genome sequencing shows that patient-to-patient transmission rarely accounts for acquisition of *Staphylococcus aureus* in an intensive care unit" Clin Infect Dis 58(5): 609-618

Prinz, T., M. Meyer, A. Pettersson and J. Tommassen (1999) "Structural characterization of the lactoferrin receptor from *Neisseria meningitidis*" J Bacteriol 181(14): 4417-4419

Proft, T. and E. N. Baker (2009) "Pili in Gram-negative and Gram-positive bacteria - structure, assembly and their role in disease" Cell Mol Life Sci 66(4): 613-635

Puleston, R., C. Beck, M. Tahir, M. Bardhan, P. Charlemagne, C. Alves, S. Ladhani, C. Watson, et al. (2012) "An unusual transmission event of *Neisseria meningitidis* serogroup W135 type 2a in a healthcare setting, England, 2012" Euro Surveill 17(44)

Quail, M. A., M. Smith, P. Coupland, T. D. Otto, S. R. Harris, T. R. Connor, A. Bertoni, H. P. Swerdlow, et al. (2012) "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers" BMC Genomics 13: 341

Qvarnstrom, Y. and G. Swedberg (2000) "Additive effects of a two-amino-acid insertion and a single-amino-acid substitution in dihydropteroate synthase for the development of sulphonamide-resistant *Neisseria meningitidis*" Microbiology-Uk 146: 1151-1156

R Development Core Team (2012). "R: A language and environment for statistical computing". "R Foundation for Statistical Computing, Vienna, Austria".

Radstrom, P., C. Fermer, B. E. Kristiansen, A. Jenkins, O. Skold and G. Swedberg (1992) "Transformational exchanges in the dihydropteroate synthase gene of *Neisseria meningitidis:* a novel mechanism for acquisition of sulfonamide resistance" J Bacteriol 174(20): 6386-6393

Ramsay, M. E., N. J. Andrews, C. L. Trotter, E. B. Kaczmarski and E. Miller (2003) "Herd immunity from meningococcal serogroup C conjugate vaccination in England: database analysis" BMJ 326(7385): 365-366

Rappuoli, R. (2007) "Bridging the knowledge gaps in vaccine design" Nat Biotechnol 25(12): 1361-1366

Read, R. C., D. Baxter, D. R. Chadwick, S. N. Faust, A. Finn, S. B. Gordon, P. T. Heath, D. J. Lewis, et al. (2014) "Effect of a quadrivalent meningococcal ACWY glycoconjugate or a serogroup B meningococcal vaccine on meningococcal carriage: an observer-blind, phase 3 randomised clinical trial" Lancet 384(9960): 2123-2131

Reingold, A. L., C. V. Broome, A. W. Hightower, G. W. Ajello, G. A. Bolan, C. Adamsbaum, E. E. Jones, C. Phillips, et al. (1985) "Age-specific differences in duration of clinical protection after vaccination with meningococcal polysaccharide A vaccine" Lancet 2(8447): 114-118

Reuter, S., M. J. Ellington, E. J. Cartwright, C. U. Koser, M. E. Torok, T. Gouliouris, S. R. Harris, N. M. Brown, et al. (2013) "Rapid bacterial whole-genome sequencing to enhance diagnostic and public health microbiology" JAMA Intern Med 173(15): 1397-1404

Richardson, E. J. and M. Watson (2013) "The automatic annotation of bacterial genomes" Brief Bioinform 14(1): 1-12

Ringuette, L., M. Lorange, A. Ryan and F. Ashton (1995) "Meningococcal infections in the Province of Quebec, Canada, during the period 1991 to 1992" J Clin Microbiol 33(1): 53-57

Rishishwar, L., L. S. Katz, N. V. Sharma, L. Rowe, M. Frace, J. Dolan Thomas, B. H. Harcourt, L. W. Mayer, et al. (2012) "Genomic basis of a polyagglutinating isolate of *Neisseria meningitidis*" J Bacteriol 194(20): 5649-5656

Rissman, A. I., B. Mau, B. S. Biehl, A. E. Darling, J. D. Glasner and N. T. Perna (2009) "Reordering contigs of draft genomes using the Mauve Aligner" Bioinformatics 25(16): 2071-2073

Robbins, J. B., R. Schneerson, S. C. Szu, A. Fattom, Y. Yang, T. Lagergard, C. Chu and U. S. Sorensen (1989) "Prevention of invasive bacterial diseases by immunization with polysaccharide-protein conjugates" Curr Top Microbiol Immunol 146: 169-180

Roberts, R. J., T. Vincze, J. Posfai and D. Macelis (2015) "REBASE--a database for DNA restriction and modification: enzymes, genes and genomes" Nucleic Acids Res 43(Database issue): D298-299

Robinson, K., K. G. Wooldridge, D. B. Wells, A. Hasan, I. Todd, A. Robins, R. James and D. A. Ala'Aldeen (2005) "T-cell-stimulating protein A elicits immune responses during meningococcal carriage and human disease" Infect Immun 73(8): 4684-4693

Rosenstein, N. E., B. A. Perkins, D. S. Stephens, L. Lefkowitz, M. L. Cartter, R. Danila, P. Cieslak, K. A. Shutt, et al. (1999) "The changing epidemiology of meningococcal disease in the United States, 1992-1996" J Infect Dis 180(6): 1894-1901

Rosenstein, N. E., B. A. Perkins, D. S. Stephens, T. Popovic and J. M. Hughes (2001) "Meningococcal disease" N Engl J Med 344(18): 1378-1388

Rothman, K. J. (2002). Epidemiology: An Introduction, Oxford University Press.

Rothman, K. J., Greenland, S., Lash, T.L. (2008). Modern Epidemiology, Lippincott Williams & Wilkins.

Rouaud, P., A. Perrocheau, M. K. Taha, C. Sesboue, A. M. Forgues, I. Parent du Chatelet and D. Levy-Bruhl (2006) "Prolonged outbreak of B meningococcal disease in the Seine-Maritime department, France, January 2003 to June 2005" Euro Surveill 11(7): 178-181

Round, A., M. R. Evans, R. L. Salmon, I. K. Hosein, A. K. Mukerjee, R. W. Smith and S. R. Palmer (2001) "Public health management of an outbreak of group C meningococcal disease in university campus residents" Eur J Public Health 11(4): 431-436

Rowe, H. A., N. J. Griffiths, D. J. Hill and M. Virji (2007) "Co-ordinate action of bacterial adhesins and human carcinoembryonic antigen receptors in enhanced cellular invasion by capsulate serum resistant *Neisseria meningitidis*" Cell Microbiol 9(1): 154-168

Roznovsky, L., J. Matuska and P. Kriz-Kuzemenska (1994) "Unusual incidence of serious meningococcal diseases in some north Moravian districts in the first half of 1993" Prakt lek 74: 101-104

Rusniok, C., D. Vallenet, S. Floquet, H. Ewles, C. Mouze-Soulama, D. Brown, A. Lajus, C. Buchrieser, et al. (2009) "NeMeSys: a biological resource for narrowing the gap between sequence and function in the human pathogen *Neisseria meningitidis*" Genome Biol 10(10): R110

Russell, J. E., K. A. Jolley, I. M. Feavers, M. C. Maiden and J. Suker (2004) "PorA variable regions of *Neisseria meningitidis*" Emerg Infect Dis 10(4): 674-678

Russell, J. E., R. Urwin, S. J. Gray, A. J. Fox, I. M. Feavers and M. C. Maiden (2008) "Molecular epidemiology of meningococcal disease in England and Wales 1975-1995, before the introduction of serogroup C conjugate vaccines" Microbiology 154(Pt 4): 1170-1177

Sacchi, C. T., A. P. Lemos, A. M. Whitney, C. A. Solari, M. E. Brandt, C. E. Melles, C. E. Frasch and L. W. Mayer (1998) "Correlation between serological and sequencing analyses of the PorB outer membrane protein in the *Neisseria meningitidis* serotyping system" Clin Diagn Lab Immunol 5(3): 348-354

Safadi, M. A., S. Gonzalez-Ayala, A. Jakel, H. Wieffer, C. Moreno and A. Vyse (2013) "The epidemiology of meningococcal disease in Latin America 1945-2010: an unpredictable and changing landscape" Epidemiol Infect 141(3): 447-458

Saleem, M., S. M. Prince, S. E. Rigby, M. Imran, H. Patel, H. Chan, H. Sanders, M. C. Maiden, et al. (2013) "Use of a molecular decoy to segregate transport from antigenicity in the FrpB iron transporter from *Neisseria meningitidis*" PLoS One 8(2): e56746

Salzberg, S. L., A. L. Delcher, S. Kasif and O. White (1998) "Microbial gene identification using interpolated Markov models" Nucleic Acids Res 26(2): 544-548

Salzberg, S. L., A. M. Phillippy, A. Zimin, D. Puiu, T. Magoc, S. Koren, T. J. Treangen, M. C. Schatz, et al. (2012) "GAGE: A critical evaluation of genome assemblies and assembly algorithms" Genome Res 22(3): 557-567

Santolaya, M. E., M. O'Ryan, M. T. Valenzuela, V. Prado, R. F. Vergara, A. Munoz, D. Toneatto, G. Grana, et al. (2013) "Persistence of antibodies in adolescents 18-24 months after immunization with one, two, or three doses of 4CMenB meningococcal serogroup B vaccine" Hum Vaccin Immunother 9(11): 2304-2310

Sarafian, S. K. and J. S. Knapp (1989) "Molecular epidemiology of gonorrhea" Clin Microbiol Rev 2 Suppl: S49-55

Saunders, N. J., J. F. Peden, D. W. Hood and E. R. Moxon (1998) "Simple sequence repeats in the Helicobacter pylori genome" Mol Microbiol 27(6): 1091-1098

Scarselli, M., D. Serruto, P. Montanari, B. Capecchi, J. Adu-Bobie, D. Veggi, R. Rappuoli, M. Pizza, et al. (2006) "*Neisseria meningitidis* NhhA is a multifunctional trimeric autotransporter adhesin" Mol Microbiol 61(3): 631-644

Schaaf, L., E. M. Baur, K. Kusterer, K. Huc, H. Bode, G. Rohr and K. H. Usadel (1989) "Meningococcemia in a group of Canadian students on a European trip" Can Dis Wkly Rep 15(46): 229-231

Schadt, E. E., O. Banerjee, G. Fang, Z. Feng, W. H. Wong, X. Zhang, A. Kislyuk, T. A. Clark, et al. (2013) "Modeling kinetic rate variation in third generation DNA sequencing data to detect putative modifications to DNA bases" Genome Res 23(1): 129-141

Schmitt, C., D. Turner, M. Boesl, M. Abele, M. Frosch and O. Kurzai (2007) "A functional two-partner secretion system contributes to adhesion of *Neisseria meningitidis* to epithelial cells" J Bacteriol 189(22): 7968-7976

Schoen, C., J. Blom, H. Claus, A. Schramm-Gluck, P. Brandt, T. Muller, A. Goesmann, B. Joseph, et al. (2008) "Whole-genome comparison of disease and carriage strains provides insights into virulence evolution in *Neisseria meningitidis*" Proc Natl Acad Sci U S A 105(9): 3473-3478

Schoen, C., L. Kischkies, J. Elias and B. J. Ampattu (2014) "Metabolism and virulence in *Neisseria meningitidis*" Front Cell Infect Microbiol 4: 114

Schoen, C., H. Tettelin, J. Parkhill and M. Frosch (2009) "Genome flexibility in *Neisseria meningitidis*" Vaccine 27: B103-B111

Scholten, R. J., B. Kuipers, H. A. Valkenburg, J. Dankert, W. D. Zollinger and J. T. Poolman (1994a) "Lipo-oligosaccharide immunotyping of *Neisseria meningitidis* by a whole-cell ELISA with monoclonal antibodies" J Med Microbiol 41(4): 236-243

Scholten, R. J., J. T. Poolman, H. A. Valkenburg, H. A. Bijlmer, J. Dankert and D. A. Caugant (1994b) "Phenotypic and genotypic changes in a new clone complex of *Neisseria meningitidis* causing disease in The Netherlands, 1958-1990" J Infect Dis 169(3): 673-676

Seemann, T. (2014) "Prokka: rapid prokaryotic genome annotation" Bioinformatics

Segal, E., P. Hagblom, H. S. Seifert and M. So (1986) "Antigenic variation of gonococcal pilus involves assembly of separated silent gene segments" Proc Natl Acad Sci U S A 83(7): 2177-2181

Seib, K. L., F. E. Jen, A. Tan, A. L. Scott, R. Kumar, P. M. Power, L. T. Chen, H. J. Wu, et al. (2015a) "Specificity of the ModA11, ModA12 and ModD1 epigenetic regulator N6-adenine DNA methyltransferases of *Neisseria meningitidis*" Nucleic Acids Res 43(8): 4150-4162

Seib, K. L., E. Pigozzi, A. Muzzi, J. A. Gawthorne, I. Delany, M. P. Jennings and R. Rappuoli (2011) "A novel epigenetic regulator associated with the hypervirulent *Neisseria meningitidis* clonal complex 41/44" FASEB J 25(10): 3622-3633

Seib, K. L., M. Scarselli, M. Comanducci, D. Toneatto and V. Masignani (2015b) "*Neisseria meningitidis* factor H-binding protein fHbp: a key virulence factor and vaccine antigen" Expert Rev Vaccines 14(6): 841-859

Seib, K. L., H. J. Tseng, A. G. McEwan, M. A. Apicella and M. P. Jennings (2004) "Defenses against oxidative stress in *Neisseria gonorrhoeae* and *Neisseria meningitidis*: distinctive systems for different lifestyles" J Infect Dis 190(1): 136-147

Seib, K. L., X. Zhao and R. Rappuoli (2012) "Developing vaccines in the era of genomics: a decade of reverse vaccinology" Clin Microbiol Infect 18 Suppl 5: 109-116

Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour and T. S. Whittam (1986) "Methods of multilocus enzyme electrophoresis for bacterial population-genetics and systematics" Appl Environ Microbiol 51(5): 873-884

Serruto, D., J. Adu-Bobie, M. Scarselli, D. Veggi, M. Pizza, R. Rappuoli and B. Arico (2003) "*Neisseria meningitidis* App, a new adhesin with autocatalytic serine protease activity" Mol Microbiol 48(2): 323-334

Serruto, D., M. J. Bottomley, S. Ram, M. M. Giuliani and R. Rappuoli (2012) "The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: immunological, functional and structural characterization of the antigens" Vaccine 30 Suppl 2: B87-97

Serruto, D., T. Spadafina, L. Ciucchi, L. A. Lewis, S. Ram, M. Tontini, L. Santini, A. Biolchi, et al. (2010) "*Neisseria meningitidis* GNA2132, a heparin-binding protein that induces protective immunity in humans" Proc Natl Acad Sci U S A 107(8): 3770-3775

Severiano, A., F. R. Pinto, M. Ramirez and J. A. Carrico (2011) "Adjusted Wallace coefficient as a measure of congruence between typing methods" J Clin Microbiol 49(11): 3997-4000

Sheppard, S. K., X. Didelot, G. Meric, A. Torralbo, K. A. Jolley, D. J. Kelly, S. D. Bentley, M. C. Maiden, et al. (2013) "Genome-wide association study identifies vitamin B5 biosynthesis as a host specificity factor in Campylobacter" Proc Natl Acad Sci U S A 110(29): 11923-11927

Sheppard, S. K., K. A. Jolley and M. C. Maiden (2012) "A gene-by-gene approach to bacterial population genomics: whole genome MLST of *Campylobacter*" Genes (Basel) 3(2): 261-277

Sierra, G. V., H. C. Campa, N. M. Varcacel, I. L. Garcia, P. L. Izquierdo, P. F. Sotolongo, G. V. Casanueva, C. O. Rico, et al. (1991) "Vaccine against group B *Neisseria meningitidis*: protection trial and mass vaccination results in Cuba" NIPH Ann 14(2): 195-207; discussion 208-110

Siguier, P., J. Perochon, L. Lestrade, J. Mahillon and M. Chandler (2006) "ISfinder: the reference centre for bacterial insertion sequences" Nucleic Acids Res 34(Database issue): D32-36

Sim, R. J., M. M. Harrison, E. R. Moxon and C. M. Tang (2000) "Underestimation of meningococci in tonsillar tissue by nasopharyngeal swabbing" Lancet 356(9242): 1653-1654

Sjolinder, H. and A. B. Jonsson (2007) "Imaging of disease dynamics during meningococcal sepsis" PLoS One 2(2): e241

Slaterus, K. W. (1961) "Serological typing of meningococci by means of micro-precipitation" Antonie Van Leeuwenhoek 27: 305-315

Smith, J. M., N. H. Smith, M. O'Rourke and B. G. Spratt (1993) "How clonal are bacteria?" Proc Natl Acad Sci U S A 90(10): 4384-4388

Snape, M. D., T. Dawson, P. Oster, A. Evans, T. M. John, B. Ohene-Kena, J. Findlow, L. M. Yu, et al. (2010) "Immunogenicity of two investigational serogroup B meningococcal vaccines in the first year of life: a randomized comparative trial" Pediatr Infect Dis J 29(11): e71-79

Sorhouet-Pereira, C., A. Efron, P. Gagetti, D. Faccone, M. Regueira, A. Corso, S. I. I. W. G. Argentinean, J. M. Gabastou, et al. (2013) "Phenotypic and genotypic characteristics of *Neisseria meningitidis* disease-causing strains in Argentina, 2010" PLoS One 8(3): e58065

Srikhanta, Y. N., S. J. Dowideit, J. L. Edwards, M. L. Falsetta, H. J. Wu, O. B. Harrison, K. L. Fox, K. L. Seib, et al. (2009) "Phasevarions mediate random switching of gene expression in pathogenic *Neisseria*" PLoS Pathog 5(4): e1000400

Srikhanta, Y. N., K. L. Fox and M. P. Jennings (2010) "The phasevarion: phase variation of type III DNA methyltransferases controls coordinated switching in multiple genes" Nat Rev Microbiol 8(3): 196-206

Srikhanta, Y. N., R. J. Gorrell, J. A. Steen, J. A. Gawthorne, T. Kwok, S. M. Grimmond, R. M. Robins-Browne and M. P. Jennings (2011) "Phasevarion mediated epigenetic gene regulation in *Helicobacter pylori*" PLoS One 6(12): e27569

Srikhanta, Y. N., T. L. Maguire, K. J. Stacey, S. M. Grimmond and M. P. Jennings (2005) "The phasevarion: a genetic system controlling coordinated, random switching of expression of multiple genes" Proc Natl Acad Sci U S A 102(15): 5547-5551

Stephens, D. S. (1999) "Uncloaking the meningococcus: dynamics of carriage and disease" Lancet 353(9157): 941-942

Stephens, D. S. and M. M. Farley (1991) "Pathogenic events during infection of the human nasopharynx with *Neisseria meningitidis* and *Haemophilus influenzae*" Rev Infect Dis 13(1): 22-33

Stephens, D. S., B. Greenwood and P. Brandtzaeg (2007) "Epidemic meningitis, meningococcaemia, and *Neisseria meningitidis*" Lancet 369(9580): 2196-2210

Stojiljkovic, I., J. Larson, V. Hwa, S. Anic and M. So (1996) "HmbR outer membrane receptors of pathogenic *Neisseria* spp.: iron-regulated, hemoglobin-binding proteins with a high level of primary structure conservation" J Bacteriol 178(15): 4670-4678

Stollenwerk, N., M. C. J. Maiden and V. A. A. Jansen (2004) "Diversity in pathogenicity can cause outbreaks of meningococcal disease" Proc Natl Acad Sci U S A 101(27): 10229-10234

Stothard, P. and D. S. Wishart (2005) "Circular genome visualization and exploration using CGView" Bioinformatics 21(4): 537-539

Struve, C., C. C. Roe, M. Stegger, S. G. Stahlhut, D. S. Hansen, D. M. Engelthaler, P. S. Andersen, E. M. Driebe, et al. (2015) "Mapping the evolution of hypervirulent *Klebsiella pneumoniae*" MBio 6(4)

Stucki, D., M. Ballif, T. Bodmer, M. Coscolla, A. M. Maurer, S. Droz, C. Butz, S. Borrell, et al. (2015) "Tracking a tuberculosis outbreak over 21 years: strain-specific single-nucleotide polymorphism typing combined with targeted whole-genome sequencing" J Infect Dis 211(8): 1306-1316

Swain, C. L. and D. R. Martin (2007) "Survival of meningococci outside of the host: implications for acquisition" Epidemiol Infect 135(2): 315-320

Swartley, J. S., A. A. Marfin, S. Edupuganti, L. J. Liu, P. Cieslak, B. Perkins, J. D. Wenger and D. S. Stephens (1997) "Capsule switching of *Neisseria meningitidis*" Proc Natl Acad Sci U S A 94(1): 271-276

Taha, M. K., M. Achtman, J. M. Alonso, B. Greenwood, M. Ramsay, A. Fox, S. Gray and E. Kaczmarski (2000) "Serogroup W135 meningococcal disease in Hajj pilgrims" Lancet 356(9248): 2159

Taha, M. K., S. T. Hedberg, M. Szatanik, E. Hong, C. Ruckly, R. Abad, S. Bertrand, F. Carion, et al. (2010) "Multicenter study for defining the breakpoint for rifampin resistance in Neisseria meningitidis by rpoB sequencing" Antimicrob Agents Chemother 54(9): 3651-3658

Taha, M. K., J. A. Vazquez, E. Hong, D. E. Bennett, S. Bertrand, S. Bukovski, M. T. Cafferkey, F. Carion, et al. (2007) "Target gene sequencing to characterize the penicillin G susceptibility of *Neisseria meningitidis*" Antimicrob Agents Chemother 51(8): 2784-2792

Takahashi, N., S. Ohashi, M. R. Sadykov, Y. Mizutani-Ui and I. Kobayashi (2011) "IS-linked movement of a restriction-modification system" PLoS One 6(1): e16554

Talavera, G. and J. Castresana (2007) "Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments" Syst Biol 56(4): 564-577

Tamura, K., G. Stecher, D. Peterson, A. Filipski and S. Kumar (2013) "MEGA6: Molecular Evolutionary Genetics Analysis version 6.0" Mol Biol Evol 30(12): 2725-2729

Tatusov, R. L., E. V. Koonin and D. J. Lipman (1997) "A genomic perspective on protein families" Science 278(5338): 631-637

Tauseef, I., Y. M. Ali and C. D. Bayliss (2013) "Phase variation of PorA, a major outer membrane protein, mediates escape of bactericidal antibodies by *Neisseria meningitidis*" Infect Immun 81(4): 1374-1380

Tettelin, H., N. J. Saunders, J. Heidelberg, A. C. Jeffries, K. E. Nelson, J. A. Eisen, K. A. Ketchum, D. W. Hood, et al. (2000) "Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58" Science 287(5459): 1809-1815

Teyssou, R. and E. Muros-Le Rouzic (2007) "Meningitis epidemics in Africa: a brief overview" Vaccine 25 Suppl 1: A3-7

Tommassen, J., P. Vermeij, M. Struyve, R. Benz and J. T. Poolman (1990) "Isolation of *Neisseria meningitidis* mutants deficient in class 1 (porA) and class 3 (porB) outer membrane proteins" Infect Immun 58(5): 1355-1359

Toros, B., S. T. Hedberg, M. Unemo, S. Jacobsson, D. M. Hill, P. Olcen, H. Fredlund, H. B. Bratcher, et al. (2015) "Genome-based characterization of emergent invasive *Neisseria meningitidis* serogroup Y isolates in Sweden from 1995 to 2012" J Clin Microbiol 53(7): 2154-2162

Toros, B., S. Thulin Hedberg, S. Jacobsson, H. Fredlund, P. Olcen and P. Molling (2014) "Surveillance of invasive *Neisseria meningitidis* with a serogroup Y update, Sweden 2010 to 2012" Euro Surveill 19(42)

Treangen, T. J., O. H. Ambur, T. Tonjum and E. P. Rocha (2008) "The impact of the neisserial DNA uptake sequences on genome evolution and stability" Genome Biol 9(3): R60

Trotter, C. L., M. Chandra, R. Cano, A. Larrauri, M. E. Ramsay, C. Brehony, K. A. Jolley, M. C. Maiden, et al. (2007) "A surveillance network for meningococcal disease in Europe" FEMS Microbiol Rev 31(1): 27-36

Trotter, C. L., H. Findlow and R. Borrow (2012) "Seroprevalence of serum bactericidal antibodies against group W135 and Y meningococci in England in 2009" Clin Vaccine Immunol 19(2): 219-222

Trotter, C. L., N. J. Gay and W. J. Edmunds (2006) "The natural history of meningococcal carriage and disease" Epidemiol Infect 134(3): 556-566

Trotter, C. L. and B. M. Greenwood (2007) "Meningococcal carriage in the African meningitis belt" Lancet Infect Dis 7(12): 797-803

Tsang, R. S., L. Kiefer, D. K. Law, J. Stoltz, R. Shahin, S. Brown and F. Jamieson (2003) "Outbreak of serogroup C meningococcal disease caused by a variant of *Neisseria meningitidis* serotype 2a ET-15 in a community of men who have sex with men" J Clin Microbiol 41(9): 4411-4414

Tsang, R. S., C. M. Tsai, A. M. Henderson, S. Tyler, D. K. Law, W. Zollinger and F. Jamieson (2008) "Immunochemical studies and genetic background of two *Neisseria meningitidis* isolates expressing unusual capsule polysaccharide antigens with specificities of both serogroup Y and W135" Can J Microbiol 54(3): 229-234

Turner, D. P., A. G. Marietou, L. Johnston, K. K. Ho, A. J. Rogers, K. G. Wooldridge and D. A. Ala'Aldeen (2006) "Characterization of MspA, an immunogenic autotransporter protein that mediates adhesion to epithelial and endothelial cells in *Neisseria meningitidis*" Infect Immun 74(5): 2957-2964

Turner, D. P., K. G. Wooldridge and D. A. Ala'Aldeen (2002) "Autotransported serine protease A of *Neisseria meningitidis*: an immunogenic, surface-exposed outer membrane, and secreted protein" Infect Immun 70(8): 4447-4461

Turner, K. M. and E. J. Feil (2007) "The secret life of the multilocus sequence type" Int J Antimicrob Agents 29(2): 129-135

Tyler, S. and R. Tsang (2004) "Genetic analysis of Canadian isolates of C:2a:P1.2,5 and B:2a:P1.2,5 *Neisseria meningitidis* strains belonging to the hypervirulent clone of ET-15" Can J Microbiol 50(6): 433-443

Tyrrell, G. J., L. Chui, M. Johnson, N. Chang, R. P. Rennie, J. A. Talbot and G. Edmonton Meningococcal Study (2002) "Outbreak of *Neisseria meningitidis*, Edmonton, Alberta, Canada" Emerg Infect Dis 8(5): 519-521

Tzanakaki, G., E. Hong, K. Kesanopoulos, A. Xirogianni, S. Bambini, L. Orlandi, M. Comanducci, A. Muzzi, et al. (2014) "Diversity of Greek meningococcal serogroup B isolates and estimated coverage of the 4CMenB meningococcal vaccine" BMC Microbiol 14: 111

Tzeng, Y. L., L. E. Martin and D. S. Stephens (2013) "Environmental survival of *Neisseria meningitidis*" Epidemiol Infect: 1-4

Unkmeir, A., K. Latsch, G. Dietrich, E. Wintermeyer, B. Schinke, S. Schwender, K. S. Kim, M. Eigenthaler, et al. (2002) "Fibronectin mediates Opc-dependent internalization of *Neisseria meningitidis* in human brain microvascular endothelial cells" Mol Microbiol 46(4): 933-946

Urwin, R., J. E. Russell, E. A. Thompson, E. C. Holmes, I. M. Feavers and M. C. Maiden (2004) "Distribution of surface protein variants among hyperinvasive meningococci: implications for vaccine design" Infect Immun 72(10): 5955-5962

van den Dobbelsteen, G. P., H. H. van Dijken, S. Pillai and L. van Alphen (2007) "Immunogenicity of a combination vaccine containing pneumococcal conjugates and meningococcal PorA OMVs" Vaccine 25(13): 2491-2496

van der Ende, A., C. T. Hopman, S. Zaat, B. B. Essink, B. Berkhout and J. Dankert (1995) "Variable expression of class 1 outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the -10 and -35 regions of the promoter" J Bacteriol 177(9): 2475-2480

van der Ley, P., J. van der Biezen and J. T. Poolman (1995) "Construction of *Neisseria meningitidis* strains carrying multiple chromosomal copies of the porA gene for use in the production of a multivalent outer membrane vesicle vaccine" Vaccine 13(4): 401-407

van der Woude, M. W. and A. J. Baumler (2004) "Phase and antigenic variation in bacteria" Clin Microbiol Rev 17(3): 581-611

van Hoek, A. J., N. Andrews, H. Campbell, G. Amirthalingam, W. J. Edmunds and E. Miller (2013) "The social life of infants in the context of infectious disease transmission; social contacts and mixing patterns of the very young" PLoS One 8(10): e76180

van Tonder, A. J., S. Mistry, J. E. Bray, D. M. Hill, A. J. Cody, C. L. Farmer, K. P. Klugman, A. von Gottberg, et al. (2014) "Defining the estimated core genome of bacterial populations using a bayesian decision model" PLoS Comput Biol 10(8): e1003788

Vasu, K. and V. Nagaraja (2013) "Diverse functions of restriction-modification systems in addition to cellular defense" Microbiol Mol Biol Rev 77(1): 53-72

Vazquez, J. A., L. de la Fuente, S. Berron, M. O'Rourke, N. H. Smith, J. Zhou and B. G. Spratt (1993) "Ecological separation and genetic isolation of *Neisseria gonorrhoeae* and *Neisseria meningitidis*" Curr Biol 3(9): 567-572

Vedros, N. A., C. Y. N. Hoke and P. Chun (1983) "*Neisseria macacae* sp. nov., a new *Neisseria* Species Isolated from the Oropharynges of Rhesus Monkeys (*Macaca mulatta*)" Int J Syst Bacterio 33(3): 515-520

Venables, W. N. and B. D. Ripley (2002). Modern Applied Statistics with S., Springer.

Vesikari, T., S. Esposito, R. Prymula, E. Ypma, I. Kohl, D. Toneatto, P. Dull, A. Kimura, et al. (2013) "Immunogenicity and safety of an investigational multicomponent, recombinant, meningococcal serogroup B vaccine (4CMenB) administered concomitantly with routine infant and child vaccinations: results of two randomised trials" Lancet 381(9869): 825-835

Vieusseux, M. (1805) "Mémoire sur la maladie qui a règné a Geneve au printemps de 1805." J Med Chir Pharmacol 11(163)

Virji, M. (1997) "Post-translational modifications of meningococcal pili. Identification of common substituents: glycans and alpha-glycerophosphate--a review" Gene 192(1): 141-147

Virji, M. (2009) "Pathogenic neisseriae: surface modulation, pathogenesis and infection control" Nat Rev Microbiol 7(4): 274-286

Virji, M., K. Makepeace, D. J. Ferguson, M. Achtman and E. R. Moxon (1993) "Meningococcal Opa and Opc proteins: their role in colonization and invasion of human epithelial and endothelial cells" Mol Microbiol 10(3): 499-510

Vogel, U., H. Claus and M. Frosch (2004) "Genetic lineages and their traits in Neisseria meningitidis" Int J Med Microbiol 294(2-3): 75-82

Vogel, U., H. Claus, M. Frosch and D. A. Caugant (2000) "Molecular basis for distinction of the ET-15 clone within the ET-37 complex of *Neisseria meningitidis*" J Clin Microbiol 38(2): 941-942

Vogel, U., C. Schoen and J. Elias (2010). Population Genetics of *Neisseria meningitidis* Bacterial Population Genetics In Infectious Disease. A. Robinson, D. Falush and E. Feil, John Wiley & Sons, Inc.: 247-267.

Vogel, U., R. Szczepanowski, H. Claus, S. Junemann, K. Prior and D. Harmsen (2012) "Ion torrent personal genome machine sequencing for genomic typing of *Neisseria meningitidis* for rapid determination of multiple layers of typing information" J Clin Microbiol 50(6): 1889-1894

Vogel, U., M. K. Taha, J. A. Vazquez, J. Findlow, H. Claus, P. Stefanelli, D. A. Caugant, P. Kriz, et al. (2013) "Predicted strain coverage of a meningococcal multicomponent vaccine (4CMenB) in Europe: a qualitative and quantitative assessment" Lancet Infect Dis 13(5): 416-425

von Gottberg, A., M. du Plessis, C. Cohen, E. Prentice, S. Schrag, L. de Gouveia, G. Coulson, G. de Jong, et al. (2008) "Emergence of endemic serogroup W135 meningococcal disease associated with a high mortality rate in South Africa" Clin Infect Dis 46(3): 377-386

Vos, M. (2009) "Why do bacteria engage in homologous recombination?" Trends Microbiol 17(6): 226-232

Vu, D. M., T. T. Wong and D. M. Granoff (2011) "Cooperative serum bactericidal activity between human antibodies to meningococcal factor H binding protein and neisserial heparin binding antigen" Vaccine 29(10): 1968-1973

Vyse, A., J. M. Wolter, J. Chen, T. Ng and M. Soriano-Gabarro (2011) "Meningococcal disease in Asia: an under-recognized public health burden" Epidemiol Infect 139(7): 967-985

Wang, J. F., D. A. Caugant, X. Li, X. Hu, J. T. Poolman, B. A. Crowe and M. Achtman (1992) "Clonal and antigenic analysis of serogroup A *Neisseria meningitidis* with particular reference to epidemiological features of epidemic meningitis in the People's Republic of China" Infect Immun 60(12): 5267-5282

Wang, J. F., D. A. Caugant, G. Morelli, B. Koumare and M. Achtman (1993) "Antigenic and Epidemiologic Properties of the Et-37 Complex of *Neisseria meningitidis*" J Infect Dis 167(6): 1320-1329

Waterhouse, A. M., J. B. Procter, D. M. Martin, M. Clamp and G. J. Barton (2009) "Jalview Version 2--a multiple sequence alignment editor and analysis workbench" Bioinformatics 25(9): 1189-1191

Watkins, E. R. and M. C. Maiden (2012) "Persistence of hyperinvasive meningococcal strain types during global spread as recorded in the PubMLST database" PLoS One 7(9): e45349

Wedege, E., E. A. Hoiby, E. Rosenqvist and G. Bjune (1998) "Immune responses against major outer membrane antigens of *Neisseria meningitidis* in vaccinees and controls who contracted meningococcal disease during the Norwegian serogroup B protection trial" Infect Immun 66(7): 3223-3231

Weiss, D. and J. Varma (2013) "Control of recent community-based outbreaks of invasive meningococcal disease in men who have sex with men in Europe and the United States" Euro Surveill 18(28)

Weyand, N. J., A. M. Wertheimer, T. R. Hobbs, J. L. Sisko, N. A. Taku, L. D. Gregston, S. Clary, D. L. Higashi, et al. (2013) "*Neisseria* infection of rhesus macaques as a model to study colonization, transmission, persistence, and horizontal gene transfer" Proc Natl Acad Sci U S A 110(8): 3059-3064

Whalen, C. M., J. C. Hockin, A. Ryan and F. Ashton (1995) "The changing epidemiology of invasive meningococcal disease in Canada, 1985 through 1992. Emergence of a virulent clone of *Neisseria meningitidis*" JAMA 273(5): 390-394

Whelan, J., S. Bambini, A. Biolchi, B. Brunelli and M. Robert-Du Ry van Beest Holle (2015) "Outbreaks of meningococcal B infection and the 4CMenB vaccine: historical and future perspectives" Expert Rev Vaccines 14(5): 713-736

Wormann, M. E., C. L. Horien, J. S. Bennett, K. A. Jolley, M. C. Maiden, C. M. Tang, E. L. Aho and R. M. Exley (2014) "Sequence, distribution and chromosomal context of class I and class II pilin genes of *Neisseria meningitidis* identified in whole genome sequences" BMC Genomics 15(1): 253

Wylie, P. A., D. Stevens, W. Drake, 3rd, J. Stuart and K. Cartwright (1997) "Epidemiology and clinical management of meningococcal disease in west Gloucestershire: retrospective, population based study" BMJ 315(7111): 774-779

Xie, O., A. J. Pollard, J. E. Mueller and G. Norheim (2013) "Emergence of serogroup X meningococcal disease in Africa: need for a vaccine" Vaccine 31(27): 2852-2861

Yazdankhah, S. P. and D. A. Caugant (2004) "*Neisseria meningitidis*: an overview of the carriage state" J Med Microbiol 53(Pt 9): 821-832

Yazdankhah, S. P., P. Kriz, G. Tzanakaki, J. Kremastinou, J. Kalmusova, M. Musilek, T. Alvestad, K. A. Jolley, et al. (2004) "Distribution of serogroups and genotypes among disease-associated and carried isolates of *Neisseria meningitidis* from the Czech Republic, Greece, and Norway" J Clin Microbiol 42(11): 5146-5153

Zerbino, D. R. and E. Birney (2008) "Velvet: Algorithms for de novo short read assembly using de Bruijn graphs" Genome Res 18(5): 821-829

Zhang, D., R. F. de Souza, V. Anantharaman, L. M. Iyer and L. Aravind (2012) "Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics" Biol Direct 7: 18

Zhou, J., L. D. Bowler and B. G. Spratt (1997) "Interspecies recombination, and phylogenetic distortions, within the glutamine synthetase and shikimate dehydrogenase genes of *Neisseria meningitidis* and commensal *Neisseria* species" Mol Microbiol 23(4): 799-812

Zhou, J. and B. G. Spratt (1992) "Sequence diversity within the argF, fbp and recA genes of natural isolates of *Neisseria meningitidis*: interspecies recombination within the argF gene" Mol Microbiol 6(15): 2135-2146

Zhu, P., G. Morelli and M. Achtman (1999) "The opcA and (psi)opcB regions in *Neisseria*: genes, pseudogenes, deletions, insertion elements and DNA islands" Mol Microbiol 33(3): 635-650

Zhu, P., A. van der Ende, D. Falush, N. Brieske, G. Morelli, B. Linz, T. Popovic, I. G. Schuurman, et al. (2001) "Fit genotypes and escape variants of subgroup III *Neisseria meningitidis* during three pandemics of epidemic meningitis" Proc Natl Acad Sci U S A 98(9): 5234-5239

Zhu, Y., J. E. Strassmann and D. C. Queller (2000) "Insertions, substitutions, and the origin of microsatellites" Genet Res 76(3): 227-236

Appendix

Appendix T1. Core-genome loci characteristic of English and Czech Republic lineage 11.2 strains

Locus	Annotation
Cellular processes and signalling	
Capsule	
NEIS0051 (NMC0051)	csc ; alpha-2,8-polysialyltransferase
NEIS0054 (NMC0054)	cssA ; N-acetylglucosamine-6-phosphate 2-epimerase
NEIS0067 (NMC0067)	lipB ; capsule polysaccharide modification protein
*NEIS0066 (NMC0066)	lipA ; capsule polysaccharide modification protein
NEIS0056 (NMC0056)	ctrB ; capsule polysaccharide export inner-membrane protein
*NEIS0059 (NMC0059)	tex ; putative transcriptional accessory protein
Pilus Associated	
NMC0003	pilS2 ; truncated pilin
Cell Wall	
NEIS1634 (NMC1634)	mtrC ; membrane fusion protein
NEIS1632 (NMC1632)	mtrE ; putative outer-membrane lipoprotein
*NMC1706	putative integral membrane efflux protein
NEIS1639 (NMC1639)	putative integral membrane protein
NEIS1637 (NMC1637)	hypothetical integral membrane protein
Cell cycle control, cell division, chromosome partitioning	
NEIS1031 (NMC1031)	putative cell-division protein
Metabolism	
Nucleotide transport and metabolism	
NEIS1419 (NMC1419)	surE ; SurE protein; Purine metabolism
*NEIS0345 (NMC0345)	pyrE ; putative orotate phosphoribosyltransferase; Pyrimidine metabolism
Coenzyme transport and metabolism	
NEIS0378 (NMC0378)	fhs ; putative formate--tetrahydrofolate ligase
Information storage and processing	
Replication, recombination and repair	
NEIS1699 (NMC1699)	MutT-related protein
NEIS1210 (NMC1210)	gcr ; site-specific recombinase
NEIS1638 (NMC1638)	recC ; exodeoxyribonuclease V
Transcription	
NEIS1635 (NMC1635)	mtrR ; transcriptional regulator
Poorly categorised	
General function prediction only	
NEIS1643 (NMC1643)	putative cytochrome oxidase subunit
NMC1640	putative cytochrome (pseudogene)
NMC0439	putative secretion protein
NMC0429	putative two-component system sensor kinase
Function unknown	
NEIS1546 (NMC1546)	conserved hypothetical protein
NMC0068	conserved hypothetical protein
NEIS0288 (NMC0288)	conserved hypothetical protein
NEIS1636 (NMC1636)	hypothetical protein
NEIS0430 (NMC0430)	hypothetical protein
NMC1201	hypothetical protein
NEIS0348 (NMC0348)	hypothetical protein

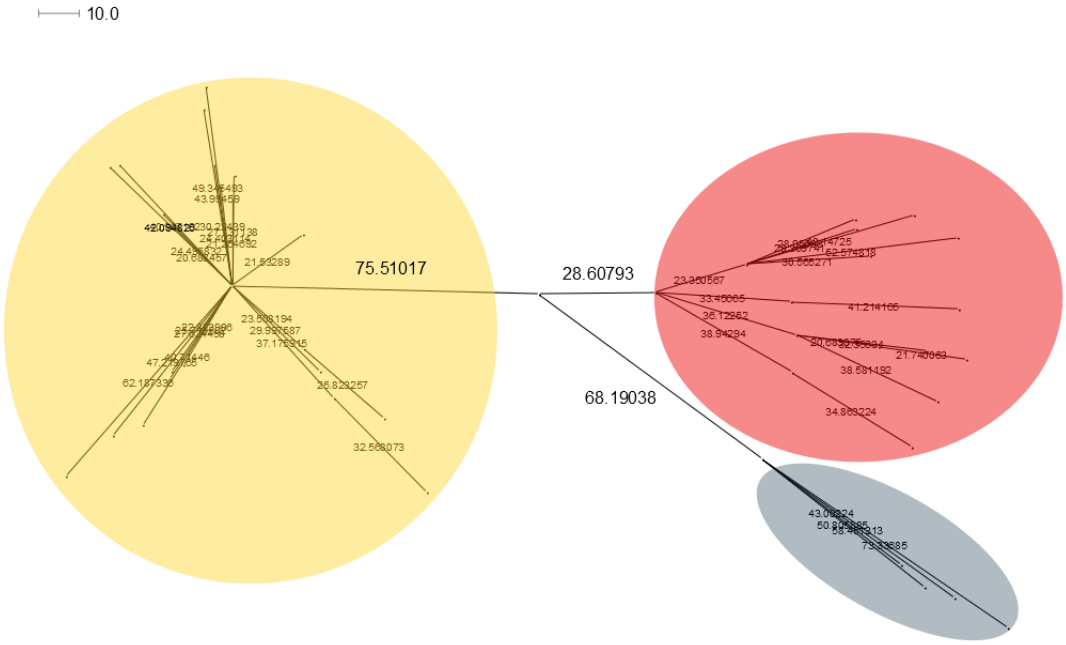
*Indicates loci that also distinguished lineage 11.2 from lineage 11.1

Appendix T2: Size of Czech Republic lineage 11.2 core- and accessory-genome from progressiveMauve alignment.

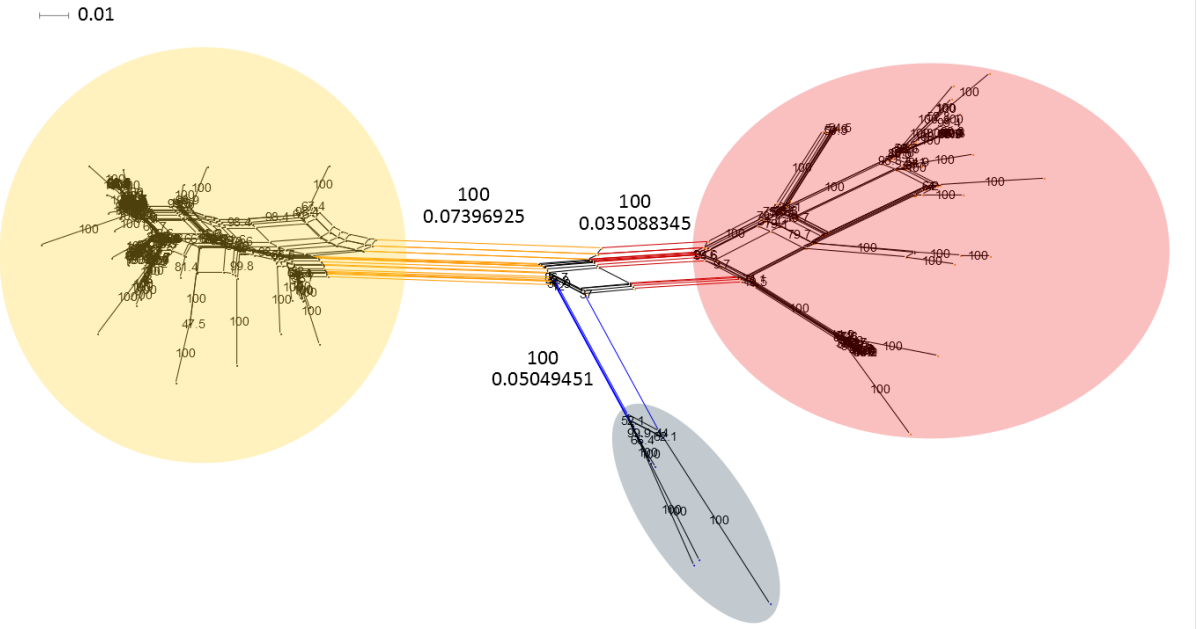
PubMLST ID	Isolate	Genome length (bp)	Sample core-genome		Sample accessory-genome		Unique regions	
			(bp)	(%)	(bp)	(%)	(bp)	(%)
14765	0002/93	2,123,517	1,887,181	88.9	152,073	7.2	84,263	4.0
8159	0383/93	2,125,702	1,887,228	88.8	68,884	3.2	169,590	8.0
8157	0381/93	2,125,901	1,887,225	88.8	110,400	5.2	128,276	6.0
15242	0351/93	2,133,120	1,887,177	88.5	166,511	7.8	79,432	3.7
939	0022/93	2,112,188	1,887,186	89.3	163,440	7.7	61,562	2.9
940	0023/93	2,118,477	1,887,182	89.1	122,123	5.8	109,172	5.2
942	0025/93	2,068,777	1,887,229	91.2	181,336	8.8	212	0.0
946	0033/93	2,101,374	1,887,171	89.8	133,894	6.4	80,309	3.8
949	0036/93	2,121,276	1,887,176	89.0	97,136	4.6	136,964	6.5
950	0037/93	2,134,257	1,887,183	88.4	188,934	8.9	58,140	2.7
952	0039/93	2,132,148	1,887,183	88.5	120,532	5.7	124,433	5.8
954	0042/93	2,138,415	1,887,181	88.3	243,423	11.4	7,811	0.4
955	0043/93	2,140,635	1,887,215	88.2	156,444	7.3	96,976	4.5
957	0045/93	2,138,576	1,887,181	88.2	245,514	11.5	5,881	0.3
958	0046/93	2,137,692	1,887,143	88.3	142,362	6.7	108,187	5.1
978	0073/93	2,129,617	1,887,190	88.6	144,405	6.8	98,022	4.6
979	0074/93	2,124,121	1,887,183	88.8	135,031	6.4	101,907	4.8
20847	0075/93	2,132,671	1,887,183	88.5	240,103	11.3	5,385	0.3
982	0079/93	2,131,934	1,887,181	88.5	200,489	9.4	44,264	2.1
992	0093/93	2,133,554	1,887,182	88.5	168,515	7.9	77,857	3.6
993	0095/93	2,090,733	1,887,145	90.3	195,741	9.4	7,847	0.4
994	0096/93	2,137,361	1,887,182	88.3	236,667	11.1	13,512	0.6
1212	0099/93	2,132,353	1,887,181	88.5	171,826	8.1	73,346	3.4
1570	0101/93	2,130,264	1,887,186	88.6	161,445	7.6	81,633	3.8
1571	0102/93	2,119,792	1,887,184	89.0	133,559	6.3	99,049	4.7
1572	0103/93	2,130,739	1,887,184	88.6	161,783	7.6	81,772	3.8
1573	0104/93	2,133,016	1,887,179	88.5	170,299	8.0	75,538	3.5
1578	0109/93	2,115,442	1,887,204	89.2	105,945	5.0	122,293	5.8
1582	0113/93	2,121,540	1,887,211	89.0	125,814	5.9	108,515	5.1
1583	0115/93	2,118,706	1,887,198	89.1	148,357	7.0	83,151	3.9
1587	0119/93	2,117,925	1,887,178	89.1	49,060	2.3	181,687	8.6
15032	0186/93	2,119,751	1,887,216	89.0	7,739	0.4	224,796	10.6
15035	0188/93	2,128,217	1,887,170	88.7	94,580	4.4	146,467	6.9
684	0259/93	2,127,990	1,887,183	88.7	230,517	10.8	10,290	0.5
15152	0263/93	2,123,565	1,887,170	88.9	152,578	7.2	83,817	3.9
15154	0264/93	2,132,327	1,887,186	88.5	234,322	11.0	10,819	0.5
1650	0270/93	2,118,890	1,887,198	89.1	172,768	8.2	58,924	2.8
1652	0272/93	2,128,386	1,887,190	88.7	161,329	7.6	79,867	3.8
15174	0284/93	2,125,794	1,887,184	88.8	106,641	5.0	131,969	6.2
15176	0287/93	2,119,060	1,887,180	89.1	153,275	7.2	78,605	3.7
15238	0343/93	2,129,286	1,887,196	88.6	97,365	4.6	144,725	6.8
15256	0393/93	2,125,390	1,887,184	88.8	190,949	9.0	47,257	2.2
8160	0395/93	2,120,595	1,887,184	89.0	46,898	2.2	186,513	8.8

15261	0400/93	2,119,337	1,887,183	89.0	147,998	7.0	84,156	4.0
15315	0489/93	2,125,980	1,887,173	88.8	172,131	8.1	66,676	3.1
15316	0490/93	2,121,934	1,887,192	88.9	115,731	5.5	119,011	5.6
15319	0500/93	2,133,554	1,887,184	88.5	131,316	6.2	115,054	5.4
15325	0512/93	2,103,790	1,887,200	89.7	71,061	3.4	145,529	6.9
948	0035/93	2,125,965	1,887,186	88.8	190,051	8.9	48,728	2.3
15330	0520/93	2,123,308	1,887,199	88.9	152,635	7.2	83,474	3.9
15336	0529/93	2,130,560	1,887,183	88.6	138,108	6.5	105,269	4.9
20848	0534/93	2,117,520	1,887,230	89.1	115,752	5.5	114,538	5.4
15342	0535/93	2,117,370	1,887,179	89.1	125,564	5.9	104,627	4.9
15344	0538/93	2,125,080	1,887,149	88.8	154,840	7.3	83,091	3.9
Average		**2,123,990**	**1,887,187**	**88.9**	**148,263**	**7.0**	**88,541**	**4.2**

a)



b)



Appendix F1: Support for lineage 23 (cc23) sub-lineages. Neighbor-Net graphs were generated from 1,592 N. meningitidis core-genome loci in a sample of lineage 23 isolates from PubMLST (n=118). Sub-lineage 23.1 isolates are indicated in yellow, 23.2 isolates in red, and 23.3 isolates in grey. a) Distances were generated from allele profiles as in Fig. 5c: only well supported/high weight splits are displayed here (weight >20). b) Distances were generated from the variable-sites (n=19,686) in concatenated alignments of the core-genome loci. Bootstrap values (1000 replicates) are displayed on edges; the weights of the edges separating sub-lineages are displayed under bootstrap values