Genomic analyses of polysaccharide capsules in *Neisseria* species

Marianne Clemence

St Edmund Hall

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2019

Acknowledgements

The thesis presented here is my own work, but it would not have been possible without the advice, encouragement and support of numerous people.

I would first like to thank my supervisor, Professor Martin Maiden, who has seen me through from a timid undergraduate to a confident DPhil student, and Dr Odile Harrison, who first nurtured my interest in Neisserial genomics and has given invaluable advice about the capsule. I would like to extend my thanks to all the members of the Maiden Lab for creating such a friendly working environment, particularly to Dr James Bray for his advice on coding and enthusiasm for "Lunch Club"; Dr Charlene Rodrigues for being a soundboard to so many ideas and her assistance with vaccine typing; Dr Holly Bratcher for supporting my lab work; Dr Jenny MacLennan for her advice about carriage isolates; Dr Keith Jolley, for all his work on PubMLST; and of course Andrea Kastner, who can always be relied on for help with administration.

I was very lucky to visit the meningococcal group at the National Institute for Biological Standards and Control, and am grateful to Professor Ian Feavers, Dr Caroline Vipond and Dr Sunil Maharjan for their efforts to make that happen, and to the rest of the meningococcal group for their help and advice. Sunil was an excellent mentor and collaborator in my time spent there, and I would also like to extend my thanks to Dr Tim Rudd, who performed the NMR on the samples. The fruit of this collaboration is found in Chapter 2.

None of this would have been possible without the financial support of the Wellcome Trust, and the organisers of the programme in Infection, Immunology and

Translational Medicine. I am grateful to all the course directors, especially Professor Chris Tang for the advice he has given me, and to all the administrators past and present. I would also like to extend my thanks to the Medical Sciences Doctoral Training Centre director Professor Robert Gilbert who, during my time as student representative, was always a conscientious listener. Further thanks is due to the organisers and teachers of the Medical Sciences Division skills programme, especially Professor Afsie Sabokbar, who was always brilliant and full of reassurance.

Finally, I would not be here without the steadfast support of all my friends and family. I am so thankful to my parents, who have always supported my choices and ambitions, whilst providing an environment that gave me the opportunity to explore my skills and interests. I am also grateful for the companionship of all my friends at Oxford. I have met people from all over the world, which has shown me the value of international cooperation and tolerance. Peer support is an important part of research, and I would like to thank my IITM course mates Hannah Behrens, Layal Liverpool, Lucy Garner, Cherelle Dacon and Richard Morter for their good humour and comradeship, and also Joe Donlan for our mutual encouragement through writing up. Particular thanks and love go to all those I have met in Hertford College Chapel Choir, especially Nick, Jane, Joe, Ali, Faaris, Hannah, Vicky, Tom, Lizzy, Kassie, Madalena, Tabitha, Nora, Mia, and Rafa, whom we all miss. Lastly, because he is so important, it is not possible to overstate my gratitude to my partner Sam Forbes who, through the ups and downs of the last four years, has been my rock, my mentor, and most of all my friend.

Declarations & Contributions

Declarations

The work herein is my own, except where explicitly stated in relevant chapters and in Contributions. No part of this work has been submitted for any other degree or professional qualification. Where work has been published during the course of this DPhil, this has been stated. Permission has been sought from co-authors of published papers to use the work that I produced as part of my thesis, including written text, figures and tables. The papers themselves are licensed under a Creative Commons Attribution License.

Contributions

This thesis made use of: the *Neisseria* Multi Locus Sequence Typing website (`https://pubmlst.org/neisseria/`) sited at the University of Oxford (Jolley *et al.* Wellcome Open Res 2018, 3(124) [version 1; referees: 2 approved]), the development of which was funded by the Wellcome Trust and European Union; the Meningitis Research Foundation Meningococcus Genome Library (`http://www.meningitis.org/research/genome`) developed by Public Health England, the Wellcome Trust Sanger Institute and the University of Oxford as a collaboration, funded by the Meningitis Research Foundation; and isolates collected as part of the UKMenCar4, carried out at the University of Oxford. All these isolates, and published sequences from a variety of sources, are listed in Appendix A.

Dr Odile Harrison first conceived of the presence of polysaccharide capsule genes in non-pathogenic *Neisseria* species, and provided advice in the development of

especially Chapter 2. All analyses, figures, tables and text pertaining to genomic analyses presented in Chapter 2 are my own work.

The biochemical analyses in Chapter 2 were carried out in collaboration with colleagues from the National Institute for Biological Standards and Control (NIBSC). Dr Sunil Maharjan developed the extraction method and supervised me in the lab. Dr Maharjan also undertook methanol purification of samples for the preparation of the ^{1}H NMR analyses in Figure 2.7. Dr Tim Rudd carried out ^{1}H NMR at the NIBSC NMR facility, and produced and assisted with the analysis of the NMR spectra in Figures 2.6 and 2.7. Dr George Kemp carried out limulus amebocyte lysate assays to identify LOS contamination. I produced all the text pertaining to these analyses.

Figures 1.1, 1.2, 1.4, 1.5, 1.6, 1.9, 1.10 and 1.11, used to illustrate the Introduction, were obtained or adapted from previously published sources. Permission was sought for images that were not in the public domain or licensed under a Creative Commons Attribution License. Licenses were granted for reuse of figures in this thesis from Elsevier, Springer Nature and Oxford University Press. Taylor & Francis offers reuse of its content in a thesis free of charge.

Abstract

Neisseria meningitidis is one of several *Neisseria* species that asymptomatically colonise the human nasopharynx. Occasionally, *N. meningitidis* invades the bloodstream, resulting in invasive meningococcal disease (IMD). In most settings, expression of a serogroup A, B, C, W, X or Y polysaccharide capsule is essential for IMD. Previously, capsules were considered unique to *N. meningitidis*, among the *Neisseria*, possibly acquired by horizontal genetic transfer (HGT). In this thesis, genomic analyses further investigate the distribution and evolution of capsules in *Neisseria* species.

Chapter 2 identifies orthologues of conserved capsule export genes in 13 non-pathogenic *Neisseria* species. Novel capsule synthesis genes are further characterised and found to resemble those present in other encapsulated bacteria including *N. meningitidis*, *Actinobacillus* and *Haemophilus*. Preliminary ^{1}H NMR results are consistent with some of the structural predictions arising from these comparisons.

Phylogenetic analyses in Chapter 3 indicate that the meningococcal capsule has undergone HGT with other *Neisseria* species on at least two occasions. Sequence analyses are consistent with hypotheses postulating that the meningococcal capsule was acquired *de novo* through HGT, with *Neisseria subflava* identified as a potential donor, but are not sufficient to definitively prove that such an event occurred.

Finally, Chapter 4 explores a more contemporary example of HGT within the meningococcal capsule locus. Capsule switching from serogroup E to serogroup B is shown to be responsible for cases of IMD in clonal complexes cc60 and cc1157. This

example is further explored to discuss the interaction between the capsule and the wider genetic profile of meningococci, and their association with IMD.

These analyses expand on previous understanding of the distribution and evolution of the polysaccharide capsule in the *Neisseria*, and the implications for the epidemiology of IMD. Discussions of this important virulence-associated factor address its enigmatic existence in organisms that principally colonise the nasopharynx without causing disease.

Contents

List of Figures

List of Tables

List of Abbreviations

ABC . . . . . ATP-binding cassette

ACT . . . . . Artemis Comparison Tool

BAST . . . . . Bexsero Antigen Sequence Typing

BIGSdb . . . . Bacterial Isolate Genome Sequence Database

cgMLST . . . Core genome multi-locus sequence typing

cnl . . . . . Capsule null locus

cps . . . . . Capsule locus

DUS . . . . . DNA uptake sequence

EDTA . . . . . Ethylenediaminetetraacetic acid

HGT . . . . . Horizontal genetic transfer

IMD . . . . . Invasive meningococcal disease

LAL assay . . Limulus amebocyte lysate assay

LOS . . . . . Lipooligosaccharide

MDA phage . Meningococcal disease-associated phage

MLST . . . . . Multi-locus sequence typing

MRF-MGL . . Meningitis Research Foundation Meningococcus Research Library

NCBI . . . . . National Center for Biotechnology Information

NGS . . . . . Next generation sequencing

NIBSC . . . . National Institution for Biological Standards and Control

NPN species . Non-pathogenic *Neisseria* species

OMV . . . . . Outer membrane vesicle

OOB error . . Out-of-bag error

ORF . . . . . Open reading frame

RDP4 . . . . . Recombination Detection Programme 4

rMLST . . . . Ribosomal multi-locus sequence typing

VR2 . . . . . Variant region 2

WGS data . . Whole genome sequence data

1 Introduction

1.1 The genus *Neisseria*

The *Neisseria* are a diverse genus of Gram-negative β-proteobacteria that asymptomatically colonise the mucosal surfaces of humans and other animals [1, 2, 3, 4, 5]. Many *Neisseria* species belong to the human oral microbiota, and have been isolated from the mouth, nose, throat and urogenital tract. The two most well studied species within the genus are those that are capable of causing human disease, *Neisseria meningitidis* and *Neisseria gonorrhoeae*. *N. meningitidis* is capable of invading the bloodstream, leading to life-threatening invasive meningococcal disease (IMD), comprising meningitis and/or septicaemia [6]. The closest relative of *N. meningitidis*, *N. gonorrhoeae*, is the second most prevalent cause of bacterial sexual disease worldwide [7]. Other *Neisseria* species occasionally act as opportunistic pathogens, but these are isolated incidences [8, 9, 10, 11]. Therefore, from a human perspective, all other *Neisseria* can be considered to be non-pathogenic *Neisseria* (NPN) species.

The characterised NPN species most closely related to *N. meningitidis* and *N. gonorrhoeae*, in order of relatedness, are: *Neisseria polysaccharea*; *'Neisseria bergeri'*; *Neisseria lactamica*; *Neisseria cinerea*; *Neisseria subflava*; and *Neisseria oralis* and *Neisseria mucosa* [13, 12]. These relationships are based on ribosomal

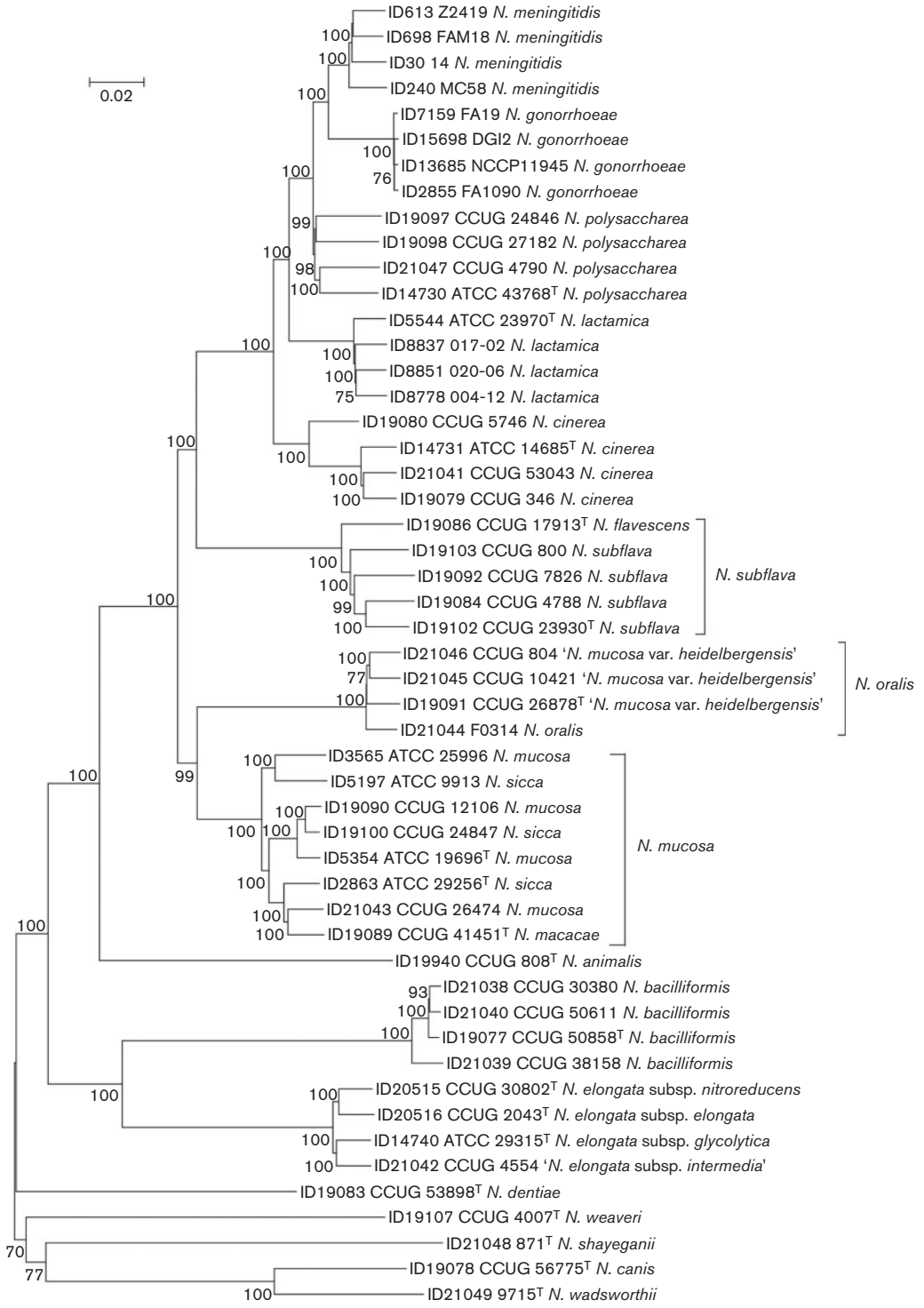

Figure 1.1: *Neisseria* species phylogeny. Neighbour-joining tree reconstructed from 246 concatenated core gene sequences. PubMLST ID, isolate and assigned species included. Species re-classifications indicated by brackets are used in this thesis. Obtained from Bennett *et al.*, 2013, Int J Syst Evol Micr 63(10) (`https://doi.org/10.1099/ijs.0.052431-0`) [12]. Licensed under a Creative Commons Attribution License.

multilocus sequence typing (rMLST) classification, using sequence fragments from ribosomal protein sub-unit gene variants to reliably classify phylogenetic relationships among bacteria [14], which has been confirmed by a 246 core genome scheme for the *Neisseria* genus (Figure 1.1) [13, 12]. *N. polysaccharea* is a paraphyly [13]. Re-classification of some *N. polysaccharea* isolates, as well as several new isolates related to *'N. bergeri'*, has been recently proposed [15], but as they have not been confirmed, the classification of Bennett *et al.* [12] will be used in this thesis. All of the species mentioned thus far have a diplococcus morphology. They have all also been isolated from the human mouth and/or throat, although differences in tropism have been identified, with *N. lactamica*, like *N. meningitidis*, more typically inhabiting the throat, compared to *N. subflava* on the tongue dorsum, and *N. mucosa* in the gingivial plaque [16]. The more distantly related species *Neisseria elongata* [17] and *Neisseria bacilliformis* [11] have a bacillus morphology. *N. elongata* has been classified into three subspecies, *elongata*, *nitroreducens* and *glycolytica*, on the basis of biochemical profiles [18]. The subspecies all belong to a single clade, but only one representative of each subspecies has been sequenced to date [12]. *N. subflava*, *N. polysaccharea* and *N. lactamica* have been most frequently isolated and sequenced during investigations into nasopharyngeal carriage of *N. meningitidis* [19].

In addition to these human-associated species, several animal-associated species have been identified. These include *Neisseria animalis* from a guinea-pig [20], *Neisseria weixii* from a Chinese rodent [21], and *Neisseria musculi* from the wild house mouse, which shows potential as a tractable animal model [2]. Human disease has been associated with infection *via* animal bites with *Neisseria weaveri*, *Neisseria canis*, *Neisseria zoodegmatis* and *Neisseria animaloris* [1]. Nevertheless, Neisserial disease is not widely reported among animals that are known to be colonised, and so presumably infection in animals is also generally asymptomatic.

1.2 The Meningococcus

In common with its NPN relatives, *N. meningitidis* is generally a commensal bacterium. It is frequently isolated from the human nasopharynx, with recorded asymptomatic carriage rates ranging from 1-40% or more, depending on age group, geographical location and setting [22]. Occasionally, *N. meningitidis* invades the bloodstream where, dependent on the possession of certain genetic factors and host-pathogen interactions, it is able to evade immune responses, causing IMD [6]. Disease is predominantly caused by the six serogroups A, B, C, W, X and Y, defined by expression of structurally distinct polysaccharide capsules [23, 24]. IMD usually presents as meningitis and/or septicaemia, which have high mortality rates, and are a public health priority in many jurisdictions.

1.2.1 Epidemiology

Unlike gonorrhoea, which was described in Egypt possibly as early as 1500 B.C. [25], the characteristics of IMD epidemics were not described until an outbreak in Geneva, Europe, in 1805 [26]. Endemic disease in Europe, the Americas and Australia was predominantly caused by serogroups B and C prior to the introduction of the MenC vaccine [22]. Cases of disease in Western countries tend to be sporadic, occurring at a rate of 1-3 cases per 100,000 population. Epidemics have been described, but they tend to result in a modest increase in cases that takes place over several years [27]. Rates of serogroup Y disease in the USA increased in the 1990s, [28], and have been reported to be increasing in the 2010s in Europe [29]. The emergence of a South American/UK lineage has resulted in increased rates of W disease with high case fatalities in Europe [30].

The first African meningococcal outbreak was recorded in Algiers in 1840 [31]. During the 20$^{\text{th}}$ Century, countries in the Sahel spanning from The Gambia to Ethiopia and Sudan, constituting the African meningitis belt [32], were subject to

frequent epidemics of predominantly serogroup A disease [33]. These epidemics were characterised by substantially higher rates of disease compared to endemic levels [22]. In the early 2000s, emergence of serogroup W disease in the belt was associated with Hajj pilgrims; pilgrims are now required to be vaccinated [34]. Since the delivery of MenAfriVac, a serogroup A vaccine, disease has declined across the belt, and cases are now predominantly caused by serogroups C, W and X [35].

1.2.2 Carriage

Meningococci isolated from IMD represent only a small proportion of the genetic diversity of meningococci, the majority of which corresponds to carriage [36]. Carriage is the more usual state in the meningococcal life-cycle, during which *N. meningitidis* asymptomatically colonises the mucosa of the nasopharynx. Carriage can persist within a host over several weeks [37, 38]. No known reservoirs besides humans exist, so carried meningococci are the major source of transmission between hosts (Figure 1.2) [39]. The highest burden of disease is in infants and children under the age of five, but carriage in this age group is relatively low [40]. By contrast, in Europe and other Western countries, carriage increases in adolescence to a peak at 19 years old [41]. Increased rates of carriage likely explain a small peak in IMD in this age group, but it is also likely that adolescent carriers are the main source of transmitted meningococci to infants, the group at the highest risk of IMD [42].

The exact process of transmission is yet to be determined, but several risk factors have been associated with carriage. Owing to the ethical concerns of infecting individuals with potentially pathogenic bacteria, and the relative intractability of animal models, most research is focused on natural carriage studies to investigate epidemiology, though these may not provide information about the duration of carriage, co-carriage of multiple strains, or rates of transmission. It has been consistently observed that overcrowding, including within communal living areas such as university halls of residence and military barracks, increases the risk of meningococcal carriage

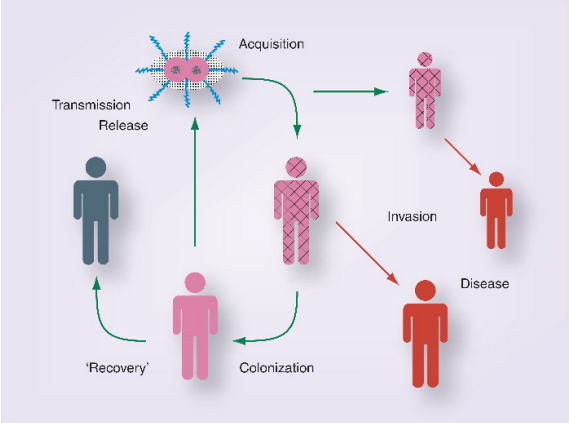

Figure 1.2: Transmission of meningococci in a typical human population. The meningococcal life-cycle predominantly consists of colonisation and transmission, as depicted by the green arrows. Obtained from Trotter and Maiden, 2009, Expert Rev Vaccines 8(7) (`https://doi.org/10.1586/erv.09.48`) [43]. Reused from Taylor & Francis.

[44, 45, 38]. Based on this observation, transmission is assumed to require close-contact [46]. Social interaction, including intimate behaviours, are also associated with carriage, and increased participation in these activities during adolescence may explain increased carriage rates in this age group [47]. Smoking and indoor cooking also increase the risk of carriage of meningococci [47, 48, 19]. Carriage is a crucial component of meningococcal epidemiology, and can inform decisions about vaccination strategies [42, 49].

1.2.3 Phylogenetics and typing

The study of meningococcal epidemiology has benefited from meningococcal typing, predominantly using the seven locus multilocus sequence typing (MLST) scheme [50, 51]. This scheme consists of sequence fragments of seven "house-keeping" genes, chosen from across the genome to minimise linkage (Figure 1.3). Each fragment is assigned an arbitrary allele number, with meningococci containing identical allele sequences being assigned the same allele number. The set of allele numbers representing the seven gene fragments is the MLST allele profile. Each unique MLST allele profile is assigned a sequence type, and based on heuristic measures, sequence types are clustered into clonal complexes. Different clonal complexes are often characterised by

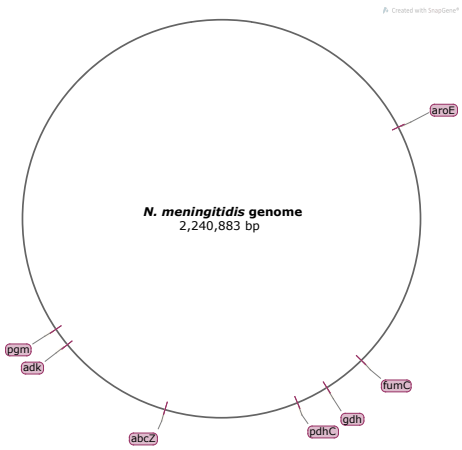

Figure 1.3: Chromosomal location of *N. meningitidis* MLST gene fragments. Fragments are mapped to the chromosome of strain H44/76.

other genotypic and phenotypic characteristics. For instance, clonal complexes cc198, cc53, cc1117 and cc1136 are found to be consistently unencapsulated and associated with carriage [52], whilst clonal complexes cc41/44, cc11, cc32, cc8 and cc269 are most frequently isolated from IMD in Europe [53]. Although Next Generation Sequencing (NGS) technologies have enabled higher resolution core genome multilocus sequence typing (cgMLST), which uses 1605 genomic loci [54], clonal complex is still a primary typing measure reported by reference laboratories.

It is difficult to determine the phylogenetic relationships among clonal complexes. Phylogenetic analyses at high and low resolution have been shown to produce star-like trees, with poorly resolved, deep internal branches [55, 56]. This is likely to be due to extensive recombination among meningococci [51], which occurs as a result of horizontal exchange of genetic material between bacteria [57].

Among bacteria, there are three major mechanisms of horizontal genetic transfer (HGT): conjugation is mediated by direct contact between two bacterial cells; transduction involves the transfer of genetic material carried by a viral phages that incorporate into the bacterial genome; and transformation involves the uptake of free DNA from the environment [58, 59]. The *Neisseria*, including *N. meningitidis*, are naturally competent for transformation. Transformation between *Neisseria* species

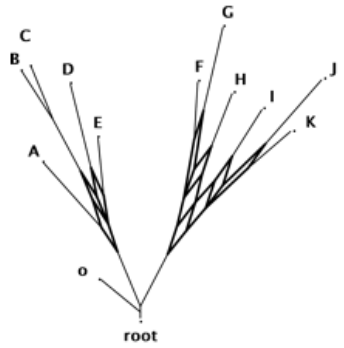

Figure 1.4: Example of a phylogenetic network. A split network representing incompatible phylogenetic signals that may arise due to recombination. Edges represent a split, separating the taxa into two groups. The length of an edge corresponds to the weight of the split. Increasing numbers of parallel edges indicates that there are conflicting splits. Obtained from Huson and Bryant, 2006, Mol Biol Evol 23(2) (`https://doi.org/10.1093/molbev/msj030`) [61]. Reused with permission from Oxford University Press.

is promoted due to the 10 bp DNA uptake sequence (DUS) 5' -GCCGTCTGAA-3', which is associated with recombinant genome sequences [46]. Uptake is mediated through binding of the DUS to the type IV pilus, followed by retraction of the type IV pilus [2]. Following uptake, DNA must be processed and integrated into the host genome to complete the HGT event.

HGT among meningococci gives rise to the concept of a global gene pool, in which any allele present in the population can be potentially transferred among unrelated isolates [60]. The gene pool includes core genes, which are essential for basic bacterial survival, as well as accessory genes, which confer additional but non-essential traits [56]. As a result of HGT, bifurcating trees are not especially reliable for within species evolutionary analyses of meningococci, except in the context of short-term evolution [51]. Evolutionary relationships of recombining sequences may more accurately be represented by phylogenetic networks (Figure 1.4) [61]. Inter-species recombination, which has been demonstrated among the *Neisseria*, including *N. meningitidis*, can blur the lines between distinct species, but it occurs insufficiently frequently to affect the established speciation of the *Neisseria* [60, 56].

Genetic analysis of meningococci has been accelerated by the NGS revolution, which enables rapid whole genome sequencing of isolates. *Neisseria* whole genome sequence (WGS) data from a variety of sources have been deposited in the PubMLST database at pubMLST.org/*neisseria*, which is hosted on the Bacterial Isolate Genome Sequence Database (BIGSdb) genomics platform [62]. As of August 2019, the public database contained WGS data from >20,006 *Neisseria* isolates, 499 of which were from defined NPN species, with a further 61 isolates classified as '*Neisseria* sp'. Most of the WGS data in PubMLST is high quality draft genome data, with sequencing reads assembled into approximately 100-300 individual contigs. The PubMLST sequence database contains defined *Neisseria* loci and allele sequences, with each locus assigned a unique NEIS number. Isolates can be annotated with NEIS loci automatically or manually through a BLAST-based process, and new alleles are assigned an arbitrary allele number. This enables genomic analyses of meningococci using both sequence based approaches, and gene-by-gene approaches based on allelic profiles, to investigate epidemiology and factors involved in immunity, colonisation and virulence.

1.2.4 Virulence

A number of processes must take place in order for *N. meningitidis* to cause IMD. Colonisation of the nasopharynx alone is not sufficient. The major processes are: invasion across the mucosal layer and into the bloodstream; survival and proliferation in the bloodstream, entailing acquisition of nutrients and evasion of the immune system; and, in the case of meningitis, crossing of the blood-brain barrier [6]. Any factor required for one of these processes could be considered a virulence factor [63], and dozens of genes have been implicated in the virulence of meningococci [64]. These genes are involved in a range of processes including mucosal attachment, iron acquisition, and immune invasion, but many of them have also been identified in NPN species [64, 65]. Whilst virulence may be attenuated in the absence of these factors, they do not necessarily represent the most important factors for defining

the pathogenic potential of meningococci compared to NPN species. Marri *et al.* described such factors as "host adaptation factors" [64], whilst Snyder *et al.* suggested that the combination of factors may define a virulence profile [65].

The unequal distribution of clonal complexes in meningococci isolated from disease, compared to meningococci isolated from carriage, indicates that certain factors or genetic variants may more directly facilitate IMD. Leading candidates include: HmbR, involved in haemoglobin acquisition, which is over-represented among disease-associated meningococci [66]; factor H binding protein (fHbp), which negatively regulates the alternative complement pathway, variant 1 of which is more likely to be associated with disease [67]; the type four pilus, involved in attachment, of which class II is associated with invasive meningococcal lineages [68, 69]; and the meningococcal disease-associated (MDA) bacteriophage, an 8 kb genetic island whose presence is associated with IMD [70, 71]. The factor most convincingly associated with IMD is the polysaccharide capsule [72].

1.3 The polysaccharide capsule

Polysaccharide capsules consist of a layer of polysaccharides that form into a structure attached covalently to phospholipids or lipid A on the cell surface [73]. In most cases, including all meningococcal capsules, the capsule is negatively charged and contains a high water content [74], constituting a protective layer around the bacterial cell surface. Capsules can function in environmental survival, including desiccation resistance [75], as well as evasion of the host immune system [76]. Capsules that confer the ability to evade the immune system are considered important virulence factors. In addition to *N. meningitidis*, capsules have been identified in many Gram-negative pathogens including *Escherichia coli*, *Actinobacillus pleuropneumoniae*, *Haemophilus influenzae*, *Pasteurella multocida*, *Mannheimia haemolytica* and *Moraxella nonliquifaciens* [76]. Gram-negative capsules are produced *via* either Wzy-dependent or ATP-binding

cassette (ABC) transporter-dependent pathways. Capsular biosynthesis has been most well characterised in *E. coli* [77], but capsules have also been well characterised in other pathogens such as *N. meningitidis*.

1.3.1 Synthesis and structure

In *N. meningitidis*, the capsule is produced *via* ABC transporter-dependent polymerisation, whereby synthesis and polymerisation of the polysaccharide take place at the bacterial inner membrane, prior to transport across the membrane and translocation to the cell surface using energy released from ATP hydrolysis [79]. These processes are encoded by genes located in the capsule locus (*cps*) [78], part of the accessory genome (Figure 1.5). The *cps* is functionally divided into several contiguous regions.

Region A contains genes involved in capsule synthesis, in particular glycosyltransferases and capsule polymerases, but also other proteins involved in additional capsule modifications such as acetyltransferases, and sometimes insertion sequences are present (Figure 1.5) [78, 76]. This region is highly variable, with 12 known variants corresponding to the 12 meningococcal serogroups A, B, C, W, X, Y, E, L, H, I, K and Z. Of these, only six (A, B, C, W, X and Y) are associated with disease [23, 24]. The structures of the capsule serogroups have been determined through extraction and purification of the capsule polysaccharides, followed by nuclear magnetic resonance imaging (NMR) analysis. A full summary of capsule serogroup structures can be found in Table 1.1.

Serogroups B, C, W and Y produce capsules containing sialic acid derivatives [81, 82]. Region A of the *cps* in all four contains the cytidine-5'-monophosphate-N-acetylneuraminic acid synthesis genes *cssABC*, and differ by the the capsule polymerase, containing either *csb*, *csc*, *csw* or *csy* respectively [78]. *csb* and *csc* nucleotide sequences are 64% identical, and determine whether $\alpha 2 \rightarrow 8$-linked (serogroup B) or $\alpha 2 \rightarrow 9$-linked (serogroup C) polysaccharides are synthesised [83]. *csw* and *csy* nucleotide sequences are 98% identical, and a single amino acid substitution

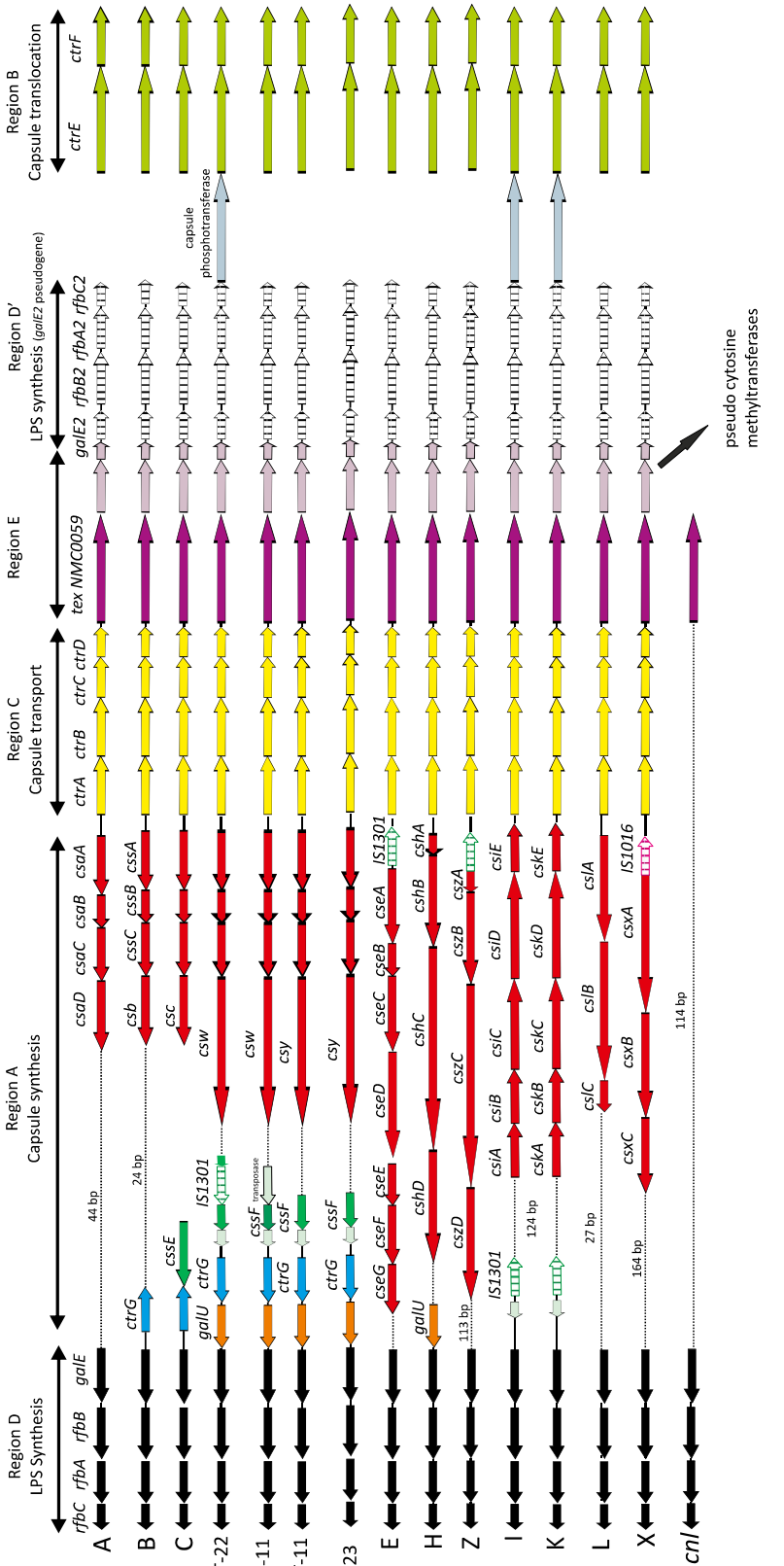

Figure 1.5: Meningococcal *cps* organisation. Obtained from Harrison *et al.* 2013, Emerg Infect Dis 19(4) (https://dx.doi.org/10.3201/eid1904.111799) [78]. Released to the public domain.

Table 1.1: Meningococcal capsular polysaccharide structures. Disease associated serogroups are underlined in bold. Obtained from Tzeng *et al.*, 2016, Crit Rev Microbiol 42(5) (`https://doi.org/10.3109/1040841X.2015.1022507`) [80]. Reused from Taylor & Francis.

Serogroup	Structural Repeating Unit[a]	Acetylation
A	ManNAc-(1-P → 6)- 3 α | OAc	(+)
B	α NeuNAc-(2 → 8)-	(-)
C	NeuNAc-(2 → 9)- | | α 7-OAc 8-OAc	(+)
Y	6-Glc-(1 → 4)- NeuNAc-(2→ 6)- α | α OAc	(+)
W	6-Gal-(1 → 4)- NeuNAc-(2→ 6)- α | α OAc	(+)
X	α GlcNAc-(1-P → 4)-	(-)
E	GalNAc-(1 → 7) – β-KDO - (2→ 3)- β | α 4, 5OAc	(+)
H	α-Gal-(1 → 2)- Gro-(3-P→ 4)-	(-)
I	α-L-GluNAcA-(1 → 3)- β-D-ManNAcA-(1→ 4)- | 4-OAc	(+)
K	β-D-ManNAcA-(1→ 4)-β-D-ManNAcA-(1 → 3)- | 4-OAc	(+)
L	β-GlcNAc-(1 → 3)-β GlcNAc-(1→ 3)- GlcNAc-(1-P-3)-	(-)
Z	α α α-GalNAc-(1 → 1')- Gro-(3'-P → 4)-	(-)

ManNAc, *N*-acetyl-D-mannosamine; NeuNAc, *N*-acetyl-neuraminic acid (sialic acid); GalNAc, *N*-acetyl-galactosamine; GlcNAc, *N*-acetyl-glucosamine; GluNAcA, *N*-acetyl-guluronic acid; ManNAcA, *N*-acetyl-mannosaminuronic acid; Glc, glucosamine; Gal, galactosamine; Gro, glycerol; KDO, 2-keto-3-deoxyoctulosonic acid.

determines whether alternating sialic acid and D-galactose (serogroup W), or sialic acid and D-glucose (serogroup Y) polysaccharides are synthesised [83, 84].

Serogroups A, X and L all contain a roughly 50% similar capsular polymerase, CsaB, CsxA or CslA respectively [85]. CsaB is responsible for the polymerisation of UDP-N-acetyl-D-mannosamine, linked by phosphodiester linkages, with subsequent modifications by other enzymes to give a final structure of α1→6-linked N-acetyl-D-mannosamine-1-phosphate [86, 87]. CsxA functions similarly in the generation of α1→4 phosphodiester linkages in the synthesis of the serogroup X capsule consisting of α1→4-linked N-acetyl-D-glucosamine 1-phosphate [85, 88]. In contrast, CslA was found to make no contribution to the synthesis of the serogroup L structure, consisting of trimeric repeats of →3-β-D-N-Acetylglucosamine-1→3-β-D-N-Acetylglucosamine-1→3-α-D-N-Acetylglucosamine-1-phosphate [89]. Instead, the two-domain polymerase CslB is responsible.

Serogroups H and Z produce teichoic-acid like capsules containing glycerol-3-phosphate residues [90, 91]. Few serogroup H isolates have been isolated from carriage studies. Both serogroups share four homologous genes within region A, the first two of which share 91% amino sequence identity [92]. Serogroups E and Z have been shown to have a degree of cross-reactivity, which initially resulted in serogroup E being referred to as serogroup Z' [93], but they do not share any homologous genes in region A [78]. The serogroup E capsule consists of 2-acetamido-2-deoxy-D-galactose and 3-deoxy-D-manno-octulosonic acid residues.

Serogroups I and K both produce capsules containing N-acetylmannosaminuronic acid, but whereas K consists of repeat O-acetylated homodisaccharides [94], I consists of O-acetylated alternating N-acetylmannosaminuronic acid and N-acetyl-guluronic acid [95]. Region A of these serogroups is highly similar, with two non-synonymous substitutions in the glycosyltransferases CsiC/CskC and CsiD/CskD, which, similarly to serogroups W and Y, may be responsible for the differences in capsule structure [78]. Nevertheless, it has been suggested that both structures should be re-examined

to confirm this difference [78]. Both serogroups are relatively poorly characterised, and rarely isolated from carriage studies.

There is some association of certain clonal complexes with particular capsular groups, for example cc23 and cc167 isolates are typically serogroup Y [96], but meningococci are also able to horizontally exchange their region A genes. This results in capsule serogroup switching among clonal complexes, which has been reported several times among *N. meningitidis* populations [97, 98, 99, 100, 101, 102, 103]. These capsule switching events have great impacts on the epidemiology for IMD. For instance, the hyperinvasive cc11-W "Hajj" strain emerged as an endemic disease-causing strain in the African meningitis belt as a result of C to W capsule switching [104]. Whilst the breakpoints of an HGT event must contain region A for capsule switching to occur, adjacent sequences in the *cps* are also frequently exchanged [103, 97].

Regions B and C of the *cps* are composed of the genes *ctrEF* and *ctrABCD* respectively, and are required for capsule translocation and transport. These regions are well conserved throughout *N. meningitidis*, unlike region A [78], and homologues are also found in other species that express an ABC transporter-dependent capsule [73]. The specific roles of the genes has been deduced based on homology to, and substitution with, ABC transporter-dependent capsule export genes in *E. coli* (Figure 1.6). CtrD is an ATP-binding protein, and together with the transmembrane protein CtrC is believed to be involved in transport across the inner membrane [105]. The auxiliary proteins CtrB and CtrA, located in the inner and outer membranes respectively, and both having a periplasmic domain, are believed to couple transport from the inner membrane through the outer membrane [105]. CtrE and CtrF are β-Kdo-transferases, which attach a lipid moiety to the capsule [106]. Knockouts of these genes have been shown to inhibit surface expression of capsule by *N. meningitidis* [107].

In contrast to many other species that express a capsule, regions A, B and C of the meningococcal *cps* are interrupted by genes without a direct role in capsule

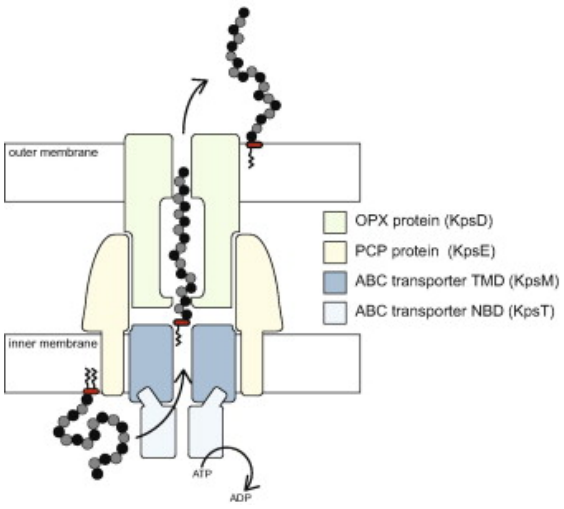

Figure 1.6: ABC transporter-dependent export of polysaccharides. A cartoon depicting a model of the capsule export complex in *E. coli*. KpsT, KpsM, KpsE and KpsD correspond to CtrD, CtrC, CtrB and CtrA respectively. PCP = polysaccharide co-polymerase; OPX = outer membrane protein. Obtained from Willis and Whitfield, 2013, Carbohyd Res 378 (`https://doi.org/10.1016/j.carres.2013.05.007`) [76]. Reused with permission from Elsevier.

synthesis [105]. Region D of the *cps* contains the genes *rfbABC* and *galE*, the latter of which is thought to play a role in lipooligosaccharide (LOS) synthesis [108]. *galE* is also required in the synthesis of serogroup E and Z capsules [109, 91]. A duplication of region D is designated region D' [110], and the *cps* contains two forms of *galE*, one of which is truncated. The truncated and full forms of *galE* in cps^+ meningococci are designated *galE2* and *galE1* respectively. The capsule locus has been shown to re-orientate itself within the genome between these two genes (Figure 1.7) [111]. It has been noted that, since the *cps* is located 54 kb downstream of the origin of replication, it is possible that these inversions resolve collisions between transcription and genome replication machinery [111], as described in *E. coli* [112], which may be important in regions where genes are highly expressed. Region E contains the gene *tex* and two pseudo cytosine methyltransferases of unknown function.

Although regions D, D' and E are not directly involved in capsule synthesis or transport, they are generally considered part of the *cps* due to their location within the locus [78]. There are several meningococcal clonal complexes, including cc198,

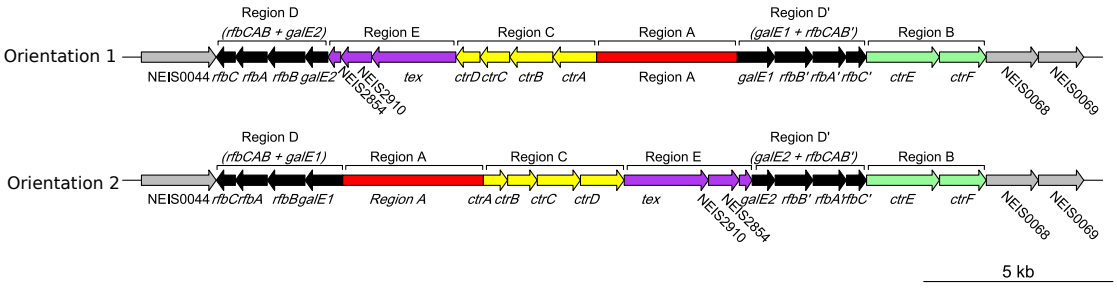

Figure 1.7: Meningococcal *cps* orientations. Organisation of genes within the two orientations of the meningococcal *cps*.

cc53, cc1117 and cc1136, that are consistently found to lack genes required for capsule synthesis, but do contain region D and at least *tex* from region E [52]. Isolates that do not contain regions A, B and C are described as capsule null, and instead possess a distinct 113–118 bp sequence located between regions D and E, the capsule null (*cnl*) locus. The *cnl* locus has also been identified in *N. gonorrhoeae* and the NPN species *N. lactamica* [52], and no encapsulated isolates from these species have been described.

1.3.2 Polysaccharide vaccines

The development of conjugated polysaccharide vaccines has been one of the greatest successes in the control of meningococcal disease. The capsule is poorly immunogenic [113], so whilst vaccines consisting of capsule polysaccharides alone can be effective at controlling outbreaks, such as the the Quebec outbreak of meningitis C disease in the 1980s/1990s [114], long-term immunity is limited, particularly in infants [115]. By covalently bonding polysaccharide to a carrier protein, conjugate vaccines have been shown to induce T-cell responses and result in higher immunogenicity, including in infants [116]. This approach has also been successful against the pneumococcus, and *H. influenzae* serotype b. Conjugate vaccines have been key to the reversal of increasing levels of meningitis C in the UK in the late 1990s [117], and a reduction in cases of meningitis A in the African meningitis belt [35]. Initial evidence suggests that the quadrivalent MenACWY conjugated vaccine may also halt the rise of cc11-W cases [118], and a pentavalent MenACWXY conjugate vaccine is in development

[119]. Therefore, conjugate vaccines may provide a means to control meningococcal disease caused by five of the six disease-associated serogroups.

It has not been possible to exploit the serogroup B capsule to develop a safe and efficacious polysaccharide vaccine, conjugated or otherwise. Like other serogroups, the B capsular polysaccharide has low immunogenicity, but it also mimics a mammalian polysaccharide that is particularly prevalent on embryonic neural cells [120], preventing the development of an effective polysaccharide vaccine, which may present a risk of autoimmunity [115]. Instead, vaccine development against meningitis B has recently focused on sub-capsular antigens. The two licensed and available "MenB" vaccines at the time of writing are: Bexsero®, an outer membrane vesicle (OMV)-based vaccine containing fHbp, *Neisseria* adhesin (NadA), Neisserial heparin-binding antigen (NHBA) and Porin A (PorA) [121]; and Trumenba®, which contains two families of lipidated fHbp [122]. These vaccines rely on the prediction that the most prevalent strains of serogroup B meningococci will have exact matches to these antigens, or cross-reactive matches to fHbp or NadA [123, 124, 125, 122]. Since sub-capsular antigens are not necessarily specific for the target serogroup, these vaccines may also target non-serogroup B meningococci [126], so "MenB" vaccines cannot be said to be specific for disease-associated meningococci in the same way as MenACWY.

1.3.3 Function in virulence

Expression of a polysaccharide is generally considered to be essential for IMD, specifically from one of serogroups A, B, C, W, X or Y [24, 23, 46], since these serogroups facilitate evasion of the immune system. There are a few rare cases where capsule null meningococci [127, 128, 129], or meningococci expressing other serogroups [130, 131], are isolated from disease, but these are the exception. For instance, 98.8% of IMD cases confirmed by culture and/or PCR in England in the epidemiological year 2017-2018 were caused by meningococci expressing serogroup B, C, W, X or Y [132]. Immunocompromised individuals may be more susceptible to

IMD from other serogroups and capsule null meningococci [133], particularly with respect to complement-mediated immune responses.

Capsular serogroups can enhance evasion of complement-mediated immunity. Studies have shown that disease-associated capsule groups are less sensitive to immune serum [134, 130], and serogroup C serum resistance is impeded in capsule knockouts [135]. Serogroup B and C meningococci are able to regulate the alternative complement pathway by enhancing recruitment of the negative regulator factor H to bound C3b, resulting in reduced activation of complement [24, 136, 137]. In this way, meningococci can avoid killing by bacteriolysis and opsonisation-mediated phagocytosis. This enables proliferation of the bacteria in the bloodstream and central nervous system [138].

Encapsulated meningococci are also more capable of surviving intracellularly. Serogroup B meningococci have been shown to be less susceptible to antimicrobial peptides than unencapsulated bacteria, which may enable intracellular survival [139, 140]. Capsule genes, in addition to efflux pumps associated with cationic antimicrobial resistance, were also found to be up-regulated during cell infection [139]. Intracellular survival may be important for the traversal of the nasopharyngeal mucosal barrier and the blood-brain barrier, which is necessary for IMD.

1.3.4 Function in asymptomatic colonisation

As *N. meningitidis* is a principally commensal organism, it would be expected that selection would favour adaptations optimised to establish and maintain colonisation in the nasopharynx, the preferred niche during the commensal life-cycle (Figure 1.8) [141]. Disease does not enhance transmission, but promotes rapid death of the host, so can be regarded as an evolutionary dead end [142, 143] (Figure 1.8D). Any traits that facilitate IMD are expected to be accidental side-effects. Therefore, the capsule is unlikely to primarily function in causing disease [141, 143]. This is consistent with the existence of six capsular groups (E, H, I, K, L and Z) that are not

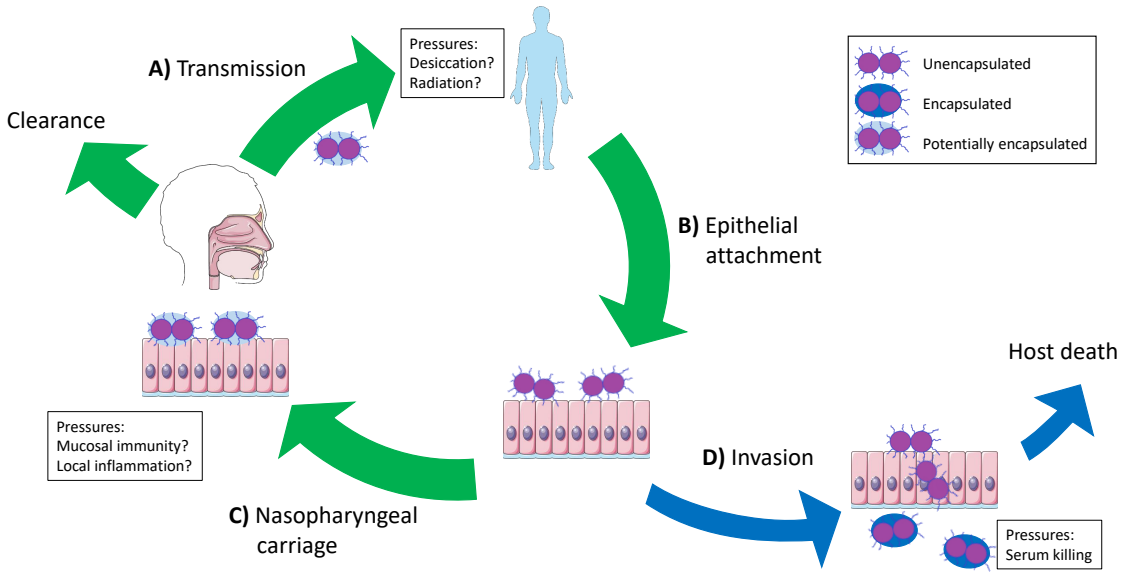

Figure 1.8: A model for capsule function in transmission and disease. The life-cycle of asymptomatic colonisation, carriage and transmission, depicted by the green arrows, is more usual. For A) transmission, the capsule may protect from desiccation or radiation, although carriage risk factor analysis and *in vitro* desiccation resistance data do not support this. During B) attachment to the epithelial layer, the capsule has been demonstrated to switch off. It is not clear whether capsule functions in C) asymptomatic colonisation and carriage in the nasopharynx, but *in vivo* mouse studies and proposals for mucosal immunity and/or local inflammation evasion indicate that it might. In rare instances of D) IMD, capsule expression is essential for survival and proliferation in the bloodstream. Some meningococci do not possess *cps*, so express a capsule at no point in the life-cycle, and are not usually associated with IMD. Image created using art obtained from Servier Medical Art, licensed under a Creative Commons Attribution License.

associated with IMD, and yet have persisted in *N. meningitidis* populations. Capsules have been identified in many free-living bacteria and human symbionts including *Bacteroides* species and non-pathogenic *Streptococcus* species [144, 145, 146, 147]. In both *N. meningitidis* and other species, association with disease is often restricted to a subset of capsular groups or types [6, 148, 149]. The role of the capsule in the asymptomatic life-cycle of *N. meningitidis* is comparatively poorly understood.

1.3.4.1 Function in transmission

A possible function for the capsule in protecting *N. meningitidis* from desiccation or radiation during the transmission stage of the organism's life-cycle (Figure 1.8A) has

been frequently referenced [143, 52, 150, 80, 141, 151]. This was proposed by Virji in 1996, although the speculative nature of this idea was emphasised [152]. The idea is consistent with the frequent occurrence of carriage isolates that are unencapsulated, despite possessing the *cps* locus [46], in contrast to disease-associated isolates, which are almost invariably encapsulated, and could explain the existence of mechanisms that regulate *N. meningitidis* capsule expression, so that it can be switched off during asymptomatic colonisation of the nasopharynx [80, 141]. If the capsule is switched on at all outside of IMD, these data are consistent with its being expressed only during transmission between hosts, when meningococci may be exposed to conditions that make capsule expression favourable.

Polysaccharide capsules have been shown to protect a number of pathogens from desiccation including *Salmonella*, *Acinetobacter* and *E. coli* [75, 153, 154]. Soil bacteria have also been demonstrated to up-regulate extracellular polysaccharide in response to desiccation [155]. The explanation for this desiccation resistance is the capsule's negative charge, which renders it highly hydrated and resistant to water loss [74]. Desiccation resistance is most commonly discussed with respect to environmental persistence and the formation of biofilms [153, 156, 157]. Whether this is an important stage in meningococcal transmission has been investigated both indirectly, through carriage risk factor analyses, and more directly through *in vitro* testing of desiccation resistance.

Carriage risk factor analyses suggest that environmental persistence is not an important phase in the meningococcal life-cycle. These analyses identify behaviours and conditions associated with carriage of meningococci, which may indicate predominant modes of transmission. As early as 1917, overcrowding was identified as a risk factor for transmission in army barracks [44]. More recent carriage studies, which have sampled teenagers or young adults for *N. meningitidis*, whilst collecting behavioural questionnaires, have identified behavioural factors including overcrowding and smoking that increase the likelihood of carriage, and in developed countries at

least, intimate behaviour such as kissing may also be an important factor [151, 47, 158]. The association with kissing and crowding is consistent with an important role for close contact during meningococcal transmission. Desiccation is unlikely to be an important pressure during close contact transmission, so the benefit of capsule expression may be limited, especially if efficiently turning off expression is important for colonisation success. Another analysis showed an association between sharing of food, drink and pacifiers with IMD, for which desiccation resistance may be advantageous, but this effect was not significant when combined with other factors [45]. Nevertheless, it is possible that these studies have missed a component of meningococcal transmission for which desiccation avoidance, and therefore capsule expression, is important, as they study transmission indirectly.

Direct *in vitro* analyses of desiccation resistance have contradicted the commonly held belief that meningococci perish rapidly outside the host. This implies that it may be possible for *N. meningitidis* to spread via fomites, a setting in which factors that increase desiccation resistance would be beneficial. Survival of meningococci on surfaces over periods from hours to as long as ten days has been demonstrated [159, 160, 161, 162], although survival decreases substantially within the first 24 hours, and viability was much lower in *N. meningitidis* compared to *Streptococcus pneumoniae* and *Acinetobacter baumannii*, suggesting fomites may not be an important mode of transmission for *N. meningitidis*, compared to other pathogens [161]. No difference was found in survival between serogroup A, B and C meningococci and their isogenic capsule mutants [161], although some differences between capsule groups have been identified [162]. Therefore, it seems unlikely that capsule would be a beneficial factor if fomites, and therefore desiccation resistance, are an important source of meningococcal transmission.

Without strong evidence from either carriage studies or direct experimental analysis of the effect of encapsulation on desiccation resistance, the idea that capsule has developed as a means to protect *N. meningitidis* from desiccation remains

speculative. It is possible that the capsule plays some unknown function during transmission, perhaps during shedding of *N. meningitidis* from the nasopharynx at a later stage in infection, but more promising insights have been gained through investigating the role of capsule in maintaining colonisation.

1.3.4.2 Function in colonisation and persistence

The alternative to a function in transmission is that the capsule has some function in colonisation and persistence in the nasopharynx (Figure 1.8C). Although meningococci isolated from carriage are more frequently unencapsulated [46], they are not exclusively so. Carriage isolates provide only a snapshot of the genetic diversity of meningococcal colonisation of the nasopharynx, and rarely account for length of colonisation, and other environmental conditions that might affect capsule expression. Further *in vitro* evidence for expression of the capsule during this stage of the lifecycle is conflicting.

Evidence from *ex situ* tissue culture experiments is generally consistent with a lack of function for the capsule during asymptomatic colonisation. The capsule has been found to be down-regulated during adhesion to epithelial cells [163, 164], and capsule expression results in reduced attachment and invasion of human airway epithelial cells, possibly attributed to increased hydrophobicity [165]. Encapsulated *N. meningitidis* has also been demonstrated to perform poorly during biofilm formation *in vitro* compared to unencapsulated meningococci [150, 166]. These studies indicate that expression of the capsule is a hindrance to colonisation, at least during attachment to the epithelial layer (Figure 1.8B). Nevertheless, another study has shown no effect of encapsulation status on *N. meningitidis* biofilm growth on human epithelial cells [167]. These differences may be attributed to differences in the experimental system. *Ex situ* models do not currently account for the complex environment of the nasopharynx, which is occupied by a multitude of different microorganisms, and

it is difficult to determine whether capsule is inhibitory during *in situ* colonisation of the human nasopharynx.

Immune protection in the nasopharynx

Real-world efficacy of capsular vaccines against meningococcal carriage indicates that capsule expression does occur at some stage during asymptomatic colonisation. Carriage studies in UK teenagers carried out during and after the implementation of the conjugate MenC vaccine at the turn of the 20$^{\text{th}}$ Century indicated that there had been a substantial reduction in carriage of serogroup C *N. meningitidis*, as well as disease [168]. This has likely increased the efficacy of the vaccination campaign, since carried meningococci serve as the key source of transmission [46]. Although the exact mechanism of protection is not understood, it must surely rely on expression of the capsule at some stage during nasopharyngeal carriage. Expression of the capsule during carriage also provides an explanation for the sophisticated host mimicry of the serogroup B capsule, which has made development of a serogroup B capsular polysaccharide vaccine difficult [169, 120]. General resistance to immunity could be explained as an inherent property of polysaccharide capsules, but it is unlikely that *N. meningitidis* would benefit from a very sophisticated immune evasion mechanism if the capsule is only expressed during transmission. These observations cannot directly support a functional role for capsule during nasopharyngeal colonisation, but they are certainly consistent with capsule being expressed during asymptomatic colonisation.

Analysis of *in vivo* mouse models of carriage are consistent with a benefit from capsule expression during nasopharyngeal colonisation in some meningococcal strains. A model established by supplementing mice with iron demonstrated that capsule deficiency impedes nasopharyngeal colonisation [170], as did a study of intranasal colonisation in infant mice [171]. Similarly, a model established using CEACAM1-humanised mice showed that capsule deficiency either reduced or had no effect on nasopharyngeal colonisation [172]. The first model demonstrated that the difference

could be partially explained by complement, but the second found that complement deficiency only increased meningococcal colonisation if neutrophils were also depleted, indicating that differences can also arise between different mouse models. Nevertheless, these models indicated that mucosal immunity may be a pressure experienced during colonisation of the nasopharynx, and *N. meningitidis* may be exposed to a variety of immunological responses in the nasopharynx [173], against which capsule expression might be protective.

Local inflammation hypothesis

Molecular evidence consistent with a function for the capsule in protecting *N. meningitidis* from mucosal immunity comes from the identification of a temperature-dependent regulatory mechanism of capsule expression. An RNA thermosensor has been identified in the 5' UTR of *cssA*, a capsule synthesis gene present in the *cps* encoding serogroups B, C, W and Y [174]. At low temperatures, the thermosensor motif causes *cssA* RNA to form a hairpin structure, typical of RNA thermsosensors [175], which prevents the ribosome from binding to the ribosome binding site (RBS), inhibiting translation (Figure 1.9). As expected for a RNA thermosensor, this hairpin was found to open up with increasing temperatures [176], facilitating access to the RBS and increasing translation. Deletion of the proposed RNA thermosensor, giving free access to the RBS, resulted in an increase in levels of CssA protein, as well as increased serum resistance, consistent with higher expression of the capsule [174]. This regulatory mechanism is likely important in the transition to meningococcal disease, since the temperatures within the bloodstream would be higher than the nasopharynx. There may also be transient pressures during asymptomatic colonisation, for which expression of a capsule, regulated by a thermosensor, may be beneficial [177].

In the nasopharynx, an increased temperature is associated with inflammation, and is an indication of infection [178]. Since *N. meningitidis* shares the nasopharynx with numerous other bacteria and viruses, including common respiratory viruses such

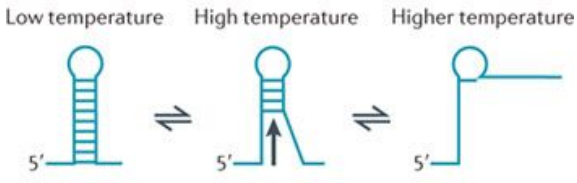

Figure 1.9: RNA thermosensor mechanism of action. Hairpin structure prevents ribosome binding at low temperatures. Obtained from Kortmann and Narberhaus, 2012, Nat Rev Microbiol 10(4) (`https://doi.org/10.1038/nrmicro2730`) [175]. Reused with permission from Springer Nature.

as the rhinoviruses and influenza, it is proposed that induction of inflammation by co-infecting pathogens exposes commensal *N. meningitidis* to the same inflammatory responses [177]. This may induce temperature-dependent up-regulation of capsule expression, consistent with the capsule having the primary purpose of protecting *N. meningitidis* from local inflammation within the nasopharynx. In this case, it would be expected that exposure to co-infecting pathogens would change the dynamics of meningococcal carriage and the levels of capsule expression, possibly resulting in changes in the rates of transition to meningococcal disease.

Both challenge studies and observational studies have investigated the effects of respiratory virus co-infection on IMD, although equivalent investigations into carriage are lacking. Strongest associations have been described for influenza, with associations with disease described in observational [179, 180], case-control [181], and time-series [182, 183, 184] analyses. Another time-series analysis found no association between influenza and meningococcal or other bacterial infections [185]. All time-series were adjusted for seasonality. Other respiratory tract infections have been less well examined, with no associations found for Rhinovirus, adenoviruses or respiratory syncytial virus [186, 187]. A time-series analysis in Burkina Faso found an association between undefined upper-, but not lower-, respiratory tract infections and IMD [188]. Overall, there is an emerging body of evidence supporting an association between influenza and IMD, with insufficient evidence for other respiratory tract infections. Therefore, local inflammation may impact on *N. meningitidis* behaviour

in the nasopharynx, including capsule expression, in some instances. Nevertheless, there may be several alternative explanations for the association between influenza and IMD, including damage to the upper respiratory tract, changes to the immune system [183], and interactions between the capsule and influenza neuraminidase [189]. Investigating the association between encapsulation of carriage isolates and previous respiratory tract infection may be more informative as to whether a key function of the capsule is to protect meningococci from local inflammation in the nasopharynx. Smokers and non-smokers may be of particular interest, since smoking can induce inflammatory responses, and has been associated with capsule expression in commensal biofilms [190].

The precise function of the polysaccharide capsule in the asymptomatic *N. meningitidis* lifecycle remains enigmatic. There is little molecular evidence to support the assumption that the capsule protects the meningococcus during transmission, and data from epidemiological studies are not conclusive. Hypotheses consistent with a protective function in the nasopharynx are only in their infancy, although these conflict with *ex situ* cellular studies that indicate a potential disadvantage during epithelial colonisation. Nevertheless, *in vivo* experiments and carriage studies indicate that meningococci can and do express a capsule during asymptomatic carriage.

1.3.5 Capsule evolution in *N. meningitidis*

In spite of the likely primarily non-virulent function of the polysaccharide capsule, it has nevertheless not been previously described in NPN species. Whilst a number of putative virulence-associated genes have been found in NPN species [65], the capsule has been considered to be unique to the meningococcus [64]. The presence of a capsule in *N. gonorrhoeae* was a source of controversy in the 1970s and 1980s as a result of conflicting molecular analyses [191], but it has since been established that *N. gonorrhoeae* does not express a capsule, and genomic analyses have revealed the absence of regions A, B and C in the equivalent genome position [110]. Since

then, no capsule genes have been identified in the WGS data of any of the thousands of *N. gonorrhoeae* isolates in PubMLST. Region C genes *ctrABCD* have previously been identified in *N. subflava*, but based on the presence of a DUS, were assumed to have been acquired by HGT [64]. Capsule genes have not been identified in any other *Neisseria* species, but have been characterised in another member of the *Neisseriacae* family, *Kingella kingae* [192]. It is therefore possible that the meningococcal capsule was acquired in an HGT event that gave rise to the potentially pathogenic variants of *N. meningitidis*.

It has been argued that the low G+C content of region A sequences in all capsule serogroups, in many cases below 40%, compared to 50% in the meningococcal core genome, is evidence of an HGT event [78, 83, 193]. Different bacterial species have different core genome G+C content, which is believed to result from codon bias optimisation across the genome [194]; elements acquired by HGT that have a different G+C content are predicted to eventually optimise to the core genome level. G+C content lower or higher than the core genome is a typical feature of pathogenicity islands, which are often involved in HGT events [195], but many of the other typical features, including repeated DNA flanking the island and association with tRNA genes, are not present in the meningococcal *cps* [143]. The low G+C content in region A may be a misleading feature, since low G+C content has also been reported in the capsule synthesis regions of *E. coli*, *Haemophilus sputorum* and *A. pleuropneumoniae* [196, 197, 198]. There is also a much less pronounced decrease in G+C content in the rest of the *cps*. Whilst it is certainly an interesting observation, and may be informative about the primordial, common ancestral history of capsule genes among Gram-negative bacteria, it may be less informative when it comes to ascertaining the evolutionary history of the capsule within the meningococcus.

Nevertheless, the alleged absence of capsules in other *Neisseria* species has left an HGT acquisition event as the most parsimonious explanation for the existence of the meningococcal capsule. This initial capsule acquisition is assumed to have taken

place after the divergence of *N. meningitidis* and *N. gonorrhoeae* [143, 199, 193]. Based on similarity to the *cps* in *Enterobacteriacae* and *Haemophilus* [73], it was proposed that region C and region A may have been co-acquired, which seems more plausible if G+C content is not a key feature of this HGT event, with a subsequent acquisition of region B during a recombination event that also duplicated region D [143]. This comparative genomics approach argued in favour of a *de novo* acquisition of the capsule in meningococci, but did not provide strong evidence for any particular mechanism or donor.

The most plausible donor must be an organism that also inhabits the nasopharynx, creating the opportunity for co-habitation and horizontal exchange. In 2008, the capsule transport genes *hexDCBA* and *phyAB* of *P.multocida* – a member of the *Pasteurellacae* family found in the nasopharynx of livestock – were found to have the highest amino acid sequence identity to *ctrABCD* and *ctrEF* respectively among genes present in the STRING database at the time [72]. The values of this sequence identity ranged from a fairly modest 50.1-74.9%. If an acquisition event did occur after the split between *N. meningitidis* and *N. gonorrhoeae* – or even between those two species and *N. polysaccharea*, taking into consideration suggestions that *N. gonorrhoeae* was an unencapsulated meningococcal strain that branched out into a new niche [200] – then given the low core genome divergence among these organisms, it is unlikely that the meningococcal capsule would have diverged so greatly from its donor. Nevertheless, it was suggested that the true donor may be another member of the *Pasteurellacae* family [72]. This was not an unreasonable claim, given that HGT of other loci between *N. meningitidis* and *H. influenzae*, another member of the *Pasteurellacae*, had been described [201]. Homology between region A genes and capsule synthesis genes in members of the *Pasteurellacae* and other species has also been observed [202, 78]. With NGS only just being introduced at the time, there was little sequence data to be searched, and a more plausible donor was yet to be found, and still there was no evidence supporting a specific mechanism.

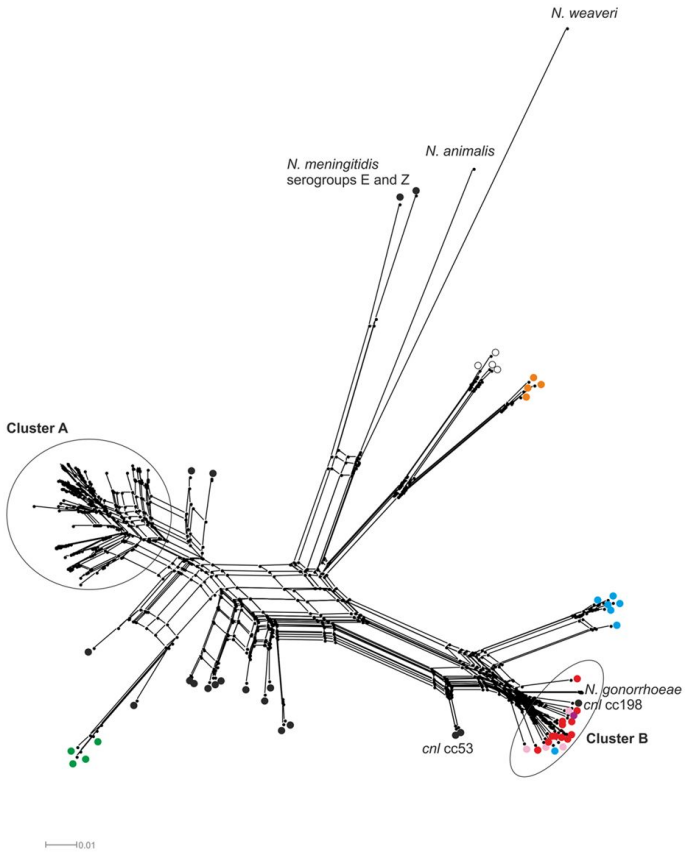

Figure 1.10: "Phylogenetic reconstruction using *galE* alleles from commensal and pathogenic *Neisseria* spp. using an unrooted neighbor-net algorithm. *N. gonorrhoeae* and commensal species are represented by circles with different colours denoting species. These were: *N. meningitidis* is black; *N. lactamica* is red; *N. polysaccharea* is pink; *'N. bergeri'* is mauve; *N. cinerea* is blue; *N. elongata* is orange; *N. animalis* and *N. weaveri* are grey; *N. mucosa* is white and *N. subflava* is green. *N. meningitidis* isolates are found in cluster A but also on branches in-between clusters A and B as well as the *cnl* in cluster B. *N. gonorrhoeae* are only found in one branch of cluster B and are labelled as *N. gonorrhoeae* without a symbol." Obtained from Bartley *et al.*, 2017, Sci Rep 7(44442) (`https://doi.org/10.1038/srep44442`) [111]. Licensed under a Creative Commons Attribution License.

One mechanism has since been put forward by Bartley *et al.* [111], describing acquisition of the capsule in a single *en bloc* HGT event, in which the site of capture was *galE2*. The homologues *galE1* and *galE2* were shown to be divergent, with *galE2* more closely related to capsule null *galE* [111] (Figure 1.10). In the same study, it was also demonstrated that *galE* is bi-functional, synthesising UDP-galactose and UDP-galactosamine. The homologue *galE1* was determined to be predominantly mono-functional, producing only UDP-galactose, but bi-functional in capsule groups Z

and E, which require this bi-functionality for capsule synthesis. The sequence of *galE1* in these serogroups also clustered with neither *galE* nor other *galE1* sequences. Bartley *et al.* proposed that *cps* regions ←*tex*←region C-region A→*galE1*→*rfbCAB'*→region B→NEIS0068→ were donated *en bloc* to a capsule null meningococcus, entailing homologous recombination at NEIS0069 and an illegitimate breakpoint in *galE*, which resulted in the truncated *galE2* (Figure 1.11). This HGT event would have initially resulted in *cps* organisation as seen in Orientation 1 (Figure 1.7).

Whilst this hypothesis provides an explanation for the phylogenetic relationships seen between *galE1*, *galE2* and *galE*, the pseudo cytosine methyltransferases downstream of *tex* are excluded from this model, and no evidence for an illegitimate breakpoint was put forward. A more compelling mechanism would be if region E had already inverted in the capsule null recipient, enabling simple homologous recombination within *tex*. Bartley *et al.* also found that *rfbABC* and *rfbABC'* were almost identical in MC58, FAM18 and NMB, which is inconsistent with the divergence between *galE2* and *galE1*, unless these genes are highly conserved between species [111]. This study focused primarily on the functionality of *galE*, and its implications for capsule acquisition, but further investigation into the rest of region D, as well as regions B and C, might provide further evidence for or against this *en bloc* acquisition hypothesis.

1.4 Outline and aims of this thesis

The polysaccharide capsule is an important genetic factor in *N. meningitidis*. Whether or not a capsule is expressed, and if so its biochemical structure, plays a crucial role in determining the potential for meningococci to cause IMD [23, 24]. Its apparent absence from all other *Neisseria* species has resulted in much discussion about the possible origins of the capsule in *N. meningitidis*, an event which has had an important impact on meningococcal epidemiology from a biomedical perspective [143,

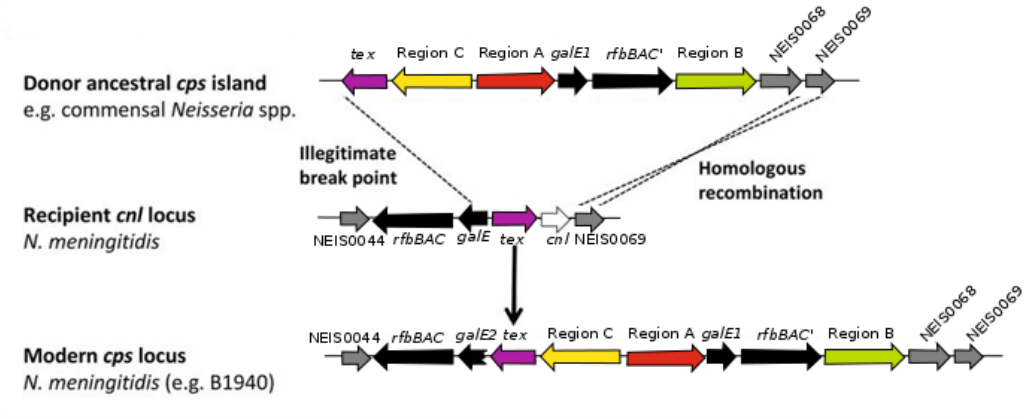

Figure 1.11: An *en bloc* acquisition model of the meningococcal *cps*. Adapted from Bartley *et al.*, 2017, Sci Rep 7(44442) (`https://doi.org/10.1038/srep44442`) [111]. Licensed under a Creative Commons Attribution License.

72, 111]. This thesis uses genomic analyses of the polysaccharide capsule to further understand its distribution among non-pathogenic bacteria, its evolutionary history in *N. meningitidis*, and how contemporary dynamics of capsule evolution impact IMD.

Chapter 2 challenges the notion that, among *Neisseria*, capsule genes are limited to *N. meningitidis*. The capsule is necessary for IMD, but *N. meningitidis* is principally a commensal organism, and it is likely that the capsule primarily functions in asymptomatic colonisation and/or transmission [143, 52, 141]. It has even been suggested that capsules are very common within the human microbiota, and relatively rare among pathogens [203], so it is perhaps unusual that capsules have not been identified in NPN species. Taking advantage of growing amounts of WGS data from NPN species, and using the conserved nature of ABC transporter-dependent capsule export sequences, the aim of Chapter 2 is to determine whether capsule genes are present in NPN species. Where capsule genes are present, the aim is to characterise NPN capsule loci and make initial predictions about possible biochemical structures.

The presence of capsule genes in NPN species challenges hypotheses postulating that the meningococcal capsule was acquired by HGT [143, 72, 111], which will be investigated in Chapter 3. The aim of Chapter 3 is to analyse recombinant sequences within the meningococcal *cps*, and evaluate hypotheses that it was acquired *en bloc*

by a capsule null organism. Using a diverse dataset, some of the challenges to meeting this objective, including intra-locus and intra-species recombination [111, 103], will be overcome to develop on previous understanding of capsule origins in *N. meningitidis*.

Finally, in Chapter 4 a more contemporary example of HGT within *N. meningitidis* will be investigated. The aim of Chapter 4 is to analyse a capsule switching event resulting in serogroup B IMD. This is an epidemiological example demonstrating how capsule serogroup can modify pathogenic potential in *N. meningitidis*. The potential future impact of such cases will be evaluated in the context of protein-based "MenB" vaccines, and possible insights into the relationship between clonal complex, capsule and invasiveness will be explored.

In Chapter 5, the major conclusions of the thesis will be summarised. Links will be drawn between the chapters, discussing the distribution of capsule in *Neisseria*, and the historic and contemporary dynamics of capsule evolution and its potential impact on the epidemiology of IMD. Whilst the biomedical importance of the capsule will be addressed, the capsule will be discussed in its context as one of many disease-associated gene products in a principally commensal organism.

2

Characterisation of putative capsules in non-pathogenic *Neisseria*

Some of this work was published in Clemence *et al.*, 2018, Microbial Genomics 4(9):e000208 (`https://doi.org/10.1099/mgen.0.000208`) [204], and presented as a talk at the 14th Congress of the European Meningococcal and Haemophilus Disease Society, Prague, 2017. Genomics work presented here was carried out by the candidate. Biochemical analyses were carried out in collaboration with colleagues at the National Institute for Biological Standards and Control (NIBSC). Contributions by collaborators are indicated.

2.1 Introduction

The genus *Neisseria* is a diverse group of bacteria, many of which are asymptomatic colonisers of the mucosal surfaces of animals and humans [1]. Whilst many *Neisseria* species belonging to the human oral microbiota have been identified, research has focused on species associated with disease: *N. meningitidis* and *N. gonorrhoeae*. In common with many other *Neisseria* species, *N. meningitidis* usually colonises the nasopharynx asymptomatically; it is only on rare occasions that it invades the bloodstream and causes IMD [6]. In contrast, there are very few case reports

of other *Neisseria* species causing invasive disease [8, 9, 10, 11], and so they can be described as NPN species.

In *N. meningitidis*, association with IMD is generally restricted to organisms that express a capsule from one of serogroups A, B, C, W, X or Y [6]. The capsule is therefore considered to be a virulence factor [205]. Whilst a number of putative virulence genes have been found in NPN species [65], the capsule has been considered to be unique to the meningococcus [64], possibly acquired in an HGT event that gave rise to the potentially pathogenic variants of *N. meningitidis* [143, 72, 111]. Nevertheless, a further six meningococcal capsule serogroups E, H, I, K, L and Z are not associated with disease, and capsules have been described in several commensal bacterial species [147, 146, 203, 145], so it would not be unprecedented if capsule genes were also present in NPN species.

The meningococcal capsule is produced *via* ABC transporter-dependent polymerisation [79]. The structurally distinct capsule serogroups are encoded by different complements of capsule synthesis proteins in region A of the *cps* [76, 78]. In contrast, the proteins involved in capsule export and surface expression are conserved among Gram-negative species, and in *N. meningitidis* are encoded by regions B and C of the *cps* [206]. Among *N. meningitidis*, the region B genes *ctrEF* and region C genes *ctrABCD* are very highly conserved, even between serogroups [78]. Therefore, it could be expected that if any NPN species express an ABC transporter-dependent capsule, homologues of regions B and C would be present in their genomes.

At the time of analysis, WGS data sequenced from a total of 20 species had been defined in the PubMLST database, using the universal and high-resolution rMLST approach [14]. These species included the human commensals *N. polysaccharea* (16 isolates), *'N. bergeri'* (one isolate), *N. lactamica* (140 isolates), *N. cinerea* (15 isolates), *N. subflava* (19 isolates), *N. oralis* (four isolates), *N. mucosa* (10 isolates), *N. elongata* subsp. (four isolates), *N. bacilliformis* (four isolates), and the pathogens *N. meningitidis* and *N. gonorrhoeae*, along with several species isolated

from animals or animal bite wounds. Although there have been case reports of other *Neisseria* species acting as opportunistic pathogens, these have usually been isolated incidences, unlike IMD [8, 9, 10, 11].

The aim of this chapter is to identify and characterise capsule genes in NPN species.

2.2 Methods

2.2.1 Genetic characterisation of *cps* in non-pathogenic *Neisseria*[1]

2.2.1.1 Isolate collection

WGS data from *Neisseria* isolates were obtained from the PubMLST database. Public NPN isolates with a defined species designation, and possessing WGS data, were surveyed for the presence of putative capsule genes. All isolates found to contain a putative *cps* were further characterised and annotated. Closed reference *Neisseria* genomes were obtained from either GenBank (`http://www.ncbi.nlm.nih.gov/genbank/index.html`) or the National Collection of Type Cultures NCTC 3000 project (`http://www.sanger.ac.uk/resources/downloads/bacteria/nctc/`). *N. meningitidis* region A sequences were sourced from WGS in PubMLST belonging to 0106/93 (serogroup L) and 0084/93 (serogroup Z), and the published *cps* of 29043 (serogroup I, GenBank accession number HF562984.1). Non-*Neisseria* capsule synthesis genes were acquired from GenBank including: *A. pleuropneumoniae* strains 14-022 (serovar 7, GenBank accession number MG780422.1), 405 (serovar 8, KJ685493.1), Femo (serovar 6, AY534316.1), 3906 (serovar 14, MG868948.1), QAS106 (serovar 12, AB701757.1) and J45 (serovar 5, AF053723.1); *H. influenzae* strains M6542 (serotype c, HM770876.1), ATCC 9008 (serotype d, HQ424464.1) and ATCC 8142 (serotype

[1] This work was published in Clemence *et al.* 2018 [204]. Work presented here was carried out by the candidate.

e, HM053635.1); and *Haemophilus parainfluenzae* strain HUB12445 (MH644108.1). A full list of isolates can be found in Table A.1.

2.2.1.2 Annotation of putative capsule genes in NPN

Meningococcal region B and C *cps* genes *ctrABCDEF* had previously been defined in the PubMLST sequence database as NEIS0055, NEIS0056, NEIS0057, NEIS0058, NEIS0066 and NEIS0067 respectively. For each gene, the deduced amino acid sequence of allele 1 was queried against isolate WGS data, using the BLASTp algorithm within the PubMLST interface. If a credible match was found, the relevant contig was downloaded and the sequence of interest annotated in Artemis [207]. G+C content was determined in MEGA7 [208]. The same approach was also used to annotate NPN isolates with region D and E genes, and any other relevant genes, where necessary. Annotations were uploaded as novel alleles in PubMLST.

The proposed *cps* regions of NPN isolates were further annotated in Artemis and visualised using genoPlotR. Undefined open reading frames (ORFs) adjacent to putative Region C genes were queried against the National Center for Biotechnology Information (NCBI) nucleotide collection using tBLASTn [209], and against the Pfam A database using BLASTp [210]. Where only hypothetical matches were found, sequences were also queried against the NCBI RefSeq database using BLASTp. Support for putative region A genes was based on homology to capsule synthesis genes from *N. meningitidis* and/or other bacterial species, or at least for gene products consistent with a function in capsule synthesis, such as glycosyltransferases. Additional guidance was based on comparisons to *cps* organisation within *N. meningitidis*. Putative region A genes were uploaded as novel NEIS loci and alleles in PubMLST.

2.2.1.3 Comparison of putative capsule synthesis genes among and between species

The putative region A sequences within genomes from isolates belonging to the same species were compared to determine whether they were equivalent. Pairwise BLASTn

comparisons of putative region A sequences were generated using the blastall tool and visualised using genoPlotR. The nucleotide sequences of homologous genes were aligned in the Emboss [211] Clustal Omega tool to quantify nucleotide identity.

Potentially homologous region A genes shared by different species were identified based on gene order, sequence length and predicted function. Pairwise tBLASTx comparisons were generated using the blastall tool or WebACT [212], and visualised using genoPlotR. The amino acid sequences of homologous genes were aligned in the Emboss Clustal Omega or Needle tool to quantify amino acid identity. *N. meningitidis* and non-*Neisseria* species containing homologues of putative NPN capsule synthesis species were chosen based on the results of NCBI BLAST results. Pairwise tBLASTx comparisons as above were used to compare *Neisseria* region A sequences to characterised capsule synthesis sequences from these species.

2.2.1.4 Phylogenetic analysis

A recombination-corrected phylogenetic tree of *Neisseria* isolates, with *Moraxella catarrhalis* strain RH4 as an outgroup, was generated based on rMLST loci [62]. The nucleotide sequences of 51 of 53 ribosomal protein subunits used in the rMLST scheme (excluding *rpmE* and *rpmJ*, as they are paralogous in some *Neisseria*), were extracted from each isolate using the BIGSdb genome comparator tool, and aligned using MAFFT [213]. A maximum-likelihood tree was generated in PhyML v3.1 [214] using the GTR+I+G substitution model, determined to be the best-fit model by jModelTest v2.1.10 [215], with 100 bootstrap replicates. The tree was corrected for recombination in ClonalFrameML [216], and rendered and annotated using the ETE3 toolkit [217]. The phylogeny included all isolates belonging to those species in which capsule genes were identified, representatives of species in which capsule genes were not identified including *N. cinerea*, *N. lactamica*, *'N. bergeri'*, *N. polysaccharea* and *N. gonorrhoeae*, and representatives of *N. meningitidis*.

2.2.2 Preliminary extraction of polysaccharides for NMR analysis[2]

Preliminary work was carried out to extract and purify capsule for NMR analysis from isolates CCUG 2043T (*N. elongata* subsp. *elongata*), CCUG 30802T (*N. elongata* subsp. *nitroreducens*), CCUG 24918 (*N. subflava*) and 12007_2012 (*N. subflava*). Isolates were grown on both 10X Mueller Hinton Blood Agar plates and in 500 mL of Mueller Hinton broth at 37°C/5% CO_2 for 16 hours.

Following incubation at 60°C to kill live bacteria, pooled cultures were boiled in a water bath for three hours to release cell surface-associated polysaccharide capsule, then centrifuged at 8000 x *g* for 60 minutes to remove cell debris. The supernatant was concentrated to a volume of 150 mL and underwent diafiltration with normal saline using the VivaFlow 30 kDa system to remove molecules smaller than 30 kDa. The retentate was ultra-centrifuged at 200,000 x *g* for one hour to remove outer membrane vesicles. The supernatant was made up to 25% ethanol and incubated at 4°C overnight to precipitate out some DNA and proteins. Samples were centrifuged at 30,000 x *g* for 60 minutes, and the supernatant was made up to 75% ethanol and left at 4°C overnight to precipitate out crude polysaccharide capsule. The supernatant was removed, and the pellet retrieved by centrifuging at 3500 x *g* for five minutes. The pellet was dried, then re-suspended in 5 mL of DNase reaction buffer (100 mM Tris HCl [pH 7.5], 25 mM $MgCl_2$, and 5 mM $CaCl_2$) and the solution underwent further DNase and protease treatment at 37°C overnight, followed by dialysis with SnakeSkin 3.5K MW in deionised water.

A method based on anion exchange chromatography was used to further purify capsule polysaccharides [218]. Anion exchange buffer was added to samples for a final concentration of 20 mM Tris/2 mM ethylenediaminetetraacetic acid (EDTA)/0.3% sodium deoxycholate (pH 8). The purpose of the detergent in the buffer was to

[2] This work was carried out in collaboration with Dr Sunil Maharjan at NIBSC. The extraction method had previously been developed by Dr Maharjan. LAL assays were carried out by Dr George Kemp. NMR was performed by the NMR facility at the NIBSC, led by Dr Tim Rudd.

discourage the formation of bivalent cations, micelles and vesicles of LOS. Samples were run through CaptoAdhere–CaptoDEAE columns in tandem, equilibrated with anion exchange buffer at 1 mL/minute, taking 60X 1 mL fractions, to separate the capsule from remaining proteins and LOS. 50 μL of each fraction was added to 50 μL of anthrone solution (2 mg anthrone/1 mL concentrated sulphuric acid) and boiled for ten minutes. Fractions that reacted with anthrone, resulting in a colour change from yellow to blue, were dialysed with SnakeSkin 3.5K MW in deionised water and tested using a limulus amebocyte lysate (LAL) assay to rule out contamination with LOS. Samples were then dried, and sent for NMR analysis.

2.3 Results

2.3.1 At least 13 NPN species possess putative capsule loci

Putative capsule loci were identified in genome sequences from a total of 13 *Neisseria* species contained within the PubMLST database, including *N. bacilliformis*, (four of four isolates); *N. elongata* subsp. (two of four isolates); *N. musculi* (one isolate); *Neisseria dentiae* (one isolate); *N. animaloris* (one isolate); *N. zoodegmatis* (one isolate); *N. weaveri* (one isolate); *N. canis* (one isolate); *Neisseria wadsworthii* (one isolate); *N. animalis* (one isolate); *N. oralis* (four of four isolates); *N. mucosa* (one of nine isolates); and *N. subflava* (seventeen of nineteen isolates) (Figure 2.1). *N. wadsworthii* and *N. animaloris* were not annotated in detail, since one or more genes within the predicted capsule locus were interrupted in the genome assembly.

None of the NPN capsule loci were organised equivalently to *N. meningitidis* (Figure 2.2): all NPN lacked the duplicated region D', and did not contain orthologues of the pseudo cytosine methyltransferases found in meningococcal region E. *N. subflava* was the only NPN species in which the putative regions A, B, C and D were contiguous, although the putative regions A, B and C were contiguous or nearly contiguous in all NPN species apart from *N. oralis*, *N. animalis*, *N. canis*

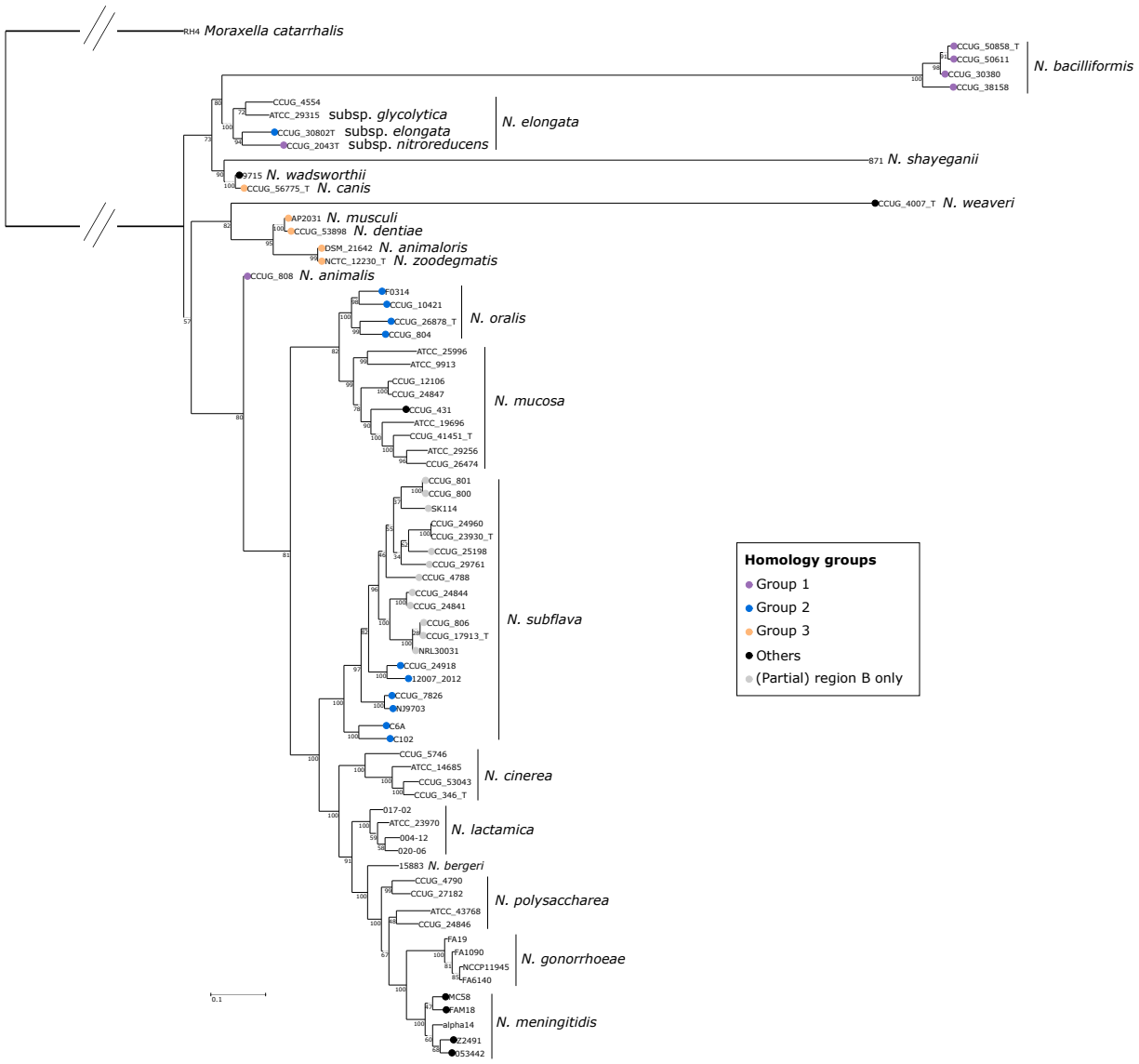

Figure 2.1: Distribution of putative capsule genes across the *Neisseria*. A maximum-likelihood phylogeny generated from an alignment of concatenated rMLST nucleotide sequences from *Neisseria*, with *M. catarrhalis* as an outgroup. Branch supports based on 100 bootstrap replicates. Corrected for recombination in ClonalFrameML. Coloured circles indicate the presence of putative capsule genes in WGS data. Isolates sharing homology in putative region A sequences, as shown in Figure 2.4, are filled with the same colour. Other isolates are black. Grey circles indicate partial putative region B genes identified in WGS data only.

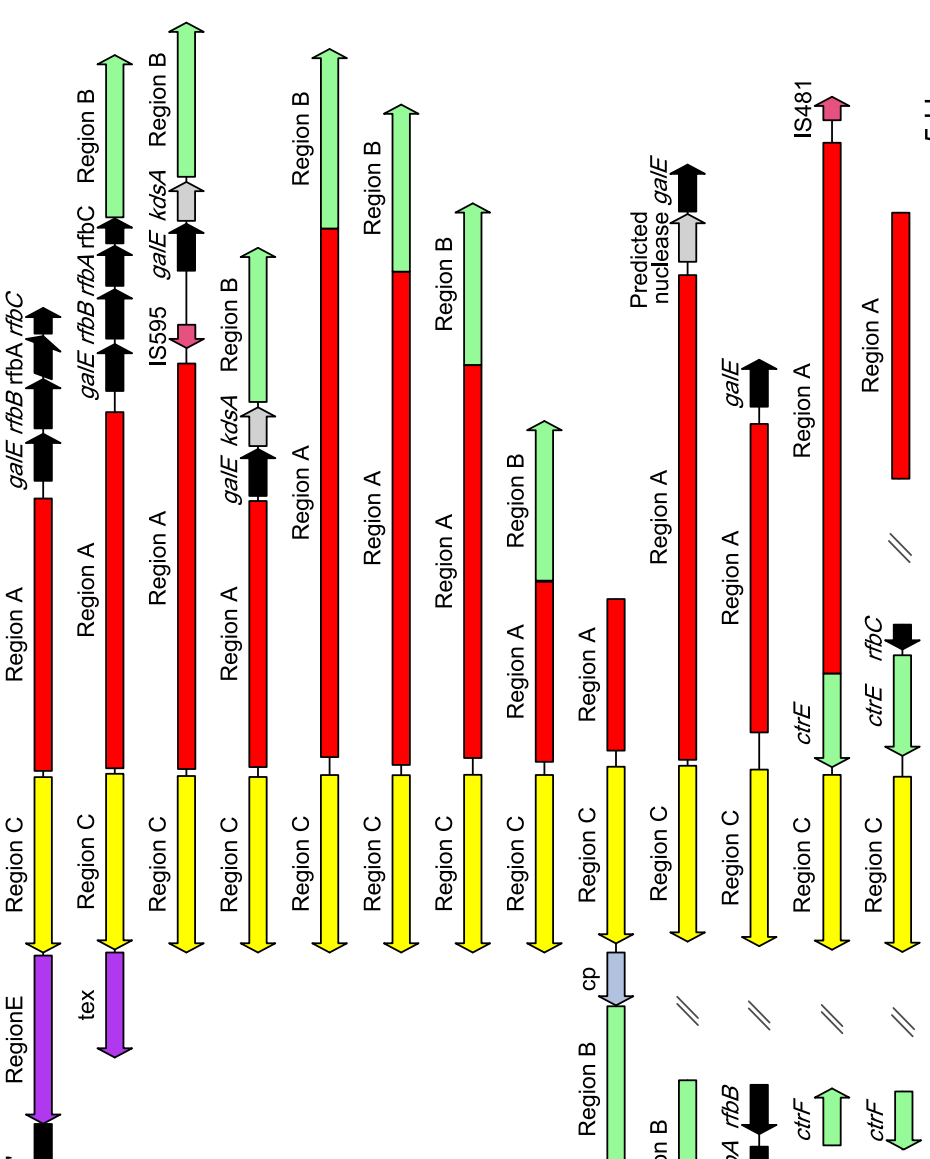

Figure 2.2: Organisation of putative capsule loci in NPN species. Region A is involved in capsule synthesis, regions B and C in capsule export. *N. animaloris* and *N. wadsworthii* were excluded, since region A was interrupted in the genome assembly for the isolates from these species. Two diagonal lines represent >5 kb between genes. cp refers to capsule phosphotransferase as seen in some W, I, and K isolates of *N. meningitidis* [78].

Table 2.1: Identification of region B and C orthologues. Annotations based on the results of queries using BLASTp against the PubMLST *Neisseria* sequence database. A-F corresponds to results of BLASTp queries of *ctrA-F* allele 1. Greyed out isolates in E and F did not contain all six genes.

A

Isolate	Species	aa ID(%)	Mismatches	Gaps	E-value	Query cover	G+C(%)
CCUG 808	*N. animalis*	79.51	75	1	0	0.95	46.68
DSM 21642	*N. animaloris*	55.44	170	2	5.00E-142	0.99	47.44
CCUG 50858 T	*N. bacilliformis*	76.38	88	2	0	0.97	49.61
CCUG 30380	*N. bacilliformis*	74.74	97	2	0	1.00	49.61
CCUG 38158	*N. bacilliformis*	74.74	97	2	0	1.00	49.70
CCUG 50611	*N. bacilliformis*	75.00	96	2	0	1.00	49.61
CCUG 56775 T	*N. canis*	69.35	113	1	7.00E-172	0.95	48.06
CCUG 53898	*N. dentiae*	56.96	163	3	2.00E-144	0.99	46.68
CCUG 30802T	*N. elongata* subsp. *elongata*	73.32	97	3	0	1.03	48.87
CCUG 2043T	*N. elongata* subsp. *nitroreducens*	73.52	102	1	0	0.99	48.94
CCUG 431	*N. mucosa*	68.34	117	1	2.00E-179	0.97	49.87
AP2031	*N. musculi*	58.65	147	3	2.00E-133	0.95	46.04
CCUG 26878 T	*N. oralis*	76.35	91	1	0	0.99	48.94
F0314	*N. oralis*	76.09	92	1	0	0.99	49.02
CCUG 10421	*N. oralis*	76.35	91	1	0	0.99	49.11
CCUG 804	*N. oralis*	76.35	91	1	0	0.99	48.85
NJ9703	*N. subflava*	76.84	90	1	0	1.01	48.09
C102	*N. subflava*	76.84	90	1	0	1.01	48.26
12007_2012	*N. subflava*	77.10	89	1	0	1.01	48.43
C6A	*N. subflava*	76.84	90	1	0	1.01	48.09
CCUG 7826	*N. subflava*	77.10	89	1	0	1.01	48.26
CCUG 24918	*N. subflava*	76.84	90	1	0	1.01	48.18
9715	*N. wadsworthii*	61.89	148	1	2.00E-165	1.00	47.24
CCUG 4007 T	*N. weaveri*	54.76	172	3	6.00E-140	0.99	47.47

B

Isolate	Species	aa ID(%)	Mismatches	Gaps	E-value	Query cover	G+C(%)
CCUG 808	*N. animalis*	77.63	86	1	0	0.97	46.96
DSM 21642	*N. animaloris*	66.40	125	0	8.00E-164	0.96	45.61
CCUG 50858 T	*N. bacilliformis*	79.11	75	0	0	0.90	48.76
CCUG 30380	*N. bacilliformis*	79.11	75	0	0	0.90	48.76
CCUG 38158	*N. bacilliformis*	78.83	76	0	0	0.90	48.68
CCUG 50611	*N. bacilliformis*	79.11	75	0	0	0.90	48.76
CCUG 56775 T	*N. canis*	73.10	99	0	1.00E-172	0.95	46.74
CCUG 53898	*N. dentiae*	68.72	117	0	4.00E-169	0.97	46.35
CCUG 30802T	*N. elongata* subsp. *elongata*	74.49	95	1	0	0.98	47.46
CCUG 2043T	*N. elongata* subsp. *nitroreducens*	74.23	96	1	0	0.98	47.46
CCUG 431	*N. mucosa*	74.68	96	1	0	1.01	48.17
AP2031	*N. musculi*	69.48	112	0	2.00E-168	0.92	46.19
CCUG 26878 T	*N. oralis*	78.57	80	2	0	0.98	47.62
F0314	*N. oralis*	78.32	81	2	0	0.98	47.70
CCUG 10421	*N. oralis*	78.57	80	2	0	0.98	47.62
CCUG 804	*N. oralis*	78.83	79	2	0	0.98	47.62
NJ9703	*N. subflava*	76.79	85	2	0	0.98	48.12
C102	*N. subflava*	76.79	85	2	0	0.98	48.21
12007_2012	*N. subflava*	76.53	86	2	0	0.98	48.29
C6A	*N. subflava*	76.79	85	2	0	1.01	48.12
CCUG 7826	*N. subflava*	76.79	85	2	0	1.01	48.12
CCUG 24918	*N. subflava*	76.79	85	2	0	1.01	47.86
9715	*N. wadsworthii*	72.05	102	0	6.00E-171	0.91	45.83
CCUG 4007 T	*N. weaveri*	66.84	124	0	2.00E-162	0.97	45.94

C

Isolate	Species	aa ID(%)	Mismatches	Gaps	E-value	Query cover	G+C(%)
CCUG 808	*N. animalis*	82.71	46	0	3.00E-142	0.97	50.50
DSM 21642	*N. animaloris*	76.32	63	0	4.00E-139	0.97	52.88
CCUG 50858 T	*N. bacilliformis*	80.08	53	0	2.00E-147	0.97	55.26
CCUG 30380	*N. bacilliformis*	80.08	53	0	6.00E-143	0.97	55.26
CCUG 38158	*N. bacilliformis*	80.83	51	0	4.00E-136	0.97	55.14
CCUG 50611	*N. bacilliformis*	80.08	53	0	7.00E-143	0.97	55.26
CCUG 56775 T	*N. canis*	75.94	64	0	3.00E-132	0.97	52.51
CCUG 53898	*N. dentiae*	77.07	61	0	1.00E-132	0.97	51.38
CCUG 30802T	*N. elongata* subsp. *elongata*	82.33	47	0	6.00E-140	0.97	51.50
CCUG 2043T	*N. elongata* subsp. *nitroreducens*	82.33	47	0	4.00E-140	0.97	51.50
CCUG 431	*N. mucosa*	78.20	58	0	6.00E-135	0.97	52.51
AP2031	*N. musculi*	75.56	65	0	5.00E-130	0.97	49.87
CCUG 26878 T	*N. oralis*	82.33	47	0	3.00E-141	0.97	52.26
F0314	*N. oralis*	82.33	47	0	1.00E-140	0.97	52.38
CCUG 10421	*N. oralis*	82.33	47	0	1.00E-140	0.97	52.38
CCUG 804	*N. oralis*	82.33	47	0	1.00E-140	0.97	52.26
NJ9703	*N. subflava*	79.70	54	0	9.00E-132	0.97	53.01
C102	*N. subflava*	79.32	55	0	1.00E-130	0.97	53.26
12007_2012	*N. subflava*	79.32	55	0	9.00E-131	0.97	53.26
C6A	*N. subflava*	79.70	54	0	3.00E-131	0.97	52.88
CCUG 7826	*N. subflava*	79.70	54	0	2.00E-131	0.97	53.01
CCUG 24918	*N. subflava*	80.08	53	0	2.00E-131	0.97	53.13
9715	*N. wadsworthii*	78.95	56	0	2.00E-139	0.97	51.50
CCUG 4007 T	*N. weaveri*	78.20	58	0	1.00E-139	0.97	52.13

D

Isolate	Species	aa ID(%)	Mismatches	Gaps	E-value	Query cover	G+C(%)
CCUG 808	*N. animalis*	90.70	20	0	3.00E-130	0.99	46.70
DSM 21642	*N. animaloris*	89.30	23	0	7.00E-130	0.99	45.16
CCUG 50858 T	*N. bacilliformis*	87.91	26	0	6.00E-129	0.99	48.39
CCUG 30380	*N. bacilliformis*	87.91	26	0	1.00E-126	0.99	48.08
CCUG 38158	*N. bacilliformis*	87.91	26	0	1.00E-125	0.99	48.39
CCUG 50611	*N. bacilliformis*	87.91	26	0	7.00E-126	0.99	48.39
CCUG 56775 T	*N. canis*	90.19	21	0	7.00E-132	0.99	47.77
CCUG 53898	*N. dentiae*	88.32	25	0	2.00E-128	0.99	47.77
CCUG 30802T	*N. elongata* subsp. *elongata*	87.91	26	0	2.00E-127	0.99	47.81
CCUG 2043T	*N. elongata* subsp. *nitroreducens*	87.91	26	0	2.00E-127	0.99	48.23
CCUG 431	*N. mucosa*	89.72	22	0	4.00E-131	0.99	49.00
AP2031	*N. musculi*	84.04	34	0	1.00E-121	0.98	47.62
CCUG 26878 T	*N. oralis*	93.09	15	0	1.00E-137	1.00	48.69
F0314	*N. oralis*	93.09	15	0	5.00E-137	1.00	48.69
CCUG 10421	*N. oralis*	93.09	15	0	5.00E-137	1.00	48.69
CCUG 804	*N. oralis*	93.09	15	0	5.00E-137	1.00	48.69
NJ9703	*N. subflava*	99.08	2	0	1.00E-144	1.00	48.08
C102	*N. subflava*	99.08	2	0	1.00E-144	1.00	48.39
12007_2012	*N. subflava*	99.08	2	0	1.00E-144	1.00	48.39
C6A	*N. subflava*	98.62	3	0	3.00E-144	1.00	48.39
CCUG 7826	*N. subflava*	98.62	3	0	7.00E-144	1.00	48.23
CCUG 24918	*N. subflava*	99.08	2	0	1.00E-144	1.00	48.08
9715	*N. wadsworthii*	92.52	16	0	5.00E-135	0.99	48.69
CCUG 4007 T	*N. weaveri*	89.25	23	0	6.00E-129	0.99	45.78

E

Isolate	Species	aa ID(%)	Mismatches	Gaps	E-value	Query cover	G+C(%)
CCUG 808	*N. animalis*	73.07	182	2	0	0.97	52.28
DSM 21642	*N. animaloris*	66.27	216	4	0	0.95	51.92
CCUG 50858 T	*N. bacilliformis*	74.12	177	2	0	1.00	52.94
CCUG 30380	*N. bacilliformis*	74.40	175	2	0	1.00	52.89
CCUG 38158	*N. bacilliformis*	73.69	180	2	0	1.00	53.01
CCUG 50611	*N. bacilliformis*	74.12	177	2	0	1.00	52.94
CCUG 56775 T	*N. canis*	51.85	302	6	0	0.92	46.67
CCUG 53898	*N. dentiae*	64.49	197	1	0	0.93	50.09
CCUG 30802T	*N. elongata* subsp. *elongata*	64.49	240	3	0	1.00	52.42
CCUG 2043T	*N. elongata* subsp. *nitroreducens*	64.91	237	3	0	1.00	52.32
CCUG 431	*N. mucosa*	81.53	119	1	0	0.93	52.25
AP2031	*N. musculi*	62.34	220	4	0	0.98	50.42
CCUG 26878 T	*N. oralis*	83.10	108	1	0	0.92	52.35
F0314	*N. oralis*	82.95	109	1	0	0.92	52.20
CCUG 10421	*N. oralis*	83.41	106	1	0	0.92	52.20
CCUG 804	*N. oralis*	83.10	108	1	0	0.92	52.35
NRL30031	*N. subflava*	96.45	25	0	0	1.00	
NJ9703	*N. subflava*	96.74	23	0	0	1.00	53.10
C102	*N. subflava*	93.71	39	1	0	0.92	53.04
CCUG 24841	*N. subflava*	96.61	23	0	0	0.96	
CCUG 24844	*N. subflava*	97.59	17	0	0	1.00	
12007_2012	*N. subflava*	95.18	34	0	0	1.00	53.19
C6A	*N. subflava*	95.74	30	0	0	1.00	53.43
CCUG 4788	*N. subflava*	96.00	15	0	0	0.53	
CCUG 29761	*N. subflava*	95.55	21	0	0	0.67	
CCUG 17913 T	*N. subflava*	96.45	25	0	0	1.00	
CCUG 806	*N. subflava*	96.45	25	0	0	1.00	
CCUG 7826	*N. subflava*	96.74	23	0	0	1.00	53.10
CCUG 24918	*N. subflava*	96.45	25	0	0	1.00	53.00
CCUG 800	*N. subflava*	92.47	11	0	6.00E-88	0.21	
CCUG 801	*N. subflava*	92.47	11	0	6.00E-88	0.21	
CCUG 25198	*N. subflava*	93.89	14	0	0	0.32	
SK114	*N. subflava*	90.64	32	2	0	0.58	
9715	*N. wadsworthii*	62.57	255	1	0	0.97	48.83
CCUG 4007 T	*N. weaveri*	67.06	223	1	0	0.96	51.23

F

Isolate	Species	aa ID(%)	Mismatches	Gaps	E-value	Query cover	G+C(%)
CCUG 808	*N. animalis*	66.42	136	0	0	0.96	55.96
DSM 21642	*N. animaloris*	63.73	147	1	5.00E-179	0.97	53.53
CCUG 50858 T	*N. bacilliformis*	60.39	139	2	7.00E-156	0.85	
CCUG 30380	*N. bacilliformis*	59.31	161	3	7.00E-158	0.96	58.15
CCUG 38158	*N. bacilliformis*	59.48	154	2	7.00E-153	0.92	58.31
CCUG 50611	*N. bacilliformis*	60.00	152	2	2.00E-153	0.92	58.23
CCUG 56775 T	*N. canis*	59.56	163	2	1.00E-162	0.97	51.66
CCUG 53898	*N. dentiae*	60.14	164	2	9.00E-171	1.00	54.57
CCUG 30802T	*N. elongata* subsp. *elongata*	60.75	156	1	9.00E-165	0.95	56.10
CCUG 2043T	*N. elongata* subsp. *nitroreducens*	60.75	156	1	1.00E-163	0.95	56.02
CCUG 431	*N. mucosa*	66.91	133	2	0	0.97	54.80
AP2031	*N. musculi*	60.57	165	1	9.00E-173	1.00	54.63
CCUG 26878 T	*N. oralis*	66.58	134	1	0	0.96	54.65
F0314	*N. oralis*	66.34	135	1	0	0.96	54.80

CCUG 10421	*N. oralis*	66.34	135	1	0	0.96	54.58
CCUG 804	*N. oralis*	66.34	135	1	0	0.96	54.73
NRL30031	*N. subflava*	97.62	10	0	0	1.00	
NJ9703	*N. subflava*	85.57	43	0	3.00E-168	0.71	
C102	*N. subflava*	91.93	33	0	0	0.97	55.40
CCUG 24841	*N. subflava*	96.90	13	0	0	1.00	
CCUG 24844	*N. subflava*	97.62	10	0	0	1.00	
12007_2012	*N. subflava*	94.29	24	0	0	1.00	54.92
C6A	*N. subflava*	92.18	32	0	0	0.97	55.32
CCUG 4788	*N. subflava*	94.67	4	0	1.00E-39	0.18	
CCUG 29761	*N. subflava*	71.05	15	2	1.00E-39	0.27	
CCUG 17913 T	*N. subflava*	97.62	10	0	0	1.00	
CCUG 806	*N. subflava*	97.62	10	0	0	1.00	
CCUG 7826	*N. subflava*	93.10	29	0	0	1.00	55.08
CCUG 24918	*N. subflava*	93.15	28	0	0	0.97	55.08
CCUG 800	*N. subflava*	84.08	32	0	3.00E-114	0.48	
CCUG 801	*N. subflava*	84.08	32	0	3.00E-114	0.48	
CCUG 25198	*N. subflava*	96.54	12	0	0	0.83	
SK114	*N. subflava*	94.67	4	0	4.00E-40	0.18	
9715	*N. wadsworthii*	60.29	160	2	4.00E-164	0.97	51.58
CCUG 4007 T	*N. weaveri*	63.48	148	1	2.00E-175	0.97	51.88

and *N. bacilliformis*. In the case of *N. canis*, *N. animalis* and *N. bacilliformis*, the different regions were not located on a single contig, but referral to closed genome sequences confirmed this separation was not an artefact. A *galE* orthologue was not present in all *cps*$^+$ isolates; in *N. animalis*, *N. oralis* and both *N. elongata* isolates, it was found to be near or adjacent to the putative region A, rather than contiguously with the other region D genes.

2.3.1.1 NPN species possess *cps* region B and C orthologues

With the exception of 11 *N. subflava* isolates, isolates in which capsule genes were identified possessed orthologous sequences of all six of the meningococcal *cps* region B and C genes (Table 2.1). The remaining 11 *N. subflava* isolates were found to contain putative orthologues for one or both of *ctrE* and *ctrF*, but no region C orthologues were identified (Table 2.1).

Orthologues shared 52-99% amino acid sequence identity with the relevant query. Alignment length covered at least 92% of the relevant query, except the orthologues of *ctrF* in CCUG 50858 (85%) and NJ9703 (71%). This reduced query coverage was attributed to an incomplete gene located at the end of a contig in WGS data from isolate CCUG 50858, and a frame-shift mutation in the sequence from isolate NJ9703.

ctrABCD orthologues were contiguous, and in the same order as those found in *N. meningitidis*. *ctrEF* orthologues were contiguous in all cps^+ isolates, except *N. canis* and *N. bacilliformis*, whose *ctrE* orthologue was adjacent to the putative region C (Figure 2.2).

2.3.1.2 Putative region A genes were identified in NPN species possessing regions B and C

In each isolate genome, with the exception of *N. bacilliformis*, a series of ORFs was identified immediately upstream of region C (*ctrE*-region C in the case of *N. canis*), which had not been previously curated in PubMLST (Figure 2.2). These ORFs were annotated as putative region A genes (Table 2.2). The ORFs are described in detail below.

In *N. mucosa*, two ORFs between putative region C and the previously annotated gene NEIS1114 (described as a membrane protein) were homologous to capsule synthesis genes from *N. meningitidis* serogroup L. In *N. weaveri*, three ORFs between putative regions B and C were homologous to capsule synthesis genes from *A. pleuropneumoniae* serovar 5. In *N. zoodegmatis*, five ORFs between putative regions B and C were homologous to capsule synthesis genes from *H. influenzae* serotype d.

N. dentiae and *N. musculi* each contained seven genes between regions B and C, of which the two immediately upstream of region C and the one immediately upstream of region B were homologous to capsule synthesis genes from *H. influenzae* serotype d. The remaining ORFs in between had only hypothetical matches. In both species, two of the four hypothetical genes were predicted glycosyltransferases, and a third in *N. musculi* was a predicted rhamnosyltransferase.

The putative region A in *N. canis* was inferred using a closed genome from the NCTC 3000 project. Upstream of the putative region C in *N. canis*, two ORFs homologous to capsule synthesis genes from *H. influenzae* serotype d, two predicted glycosyltransferases, a predicted rhamnosyltransferase, two hypothetical genes, and

Table 2.2: Annotation of putative region A genes. Annotations based on the results of queries using tBLASTn against the NCBI nucleotide collection, and BLASTp against Pfam-A. Hypothetical matches also queried using BLASTp against the NCBI RefSeq database. One representative isolate per NPN species.

Isolate/Species	NEIS	NCBI nucleotide (coverage; aa ID; Accession)	Pfam-A matches	G+C
CCUG 808 ***N. animalis***	2173	*cszA [N. meningitidis]* (100%; 81%; HF562983.1) *cps17B/8B/6B/2B/13B/7B/3B [A. pleuropneumoniae]* (100%; 78%; MG780422.1/KJ685493.1/AY534316.1/ AY357726.1/MG868947.1/ AY534317.1/ CP000687.1)	"Cytidylyltransferase-like"	43.6
	2947	*cszB [N. meningitidis]* (95%; 56%; HF562983.1) *cps17C/8C/6C/2C/13C/7C/3C [A. pleuropneumoniae]* (88-9%; 46-7%; MG780422.1/KJ685493.1/ AY534316.1/AY357726.1/MG868947.1/ AY534317.1/ CP000687.1)	No matches	43.8
	2939	*cps17D/8D/6D [A. pleuropneumoniae]* (99%; 61%; MG780422.1/KJ685493.1/AY534316.1)	"CDP-Glycerol: Poly(glycerophosphate) glycerophosphotransferase" "Glycosyl transferases group 1"	44.2
	2940	*cps17E/8E [A. pleuropneumoniae]* (100%; 72%; MG780422.1/KJ685493.1) *cps6E [A. pleuropneumoniae]* (84%; 76%; AY534316.1)	"Glycosyl transferase family 2"	40.6
	2966	*cps17F [A. pleuropneumoniae]* (29%; 46%; MG780422.1)	No matches	41.1
	2967	*ccs3 [H. influenzae]* (91%; 27%; HQ651151.1) *pcsC [H. parainfluenzae]* (90%; 27%; MH644108.1] *cps14C [A. pleuropneumoniae]* (94; 28%; AB810251.1)	"Accessory Sec system GspB-transporter"	45.8
CCUG 38158 ***N. bacilliformis***	2173	*cszA [N. meningitidis]* (96%; 75%; HF562983.1) *cps17B/8B/6B/2B/13B/7B/3B [A. pleuropneumoniae]* (96-7%; 78%; MG780422.1/ KJ685493.1/AY534316.1/AY357726.1/ MG868947.1/AY534317.1/ CP000687.1)	"Cytidylyltransferase-like"	37.6
	2938	*cszB [N. meningitidis]* (98%; 49%; HF562983.1) *cps17C/8C/7C/3C [A. pleuropneumoniae]* (81%; 45-6%; MG780422.1/KJ685493.1/ AY534317.1/ CP000687.1) *cps6C/2C/13C [A. pleuropneumoniae]* (90-1%; 42%; AY534316.1/AY357726.1/MG868947.1)	"LicD family"	36.7
	2939	*cps17D/8D/6D [A. pleuropneumoniae]* (99%; 57%; MG780422.1/KJ685493.1/AY534316.1)	"Tetratricopeptide repeat" "CDP-Glycerol: Poly(glycerophosphate) glycerophosphotransferase" "Glycosyl transferases group 1"	37.9
	2940	*cps17E/8E [A. pleuropneumoniae]* (100%; 59%; MG780422.1/KJ685493.1) *cps6E [A. pleuropneumoniae]* (84%; 61%; AY534316.1)	"Glycosyl transferase family 2"	39.4
CCUG 30802T ***N.elongata*** **subsp.** ***nitroreducens***	2173	*cszA [N. meningitidis]* (99%; 79%; HF562983.1) *cps17B/8B/6B/2B/13B/7B/3B* [*A. pleuropneumoniae*] (98-100%; 80-81%; MG780422.1/KJ685493.1/AY534316.1/ AY357726.1/MG868947.1/ AY534317.1/ CP000687.1)	"Cytidylyltransferase-like"	39.2

Isolate/Species	NEIS	NCBI nucleotide (coverage; aa ID; Accession)	Pfam-A matches	G+C
CCUG 30802T ***N.elongata*** **subsp.** ***nitroreducens***	2948	*cszB [N. meningitidis]* (88%; 47%; HF562983.1) *cps17C/8C/7C/3C [A. pleuropneumoniae]* (84%; 42-3%; MG780422.1/KJ685493.1/ AY534317.1/ CP000687.1) *cps6C/2C/13C [A. pleuropneumoniae]* (75%; 44%; AY534316.1/AY357726.1/ MG868947.1)	No matches	33.7
	2939	*cps17D/8D/6D [A. pleuropneumoniae]* (98-9%; 50%; MG780422.1/KJ685493.1/ AY534316.1)	"Tetratricopeptide repeat" "CDP-Glycerol: Poly(glycerophosphate) glycerophosphotransferase" "Glycosyl transferases group 1"	36.5
	2940	*cps17E/8E [A. pleuropneumoniae]* (98%; 68%; MG780422.1/KJ685493.1) *cps6E [A. pleuropneumoniae]* (82%; 67%; AY534316.1)	"Glycosyl transferase family 2"	37.4
CCUG 56775 T ***N. canis***	2157	*csaA [N. meningitidis]* (98%; 65%; MF503101.1) *dcs1/ecs1 [H. influenzae]* (98%; 69-70%; HQ424464.1/HM053635.1) *csiA [N. meningitidis]* (97%; 57%; HF562985.1) *bcbA [P. multocida]* (98%; 53%; AF169324.1) *wecB [V. cholerae]* (98%; 61%; GU576498.1)	"UDP-N-acetylglucosamine 2-epimerase"	40.3
	2737	*csiB [N. meningitidis]* (99%; 81%; HF562985.1) *dcs2/ecs2 [H. influenzae]* (99%; 65%; HQ424464.1/HM053635.1) *bcbB [P. multocida]* (99%; 78%; AF169324.1) *wecC [V. cholerae]* (99%; 69%; GU576498.1)	"UDP-glucose/GDP-mannose dehydrogenase family" (NAD, central and UDP binding)	43.4
	2738	hypothetical Refseq: glycosyltransferase *[A. rudis]* (70%; 48%; WP_016657901.1)	"Glycosyl transferases group 1"	37.1
	2739	hypothetical Refseq: glycosyltransferase *[M. haemolytica]* (100%; 74%; WP_080869760.1)	"Glycosyl transferase family 2"	38.9
	2740	hypothetical	"Putative rhamnosyl transferase"	31.3
	2968	hypothetical RefSeq: alpha/beta hydolase *[Alcaligenes sp.* RS4] (62%; 35%; WP_128393077.1)	No matches	30.2
	2741	*dcs3 [H. influenzae]* (57%; 37%; HQ424464.1)	No matches	34.2
	2742	*csiE [N. meningitidis]* (98%; 80%; HF562985.1) *dcs5/ecs8 [H. influenzae]* (99%; 60-62%; HQ424464.1/HM053635.1) *bcbI [P. multocida]* (100%; 81%; AF169324.1)	"Glycosyl transferases group 1"	38.5
CCUG 53898 ***N. dentiae***	2736	*csiA [N. meningitidis]* (96%; 70%; HF562985.1) *bcbA [P. multocida]* (96%; 71%; AF169324.1) *csaA [N. meningitidis]* (95%; 54%; MF503101.1) *dcs1/ecs1 [H. influenzae]* (95%; 55%; HQ424464.1/HM053635.1) *wecB [V. cholerae]* (97%; 60%; GU576498.1)	"UDP-N-acetylglucosamine 2-epimerase"	50.2
	2937	*csiB [N. meningitidis]* (93%; 86%; HF562985.1) *bcbB [P. multocida]* (94%; 79%; AF169324.1) *dcs2/ecs2 [H. influenzae]* (92%; 65-6%; HQ424464.1/HM053635.1) *wecC [V. cholerae]* (92%; 70%; GU576498.1)	"UDP-glucose/GDP-mannose dehydrogenase family" (NAD, central and UDP binding)	49.9

Isolate/Species	NEIS	NCBI nucleotide (coverage; aa ID; Accession)	Pfam-A matches	G+C
CCUG 53898	2738	hypothetical	"Glycosyltransferases group 1"	35.4
N. dentiae		RefSeq: glycosyltransferase *[M. haemolytica]* (99%; 89%; WP_052500173.1)		
	2739	hypothetical RefSeq: glycosyltransferase *[R. baltica]* (81%; 47%; WP_027672790.1)	"Glycosyl transferase family 2"	38.5
	2750	hypothetical	No matches	31.6
	2741	*dcs3 [H. influenzae]* (59%; 37%; HQ424464.1)	No matches	35.1
	2942	*csiE [N.meningitidis]* (98%; 87%; HF562985.1) *dcs5/ecs8 [H. influenzae]* (99%; 62-4%; HQ424464.1/HM053635.1) *bcbI [P. multocida]* (100%; 77%; AF169324.1)	"Glycosyl transferases group 1"	46.4
AP2031	2736	*csiA [N. meningitidis]* (97%; 69%; HF562985.1) *bcbA [P. multocida]* (97%; 72%; AF169324.1) *csaA [N. meningitidis]* (97%; 54%; MF503101.1) *dcs1/ecs1 [H. influenzae]* (96%; 54-5%; HQ424464.1/HM053635.1) *wecB [V. cholerae]* (98%; 58%; GU576498.1)	"UDP-N-acetylglucosamine 2-epimerase"	46.1
N. musculi	2737	*csiB [N. meningitidis]* (99%; 84%; HF562985.1) *bcbB [P. multocida]* (99%; 77%; AF169324.1) *dcs2/ecs2 [H. influenzae]* (99%; 64-5%; HQ424464.1/HM053635.1) *wecC [V. cholerae]* (99%; 68%; GU576498.1)	"UDP-glucose/GDP-mannose dehydrogenase family" (NAD, central and UDP binding)	49.5
	2738	hypothetical Refseq: glycosyltransferase *[A. rudis]* (73%; 46%; WP_016657901.1)	"Glycosyl transferases group 1"	39.6
	2739	hypothetical Refseq: glycosyltransferase *[M. haemolytica]* (100%; 84%; WP_080869760.1)	"Glycosyl transferase family 2"	37.1
	2740	hypothetical	"Putative rhamnosyl transferase"	30.1
	2741	*dcs3 [H. influenzae]* (58%; 36%; HQ424464.1)	No matches	40.8
	2742	*csiE [N.meningitidis]* (98%; 82%; HF562985.1) *dcs5/ecs8 [H. influenzae]* (98%; 58-9%; HQ424464.1/HM053635.1) *bcbI [P. multocida]* (100%; 72%; AF169324.1)	"Glycosyl transferases group 1"	46.0
CCUG 12230 T	2736	*csiA [N. meningitidis]* (96%; 72%; HF562985.1) *bcbA [P. multocida]* (96%; 72%; AF169324.1) *csaA [N. meningitidis]* (94%; 54%; MF503101.1) *dcs1/ecs1 [H. influenzae]* (95%; 57%; HQ424464.1/HM053635.1) *wecB [V. cholerae]* (96%; 60%; GU576498.1)	"UDP-N-acetylglucosamine 2-epimerase"	40.7
N. zoodegmatis	2737	*csiB [N. meningitidis]* (99%; 82%; HF562985.1) *bcbB [P. multocida]* (99%; 80%; AF169324.1) *dcs2/ecs2 [H. influenzae]* (99%; 66-7%; HQ424464.1/HM053635.1) *wecC [V. cholerae]* (99%; 70%; GU576498.1)	"UDP-glucose/GDP-mannose dehydrogenase family" (NAD, central and UDP binding)	42.3
	2949	*dcs3 [H. influenzae]* (100%; 69%; HQ424464.1)	"Glycosyl transferases group 1" "Glycosyl transferase family 2"	37.6
	2971	*dcs4 [H. influenzae]* (97%; 59%; HQ424464.1)	No matches	34.7
	2742	*csiE [N.meningitidis]* (98%; 77%; HF562985.1) *dcs5/ecs8 [H. influenzae]* (99%; 61-2%; HQ424464.1/HM053635.1) *bcbI [P. multocida]* (100%; 81%; AF169324.1)	"Glycosyl transferases group 1"	37.8

Isolate/Species	NEIS	NCBI nucleotide (coverage; aa ID; Accession)	Pfam-A matches	G+C
CCUG 2043 T	2965	*cslA [N. meningitidis]* (100%; 60%; HF562986.1)	"Stealth protein conserved	31.4
N. elongata		*cps14A/18A/4A/12A/15A [A. pleuropneumoniae]*	region" (1/2)	
subsp. *elongata*		(99%; 54-6%; AB810251.1/MG780424.1/		
		GU585380.1/AB701757.1/LC012002.1)		
		cps1A [A. pleuropneumoniae]		
		(89%; 56%; AF518558.1)		
		ccs1 [H. influenzae] (99%; 55%; HQ651151.1)		
		fcs1 [H. influenzae] (94%; 54%; AF549211.1)		
		pcsA [H. parainfluenzae] (99%; 55%; MH644108.1)		
	2941	*cslB [N. meningitidis]* (67%; 34%; HF562986.1)	"Glycosyl transferase family 2"	36.1
		cps14B1/2 [A. pleuropneumoniae]	"CDP-Glycerol:	
		(88%; 60%; AB810251.1)	Poly(glycerophosphate)	
		cps4B [A. pleuropneumoniae]	glycerophosphotransferase"	
		(69% 50%; GU585380.1)		
		cps1B/C [A. pleuropneumoniae]		
		(63%; 48%; AF518558.1)		
		ccs2 [H. influenzae] (99%; 49%; HQ651151.1)		
		pcsB [H. parainfluenzae] (99%; 46%; MH644108.1)		
	2942	*ccs4 [H. influenzae]* (56%; 57%; HQ651151.1)	"Bacterial transferase	37.9
		pcsD [H. parainfluenzae] (59%; 56%; MH644108.1)	hexapeptide"	
	2943	*cshD [N. meningitidis]* (67%; 26%; HF562983.1)	No matches	34.6
		csxB [N. meningitidis] (39; 30%; HF562988.1)		
		xcbB [E. coli] (43%; 31%; MG736912.1)		
	2974	hypothetical	"AMP-binding enzyme"	37.7
F0314	2941	*cps14B1/2 [A.pleuropneumoniae]*	"Glycosyl transferase family 2"	39.6
N. oralis		(89%; 48%; AB810251.1)	"CDP-Glycerol:	
		cps4B [A. pleuropneumoniae]	Poly(glycerophosphate)	
		(69% 48%; GU585380.1)	glycerophosphotransferase"	
		cps1B/C [A.pleuropneumoniae]		
		(63%; 56%; AF518558.1)		
		ccs2 [H. influenzae] (99%; 42%; HQ651151.1)		
		pcsB [H. parainfluenzae] (99%; 55%; MH644108.1)		
	2943	*cshD [N. meningitidis]* (67%; 25%; HF562983.1)	No matches	36.3
		csxB [N. meningitidis] (43; 28%; HF562988.1)		
		xcbB [E. coli] (43%; 31%; MG736912.1)		
	2974	hypothetical	"AMP-binding enzyme"	38.0
NJ9703	2184	*cslA [N. meningitidis]* (99%; 72%; HF562986.1)	"Stealth protein conserved	43.0
N. subflava		*cps14A/18A/4A/12A/1A/15A [A. pleuropneumoniae]*	region" (1/2/3)	
		(98-9%; 54-6%; AB810251.1/MG780424.1/		
		GU585380.1/AB701757.1/		
		AF518558.1/LC012002.1)		
		ccs1 [H. influenzae] (99%; 56%; HQ651151.1)		
		fcs1 [H. influenzae] (94%; 56%; AF549211.1)		
		pcsA [H. parainfluenzae] (99%; 56%; MH644108.1)		
	2941	*cslB [N. meningitidis]* (67%; 32%; HF562986.1)	"Glycosyl transferase family 2"	39.7
		cps14B1/2 [A. pleuropneumoniae]	"CDP-Glycerol:	
		(88%; 49%; AB810251.1)	Poly(glycerophosphate)	
		cps4B [A. pleuropneumoniae]	glycerophosphotransferase"	
		(69% 49%; GU585380.1)		
		cps1B/C [A. pleuropneumoniae]		
		(63%; 57%; AF518558.1)		
		ccs2 [H. influenzae] (99%; 43%; HQ651151.1)		
		pcsB [H. parainfluenzae] (99%; 55%; MH644108.1)		
	2942	*ccs4 [H. influenzae] (56%; 57%; HQ651151.1)*	"Bacterial transferase	37.2
		pcsD [H. parainfluenzae] (59%; 56%; MH644108.1)	hexapeptide"	

Isolate/Species	NEIS	NCBI nucleotide (coverage; aa ID; Accession)	Pfam-A matches	G+C
NJ9703	2943	*cshD [N. meningitidis]* (67%; 27%; HF562983.1)	No matches	36.1
N. subflava		*csxB [N. meningitidis]* (39; 34%; HF562988.1)		
		xcbB [E. coli] (58%; 29%; MG736912.1)		
	2974	hypothetical	"AMP-binding enzyme"	40.6
CCUG 431	2972	*cslA [N. meningitidis]* (97%; 58%; HF562986.1)	"Glycosyl transferase family 2"	29.9
N. mucosa		*cps12A [A. pleuropneumoniae]* (95%; 57%; AB701757.1)	"CDP-Glycerol: Poly(glycerophosphate) glycerophophotransferase"	
	2973	*cslB [N. meningitidis]* (97%; 56%; HF562986.1)	No matches	33.5
CCUG 4007 T	2944	*cps5A [A. pleuropneumoniae]* (97%; 62%; AF053723.1)	"Glycosyl transferases group 1"	30.2
N. weaveri	2945	*cps5B [A. pleuropneumoniae]* (99%; 57%; AF053723.1)	"Capsule polysaccharide biosynthesis"	28.6
	2946	*cps5C [A. pleuropneumoniae]* (99%; 47%; AF053723.1)	No matches	40.7

A. rudis: Acinetobacter rudis; R. baltica: Rheiheimera baltica; V. cholerae: Vibrio cholerae

an additional ORF homologous to a capsule synthesis gene from *H. influenzae* serotype d, were annotated as putative region A genes. Downstream of these genes were a predicted IS*481* insertion sequence, and a locus that had been annotated *kdsA*; these were not included in the putative region A. No further ORFs were investigated.

N. oralis, *N. subflava*, *N. elongata* subsp. *nitroreducens*, *N. elongata* subsp. *elongata* and *N. animalis* possessed two, four/five, four, six or seven ORFs respectively between region C and a *galE* orthologue. *N. elongata* subsp. *elongata* possessed a predicted IS*565* insertion sequence immediately upstream of *galE*, which was not included in the putative region A. Adjacent to this was a predicted AMP-binding enzyme, which was also present immediately upstream of *galE* in *N. subflava* and *N. oralis*. *N. animalis* possessed a predicted nuclease immediately upstream of *galE*, which was not included in the putative region A. It also included an ORF between NEIS2966 and NEIS2977, which appeared to be a truncated pseudogene of NEIS2966; it was not assigned a NEIS number but left annotated as ORF6. The remaining ORFs immediately upstream of region C in all five species were homologous to capsule synthesis genes from various species, and were annotated as putative region A genes. None of the putative region A genes from *N. subflava* were found in *N. subflava* isolates that lacked region C.

None of the putative region A genes described were identified through BLAST searches in the genomes from either *N. subflava* isolates that contained a putative region B, but not region C, or species totally lacking putative regions B and C. Orthologues of the four putative *N. elongata* subsp. *nitroreducens* region A genes were identified in *N. bacilliformis* through BLASTn searches. In three of four *N. bacilliformis* isolate genomes, these genes were contiguous, as in *N. elongata* subsp. *nitroreducens*, but were located on different contigs from regions B and C. In the fourth isolate, the genes were disrupted, likely due to poor assembly.

2.3.1.3 NPN species share homologous putative region A sequences

In *N. bacilliformis* (four isolates), *N. subflava* (six isolates), *N. oralis* (four isolates), and *N. elongata* (two isolates), genomes from more than one isolate of the same species possessed a putative *cps*. The putative region A nucleotide sequences from the *N. bacilliformis* genomes all contained four genes that were each 98% identical between isolate genomes, indicating that all four *N. bacilliformis* genomes had an almost identical putative region A sequences (Figure 2.3A). The putative region A nucleotide sequences from the *N. subflava* genomes all contained four genes that were >97% identical between isolate genomes (Figure 2.3B). Three *N. subflava* putative region A sequences contained an additional gene homologous with a predicted acetyltransferase [202], NEIS2942, that was >98% identical among isolate genomes. The putative region A nucleotide sequences from all six *N. subflava* genomes also contained a predicted glycosyltransferase, NEIS2941, which was either 71-73% or 97-100% identical between isolate genomes, indicating that there were two versions of this gene (Figure 2.3B). NEIS2941 version 1 was found in C102, C6A, CCUG 7826 and CCUG 24918. NEIS2941 version 2 was found in NJ9703 and 12007_2012. The putative region A nucleotide sequences from the *N. oralis* genomes all contained three genes, two of which were >98% identical between isolate genomes (Figure 2.3C). NEIS2943 was >98% identical among three of the four isolates genomes, with

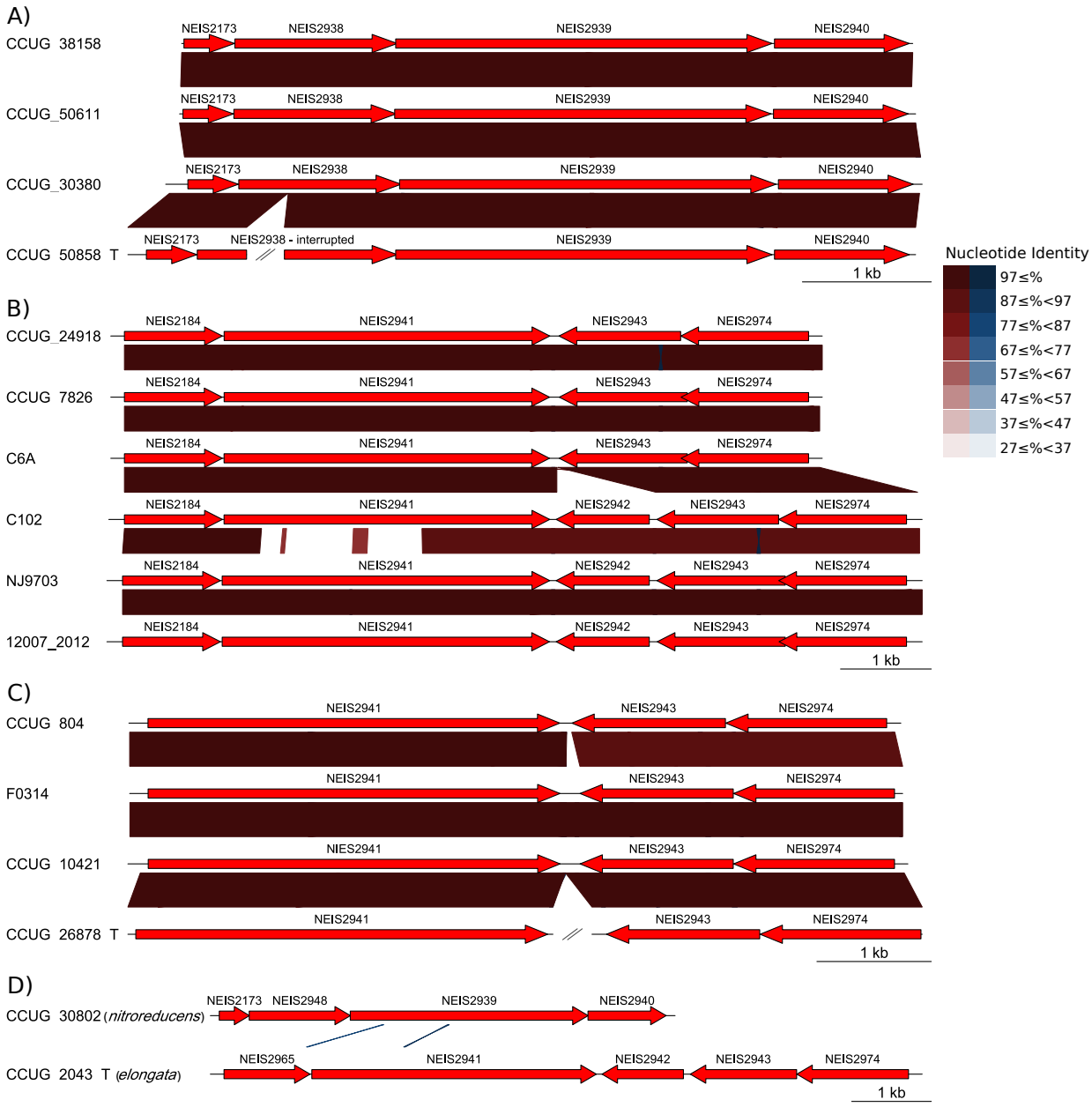

Figure 2.3: Within species comparisons of putative region A. Pairwise BLASTn comparisons of putative region A sequences within species A) *N. bacilliformis*, B) *N. subflava*, C) *N. oralis* and D) *N. elongata*. Colour intensity indicates sequence identity. Blue indicates inversion.

the version from CCUG 804 only sharing 81% nucleotide identity with the others. The putative region A from *N. elongata* subsp. *elongata* did not resemble that of *N. elongata* subsp. *nitroreducens* (Figure 2.3D).

NPN species were sorted into three groups based on shared homology within region A sequences (Figure 2.4). Homologous genes were only assigned the same

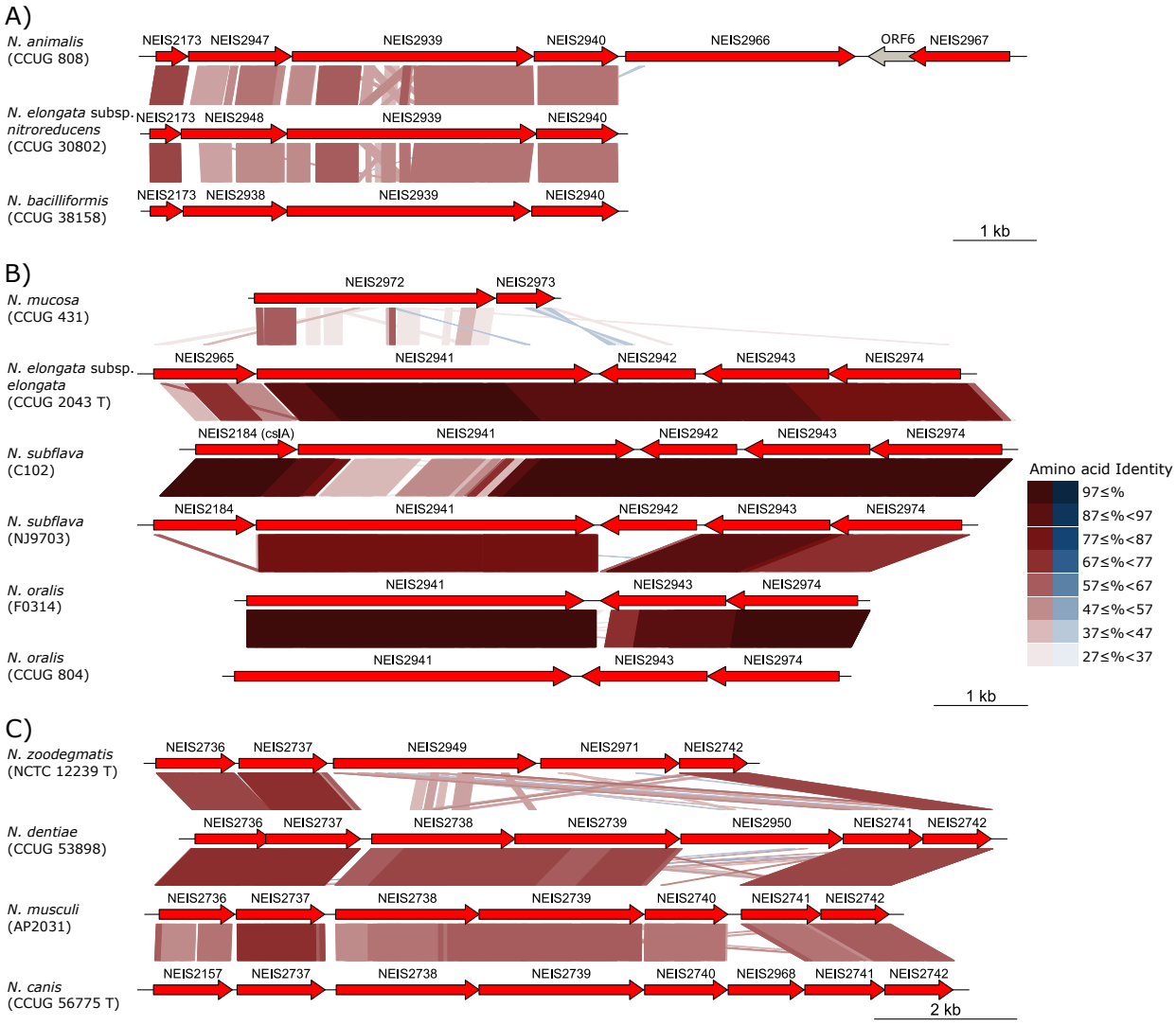

Figure 2.4: Homologous putative region A sequences between NPN species. Pairwise tBLASTx comparisons of homologous putative region A sequences between representative NPN species belonging to putative region A homology groups A) Group 1, B) Group 2 and C) Group 3. Colour intensity indicates sequence identity. Blue indicates inversion.

NEIS number if they were sufficiently identical for additional alleles to be assigned by the PubMLST genome scanner. Details of this homology are described below.

Group 1 contained putative region A sequences from *N. animalis*, *N. elongata* subsp. *nitroreducens* and *N. bacilliformis*, which shared four homologous genes with 78-83%, 42-47%, 56-58% and 61-64% amino acid identity between species (Figure 2.4A). The putative *N. animalis* region A contained two additional genes.

Group 2 contained putative region A sequences from *N. elongata* subsp. *elongata*, *N. subflava* and *N. oralis*, which shared up to five homologous genes with 58%, 60-95%, 97%, 86-95% and 75-99% amino acid identity between species. The *N. oralis* putative region A did not contain a NEIS2184 homologue (Figure 2.4B). The putative *N. mucosa* region A was also partially homologous with other Group 2 putative region A sequences. *N. mucosa* gene NEIS2972 was 32% identical to group 2 gene NEIS2941, and *N. mucosa* gene NEIS2973 was 25% identical to group 2 gene NEIS2942 at the amino acid level; neither gene was equivalent in length to its homologue (Figure 2.4B).

Group 3 contained putative region A sequences from *N. dentiae*, *N. musculi*, *N. canis*, *N. zoodegmatis* and *N. animaloris*, although *N. animaloris* putative region A was not annotated further due to its incomplete assembly. The remaining Group 3 putative region A sequences shared up to seven homologous genes with 56-96%, 86-94%, 70-82%, 73-85%, 65%, 68-83% and 79-90% amino acid identity between species (Figure 2.4C). The putative *N. zoodegmatis* region A only contained three of the homologous genes, and an additional two genes. *N. dentiae* only contained six of the homologous genes, and one additional gene. The putative *N. canis* region A contained one additional gene.

2.3.1.4 Putative NPN region A sequences resemble previously characterised capsule synthesis regions

The putative region A sequences belonging to representatives from each homology group were found to be homologous in whole or in part to capsule synthesis regions in *N. meningitidis* and other Gram-negative bacteria (Figure 2.5). Details of this homology are described below.

Putative region A sequences in Group 1 resembled the capsule synthesis region of *N. meningitidis* serogroups H and Z, and *A. pleuropneumoniae* serovars 17, 8 and 6 (Figure 2.5A). The putative region A from *N. animalis* was chosen as a representative of Group 1, since it contained additional genes not found in other members of the

group. Amino acid sequences were compared. The first two genes, NEIS2173 and NEIS2947, were 81% and 54% identical to the *N. meningitidis* capsule synthesis genes *cszA;cshA* and *cszB;cshB*. NEIS2173, NEIS2947, and the next two genes, NEIS2939 and NEIS2940, were 78%, 45%, 64% and 72-76% identical to the *A. pleuropneumoniae* capsule synthesis genes *cps17B/cps8B/cps6B*, *cps17C/cps8C/cps6C*, *cps17D/cps8D/cps6D* and *cps17E/cps8E/cps6E* respectively. NEIS2966 and ORF6 were partially homologous, with 46% and 37% amino acid identity respectively, to *cps17F*, which is not present in *A. pleuropneumoniae* serovars 8 or 6. The last gene, NEIS2967, was not homologous with any gene in the capsule synthesis region of *N. meningitidis* serogroups H or Z, or *A. pleuropneumoniae* serovars 17, 8 or 6.

Putative region A sequences in Group 2 resembled the capsule synthesis region of *N. meningitidis* serogroup L, *H. influenzae* serotype c, *H. parainfluenzae* strain HUB12445 (recently found to be encapsulated) [219], and *A. pleuropneumoniae* serovars 12 and 14 (Figure 2.5B). The putative region A from *N. elongata* subsp. *elongata* was chosen as a representative of Group 2, since it contained all five homologous genes. Amino acid sequences were compared. The first two genes, NEIS2965 and NEIS2941, were 47-58% and 32-49% identical to the *N. meningitidis* capsule synthesis genes *cslA* and *cslB*, the *H. influenzae* capsule synthesis genes *ccsA* and *ccsB*, the *H. parainfluenzae* capsule synthesis genes *pcsA* and *pcsB*, and the *A. pleuropneumoniae* capsule synthesis genes *cps14A* and *cps14B1B2B3*. The third gene, NEIS2942, was partially homologous to *cslC* (18% identity) and *ccsD/pcsD* (42-60% identity). The last two genes, NEIS2943 and NEIS2974, were not homologous with any gene in the capsule synthesis region of *N. meningitidis* serogroup L, *H. influenzae* serotype c, *H. parainfluenzae* strain HUB12445, or *A. pleuropneumoniae* serovar 14.

The putative region A from *N. mucosa* was chosen as a second representative of Group 2, since it shared comparatively little homology with the other Group 2 putative region A sequences (Figure 2.4B). Amino acid sequences were compared

(Figure 2.5C). The first gene was 59% identical to the *N. meningitidis* capsule synthesis gene *cslB*, and the *A. pleuropneumoniae* serovar 12 gene *cps12B*. The second gene was 23% identical to *cslC*. The *N. mucosa* putative region A did not contain any sequences homologous to *cslA*, but *cslA* was 34% identical to *cps12A*.

Putative region A sequences in Group 3 resembled the capsule synthesis region of *N. meningitidis* serogroup I, and *H. influenzae* serotype d (Figure 2.5D). *N. musculi* was chosen as a representative of Group 3, since its putative region A contained all seven homologous genes. Amino acid sequences were compared. The first two genes, NEIS2736 and NEIS2737, and the final gene, NEIS2743, were 71%, 87% and 84% identical to the *N. meningitidis* capsule synthesis genes *csiA*, *csiB* and *csiE*, and 56%, 66% and 61% identical to the *H. influenzae* serotype d capsule synthesis genes *dcs1*, *dcs2* and *dcs5*. NEIS2738 was partially homologous to *dcs3* (27% identity) and *dcs5* (32% identity). NEIS2739, NEIS2940 and NEIS2741 were not substantially homologous with any gene in the capsule synthesis region of *N. meningitidis* serogroup I, or *H. influenzae* serotype d.

N. zoodegmatis was chosen as a second representative of Group 3, since its putative region A contains different genes from other members of the group (Figure 2.4C). Amino acid sequences were compared (Figure 2.5E). The five genes were 58%, 69%, 56%, 57% and 61% identical to the genes *dcs1*, *dcs2*, *dcs3*, *dcs4* and *dcs5* respectively, which constitute the entire capsule synthesis region of *H. influenzae* serotype d.

The putative region A sequence from *N. weaveri* resembled the capsule synthesis region of *A. pleuropneumoniae* serovar 5 (Figure 2.5F). Amino acid sequences were compared. The three genes were 59%, 59% and 46% identical to *cps5A*, *cps5B* and *cps5C* respectively, which constitute three of the four genes in the capsule synthesis region of *A. pleuropneumoniae* serovar 5.

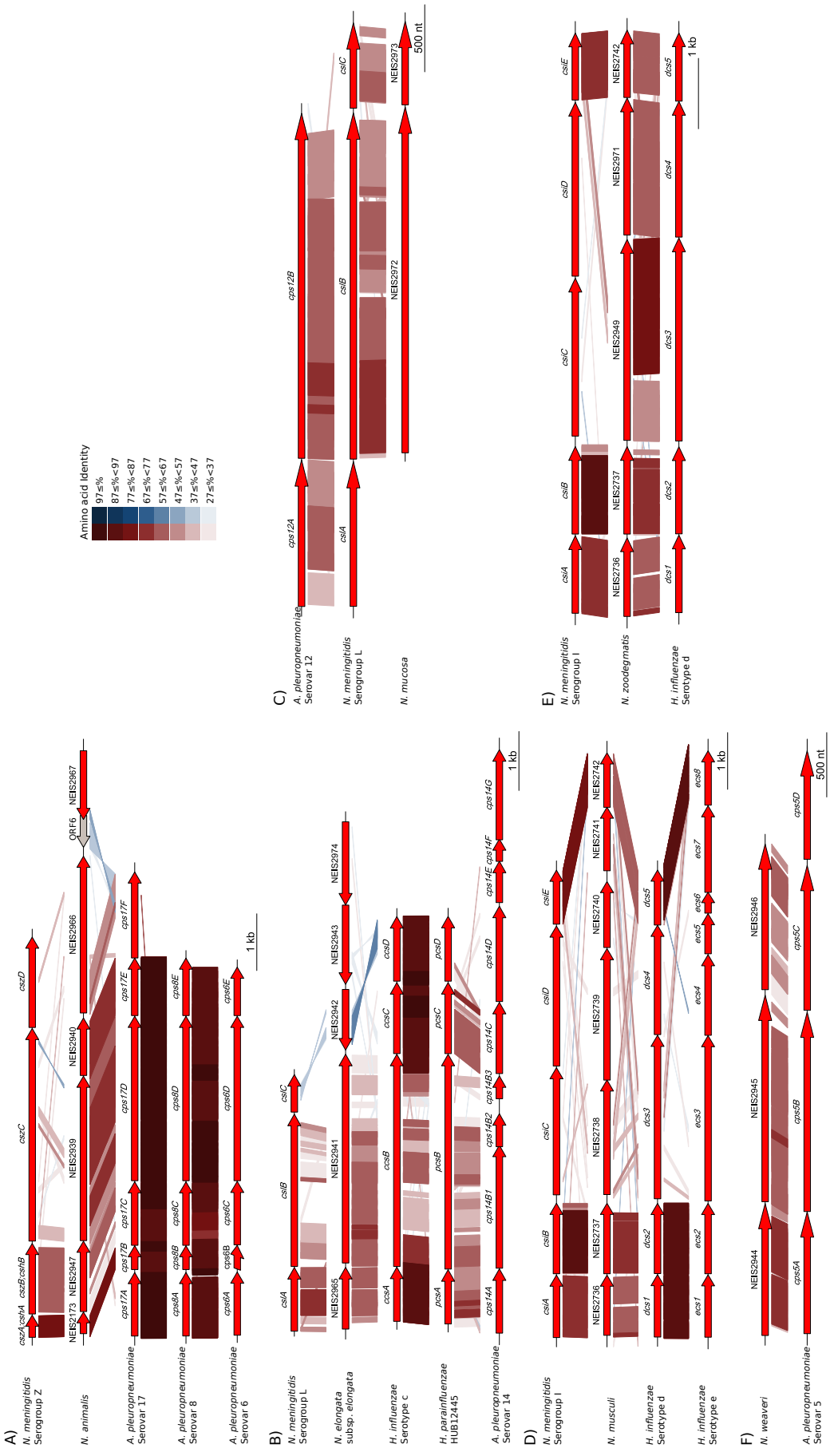

Figure 2.5: Comparisons with characterised capsule synthesis genes. Pairwise tBLASTx comparisons of previously characterised capsule synthesis sequences and putative region A sequences from A) *N. animalis*, B) *N. elongata* subsp. *elongata*, C) *N. mucosa*, D) *N. musculi*, E) *N. zoodegmatis* and F) *N. weaveri*. Colour intensity indicates sequence identity. Blue indicates inversion.

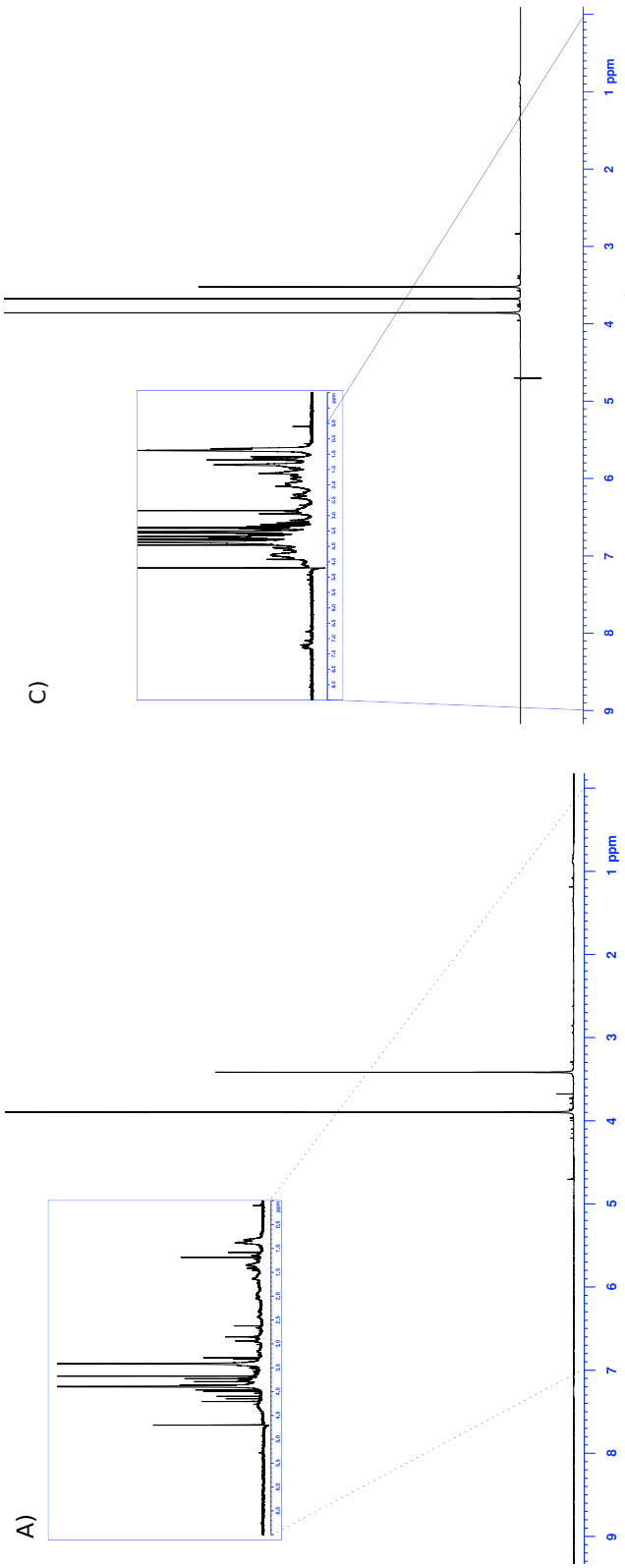

Figure 2.6: ^{1}H NMR spectra of capsular polysaccharides. Full and zoomed spectra of A) CCUG 30802T (*N. elongata* subsp. *nitroreducens*) B) CCUG 2043T (*N. elongata* subsp. *elongata*) C) CCUG 24918 (*N. subflava* lacking NEIS2942, NEIS2941 version 1) D) 12007_2012 (*N. subflava*, NEIS2941 version 2). Samples contain contamination. The CCUG 2043T spectrum is consistent with the presence of a glycerol residue. The peak at 2 ppm in CCUG 24918 is consistent with acetylation. NMR was performed by Dr Tim Rudd at the NIBSC.

2.3.2 Preliminary extraction of polysaccharides from non-pathogenic *Neisseria* species

The extraction method succeeded in extracting polysaccharides from *Neisseria* species, and some preliminary results were obtained from NMR analyses. Samples were determined to have negligible levels of LOS, according to a LAL assay. Nevertheless, the samples were not found to be as pure as desired (Figure 2.6). It was apparent that samples still contained some contamination from the anion exchange buffer, even after dialysis. Nevertheless, it was possible to detect that the NMR spectrum of CCUG 30802T (*N. elongata* subsp. *nitroreducens*) was consistent with the presence of a glycerol residue in the sample (Figure 2.6A). Also detectable was the fact that, of all the samples analysed, only 12007_2012 (*N. subflava*) contained a peak at 2 ppm, consistent with an acetyl group (Figure 2.6C). The method was unsuccessful in extracting polysaccharide in the quantities required for ^{13}C NMR (~1 mg).

2.4 Discussion

Among the *Neisseria*, the polysaccharide capsule has been considered to be a virulence factor unique to *N. meningitidis*. Although region B and C genes had been identified previously in an isolate of *N. subflava*, in the absence of further evidence at the time this was attributed to an HGT event from *N. meningitidis* to *N. subflava*, facilitated by a DNA uptake sequence in *ctrA* [64]. In this chapter, homologues of all the conserved region B and C genes in multiple NPN species from across the genus have been identified, with accompanying putative capsule synthesis loci.

There is good evidence to support the existence of region B and C genes, with the potential to encode for capsule export proteins, within 13 NPN species (Table 2.1). These genes are highly conserved across all encapsulated meningococci, and among Gram-negative bacteria that express an ATP transporter-dependent capsule [78, 76]. The amino acid sequence identity between meningococcal *ctrE* and *ctrF* and

the *E. coli* homologues *kpsC* and *kpsS* is 68% and 70% respectively [73], and it has been demonstrated that these orthologues can be functionally interchanged between species [106]. Therefore, the moderately to very high amino acid sequence identity (52-99%) between proposed *ctrABCDEF* genes in NPN species and homologues in meningococci (Table 2.1), and the fact that in most cases they co-localise into regions equivalent to meningococcal *cps* regions B and C (Figure 2.2) – although there is reduced certainty with respect to the putative *ctrF* in *N. canis* and *N. bacilliformis*, because it is not co-localised with *ctrE* in these isolates – is consistent with their coding for capsule export proteins similar to those belonging to *N. meningitidis*.

Putative region A genes were more difficult to characterise, since different complements of enzymes are required to produce the structurally distinct capsules that vary within and between species [76]. With the exception of *N. bacilliformis*, the putative region A genes were consistently adjacent to region C genes at one end, and in many cases were also flanked or closely flanked by region B (Figure 2.2). Region C is also co-located with region A in the meningococcal *cps* [78], and the three homologous regions are frequently co-located in capsule loci belonging to species from other genera [196], so in most NPN species the organisation of the putative capsule genes is consistent with a capsule locus.

Three species groups were identified that shared homologous putative region A sequences (Figure 2.4). There is little synteny across the genomes of these species, so the comparable co-localisation of putative capsule gene homologues further supports the inference that they form an arranged capsule locus. Homologous genes were not always assigned to the same NEIS locus; by default, PubMLST has very conservative rules on nucleotide sequence identity and length equivalency relative to existing alleles when adding new allele sequences to a NEIS locus. These could be overruled in the future if there is biochemical evidence indicating that the genes are functionally identical. The putative region A sequences also resembled or partially resembled

capsule biosynthesis genes characterised in other species (Table 2.2, Figure 2.5), further supporting their predicted function.

The putative capsules of Group 1 species *N. animalis*, *N. bacilliformis* and *N. elongata* subsp. *nitroreducens* are predicted to resemble teichoic acid polymers, based on their resemblance to *A. pleuropneumoniae* type I capsule serovars 17, 8 and 6 [220], and *N. meningitidis* serogroup Z [78] (Figure 2.5A). Preliminary structural analyses in *N. elongata* subsp. *nitroreducens* were consistent with this prediction, indicating that a glycerol residue was present in the capsule polysaccharide of CCUG 30802T (Figure 2.6A). There is decreased certainty about the three extra genes in *N. animalis*, two of which only partially resembled *cps17F* from *A. pleuropneumoniae* (Figure 2.5A), and the last of which only partially resembled capsule synthesis genes from *H. influenzae* serotype c and *A. pleuropneumoniae* serovar 14 (Table 2.2). Nevertheless, this is not necessarily unusual, since many of the *A. pleuropneumoniae* type I capsule synthesis regions contain unique genes [220], and so these genes have been tentatively annotated as putative region A genes at this time. Higher confidence can be assigned to the rest of the genes in the putative region A sequences of these species.

The putative capsules of Group 2 species *N. elongata* subsp. *elongata*, *N. subflava*, *N. oralis* and *N. mucosa* are predicted to consist of oligosaccharide polymers with phosphodiester linkages, which is a unifying property of *N. meningitidis* serogroup L, *H. influenzae* serotype c and *A. pleuropneumoniae* type II capsule serovars, including serovars 12 and 14 [220, 89, 202], with which these putative region A sequences shared some homology (Figure 2.5B). It is not possible to make any more specific predictions about the structure of *N. elongata* subsp. *elongata*, *N. subflava* or *N. oralis* capsules without further analyses. There is extra uncertainty about NEIS2943 and NEIS2974, which share no homology with capsule synthesis genes from any of the aforementioned species. Nevertheless, NEIS2943 partially resembles capsule synthesis genes from *N. meningitidis* serogroups H and X. NEIS2974 contained a predicted

AMP-binding domain according to Pfam-A, which does not preclude a role in capsule synthesis. Neither of these genes, or indeed any putative region A genes, are present in the WGS of *N. subflava* genomes lacking region C, consistent with a non-essential function, in common with all capsule loci. The putative region A of *N. mucosa* more closely resembles meningococcal *cslBC* (Figure 2.5C). *cslA* has been shown to have no polymerase activity in the synthesis of the serogroup L capsule [89]. Therefore, although the putative *N. mucosa* region A is lacking *cslA*, it is possible that it produces a capsule similar to the trimeric structure of *N. meningitidis* serogroup L, consisting of trimeric repeats of →3-β-D-N-Acetylglucosamine-1→3-β-D-N-Acetylglucosamine-1→3-α-D-N-Acetylglucosamine-1-phosphate [89]. Overall, the sequences comparisons support the identity of the suspected region A sequences in these species.

The putative capsules of Group 3 species *N. musculi*, *N. dentiae*, *N. canis* and *N. zoodegmatis* are predicted to resemble structures consisting of N-acetylmannosaminuronic acid, N-acetylguluronic acid and/or N-acetylglucosamine, as seen in *N. meningitidis* serogroups I and K, and *H. influenzae* serotypes d and e [78, 202] (Figure 2.5D/E). *H. influenzae* serotypes d and e share the same three homologues found in all group 3 NPN putative region As, but the genes in between are not homologues (Figure 2.5D). Despite this, both *H. influenzae* serotypes d and e consist of N-acetyl-D-glucosamine and N-acetyl-D-mannosaminuronic acid, differing only by the linkages [202]. Therefore, there is a reasonable chance that the capsule structures of the group three NPN species are relatively similar, despite possessing different complements of putative region A genes, particularly *N. dentiae*, *N. musculi* and *N. canis*, which differ by just one gene (Figure 2.4). It is possible that the putative *N. zoodegmatis* capsule has the same structure and perhaps even linkages as *H. influenzae* serotype d, given that it contains homologues of all the serotype d capsule synthesis genes (Figure 2.5E). Nevertheless, the region A sequences of *N. meningitidis* serogroups I and K are almost identical, and yet produce O-acetylated N-acetylmannosaminuronic acid homodisaccharide [94], and O-acetylated alternating N-acetylmannosaminuronic

acid and N-acetyl-guluronic acid capsule structures respectively, although is has been suggested that this difference may be an artefact of the extraction conditions [78]. Whilst the particulars of the capsule structures are hard to predict, comparative genomics supports the identity of the putative region A sequences in these species.

N. weaveri did not possess any putative capsule synthesis genes homologous to those found in other NPN species. It did contain three of the the four capsule synthesis genes belonging to *A. pleuropneumoniae* type III capsule serovar 5 (Figure 2.5F). The commonality of *A. pleuropneumoniae* type III serovars is only that they are glycosidically linked sugar polymers [220]. *N. weaveri* has one of the best examples of typical capsule locus organisation (Figure 2.2), so combined with its similarity to *A. pleuropneumoniae* serovar 5, there is good support for the identity of its putative region A.

Region A annotations were consistent with the potential for more than one capsular group within *N. elongata*, and possibly *N. subflava* and *N. oralis* (Figure 2.4). *N. subflava* isolates possessed one of two versions of the gene NEIS2941, predicted to code for a glycosyltransferase (Table 2.2), whose nucleotide sequences were approximately 72% identical (Figure 2.3). This is comparable to the divergence between the polysialyltransferase genes *csb* and *csc*, which are 64% identical, and determine the structural differences between meningococcal serogroups B and C respectively [78, 84]. In addition to *N. meningitidis*, the presence of multiple capsule groups or types is commonplace among Gram-negative bacteria, including *E. coli*, *H. influenzae* and *M. haemolytica* [73, 148, 149]. In *E. coli*, over 80 structurally different capsules exist, some of which are associated with specific pathologies, or are only expressed at certain temperatures [73]. The range of niches exploited by *E. coli*, including different hosts and tissues, as well as free-living environments, may be responsible for this diversity [76]. Individual *Neisseria* species do not demonstrate such a wide exploitation of niches, but it has been demonstrated that different

species have tropisms for specific nasopharyngeal sites [16]. There maybe more subtle intra-species niche variation, which requires an optimum capsule, or no capsule at all.

The conclusions thus far have relied on genomic data. The great advantage of genomic data is the ability to characterise putative capsule genes with high throughput, as demonstrated by the novel tool CapsuleFinder, which is claimed to have been able to search 2500 bacterial genomes and identify whether a capsule may be present [203]. Such an approach is particularly effective in a thoroughly researched field such as capsule biology, where conserved genes are well defined, and many synthesis genes have previously been identified [76]. Consistent with the genetic findings described in this chapter, indicating the potential for capsule expression in NPN species, results of India ink and alcian blue staining indicate the existence of a high molecular weight polysaccharide surrounding *N. musculi* bacteria [221]. Further biochemical analyses are required to confirm capsule expression, fully determine which genes are involved in capsule production and accurately predict capsule structure.

In preliminary attempts to address questions about the structure of NPN capsules, an extraction method involving anion exchange chromatography was used. Traditional methods of capsule extraction have included a protein removal step using phenol chloroform [222], which is highly hazardous. The method described here, based on anion exchange chromatography [218], successfully extracted capsular polysaccharides from NPN, which were shown to have LOS contamination below detectable limits. Nevertheless, the anion exchange buffer was not successfully removed, resulting in artefacts in ^{1}H NMR spectra (Figure 2.6).

Subsequent methanol precipitation of anion exchange fractions performed by Dr Maharjan successfully removed this contamination, and corroborated the initial findings described in the results of this chapter, including the presence of glycerol in CCUG 30802T (*N. elongata* subsp. *nitroreducens*), predicted from sequence comparisons. Spectra also supported an acetyl group in 12007_2012 (*N. subflava*) (Figure 2.7), consistent with the presence of NEIS2941 in 12007_2012 but not CCUG

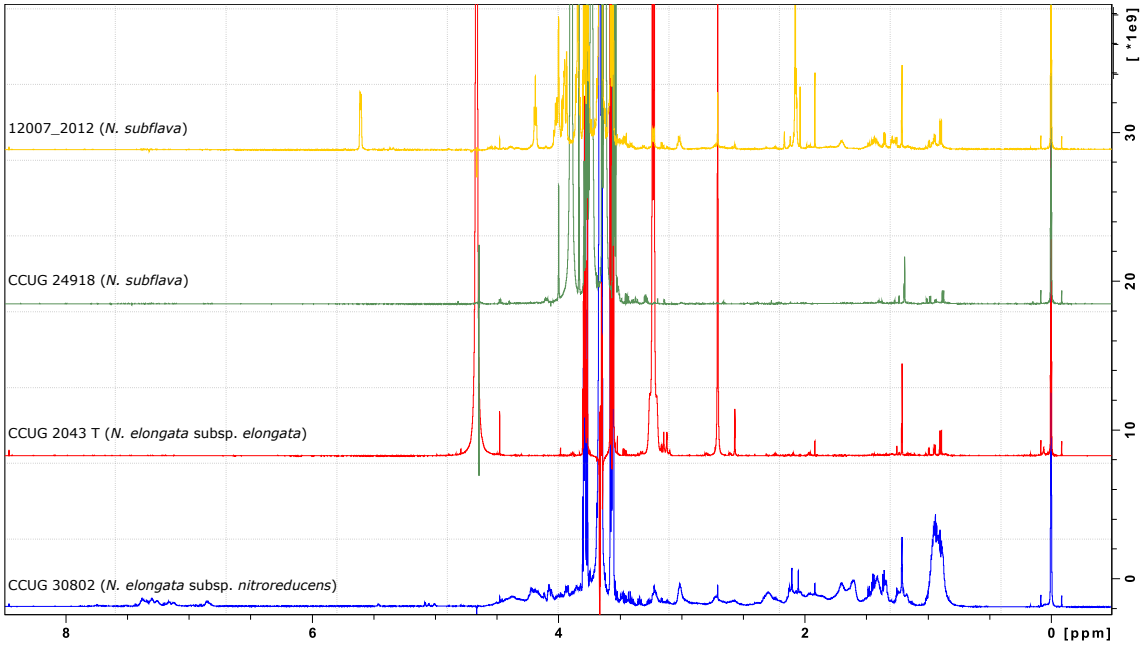

Figure 2.7: Repeat ^{1}H NMR spectra. Overlay of NMR spectra for CCUG 30802T (blue), CCUG 2043T (red), CCUG 24918 (green) and 12007_2012 (yellow) following methanol purification step. Additional purification was carried out by Dr Sunil Maharjan, and NMR was performed by Dr Tim Rudd, at the NIBSC. Reproduced with permission and included for discussion purposes.

24918; NEIS2491 is homologous to *H. influenzae* serotype c gene *ccs4* (Table 2.2), a predicted O-acetyltransferase. The cleaner spectra also show that the CCUG 30802T polysaccharide is distinctly different from that of the other *N. elongata* isolate (Figure 2.7), as predicted by sequence comparisons (Figure 2.3D). These initial findings are consistent with predictions from genomic analyses.

^{13}C NMR can provide much more detailed information about the structure of polysaccharides [223]. The method described here did not produce capsule polysaccharide in the quantities required for ^{13}C NMR. This may have been a result of capsule fragmentation during the boiling step, which resulted in small capsule fragments being lost during the filtration step. A possible solution to this step may be to use sodium deoxycholate/EDTA solubilisation of cells to increase yields, and to use a chemically defined media containing ^{13}C to reduce the quantity of polysaccharide required for ^{13}C NMR (S. Maharjan, personal communication). In this way, more detailed structural analyses could be obtained.

On the balance of evidence from genomic data, including the comparable synteny between the putative *cps* in most NPN species and the meningococcal *cps* (Figure 2.2), and the high homology of several putative capsule synthesis genes to those of *N. meningitidis* and other species (Figure 2.5, Table 2.2), combined with initial biochemical analyses, the genes identified were most likely to function in capsule synthesis. The discovery of capsule genes in non-pathogenic bacteria is not unprecedented, with a similar finding in the mitis group streptococci overturning the assumption that capsule production was unique to the pathogenic *S. pneumoniae* [147]. Capsules have also been identified in *Bacteroides* species that inhabit the human gut [145, 146], and many of the non-pathogenic members of the microbiota surveyed by CapsuleFinder contained matches for capsule genes [203]. In common with many virulence factors, including the type IV pilus [65, 64], the meningococcal capsule might be better described as a "host adaptation" [64] or "niche" [224] factor, with effects on pathogenic potential being incidental.

Nevertheless, uniquely in *N. meningitidis*, certain capsule serogroups permit survival in the bloodstream, and so the evolution of the meningococcal capsule is a key factor for understanding the emergence of meningococcal pathogenesis. The discovery of capsule genes in NPN species challenges models presented previously, hypothesising that *N. meningitidis* must have acquired a capsule by HGT [72, 111, 199]. Nevertheless, it does not necessarily preclude the acquisition of capsule genes in *N. meningitidis* by HGT. Notably, capsule genes have still not been identified in any isolates belonging to the monophyletic group that contains *N. cinerea*, *N. lactamica*, *'N. bergeri'*, *N. polysaccharea*, *N. gonorrhoeae* and *N. meningitidis*, with the exception of *N. meningitidis* itself (Figure 2.1). This uneven distribution of capsule genes must be explained by acquisition and/or loss [225, 226]. For the capsule to be present only in certain *N. meningitidis* isolates, the capsule genes must either have been lost independently as many as six times, or lost once in a common ancestor of the monophyly and re-acquired in *N. meningitidis*. The latter is the

most evolutionarily parsimonious explanation, although a scenario between these two extremes is also possible. These models have been further explored in Chapter 3 using phylogenetic analysis of WGS.

2.4.1 Conclusions

In this chapter, NPN species have been shown to possess genes that are likely to encode polysaccharide capsules. Capsules are predicted to be produced *via* ABC-transporter dependent capsule polymerisation similarly to *N. meningitidis*, but putative region A genes are predicted to synthesise novel and distinct capsule structures. This discovery of capsule genes in NPN species reflects the polysaccharide capsule's role in asymptomatic colonisation and transmission, an important stage in meningococcal epidemiology. Nevertheless, it also raises questions about the evolutionary history of the meningococcal capsule, which will be addressed in the next chapter.

3

Re-examining models of capsule acquisition in *N. meningitidis*

This work was published in Clemence *et al.*, 2019, Wellcome Open Res 4:99 (`https://doi.org/10.12688/wellcomeopenres.15333.2`) [227], and presented as a poster at the 15$^{\text{th}}$ Congress of the European Meningococcal and Haemophilus Disease Society, Lisbon, 2019. Work presented here was carried out by the candidate.

3.1 Introduction

Prior to the work in this thesis, it was hypothesised that the meningococcal *cps* was acquired as a result of an HGT event, resulting in the duplication of region D [111, 199, 72, 143, 193]. In the absence of the data collected in Chapter 2, which identified capsule sequences in NPN species, this hypothesis depended on the understanding that *N. meningitidis* was the only member of the *Neisseria* to possess a capsule, inferring that it was most likely acquired by HGT. In light of the discovery of putative capsule sequences in NPN species, the capsule acquisition hypothesis must be evaluated more critically. Chapter 2 has already proposed that the presence of a capsule in meningococci may be attributable to an HGT event, based on the most evolutionarily parsimonious explanation. The aim of this chapter is to use

WGS analyses to further evaluate the credibility of HGT events taking place within the meningococcal *cps*, and whether sequence data are consistent with a *de novo* acquisition of the meningococcal capsule.

N. meningitidis is a highly competent bacterium [57], which co-exists in the nasopharynx with many other species from the *Neisseria* and other genera [16]. In addition to frequent recombination events within meningococcal populations, there have been several accounts of HGT from NPN species to *N. meningitidis*, as well as from non-*Neisseria* species to *N. meningitidis*, including genes associated with virulence and antibiotic resistance [228, 229, 230, 231, 64, 232, 201]. Therefore, the acquisition of meningococcal *cps* sequences from other species is biologically plausible.

Several pieces of evidence have been put forward in support of *de novo* acquisition of the capsule locus, besides the observation, refuted in Chapter 2, that the capsule was unique to *N. meningitidis* among the *Neisseria*. The low G+C content within region A of the capsule locus has been cited as evidence that the capsule may have been acquired by HGT [78, 83, 193], a property that has been associated with islands of HGT [195]. Nevertheless, low G+C content has previously been described in the capsule synthesis regions of *E. coli*, *Haemophilus sputorum* and *A. pleuropneumoniae* [196, 197, 198], and now also in the putative capsule synthesis regions of NPN species in Chapter 2. Therefore, G+C content may not directly inform on the recent evolutionary history of the meningococcal capsule. Further evidence put forward includes: the deduced common ancestry between *cps* sequences within regions A, B and C, and homologous sequences in other Gram-negative bacteria [143, 72, 202]; and the organisation of genes within the *cps* [111, 72, 143], which differs from the typical arrangement in Gram-negative bacteria, wherein genes involved in capsule synthesis and export are contiguous.

As described in previous chapters, the meningococcal *cps* is comprised of the functional regions A, B and C [78]. Also co-located in *cps* is region D, which consists of *galE* and *rfbABC*, and region D', a duplicated version of region D. Region E consists

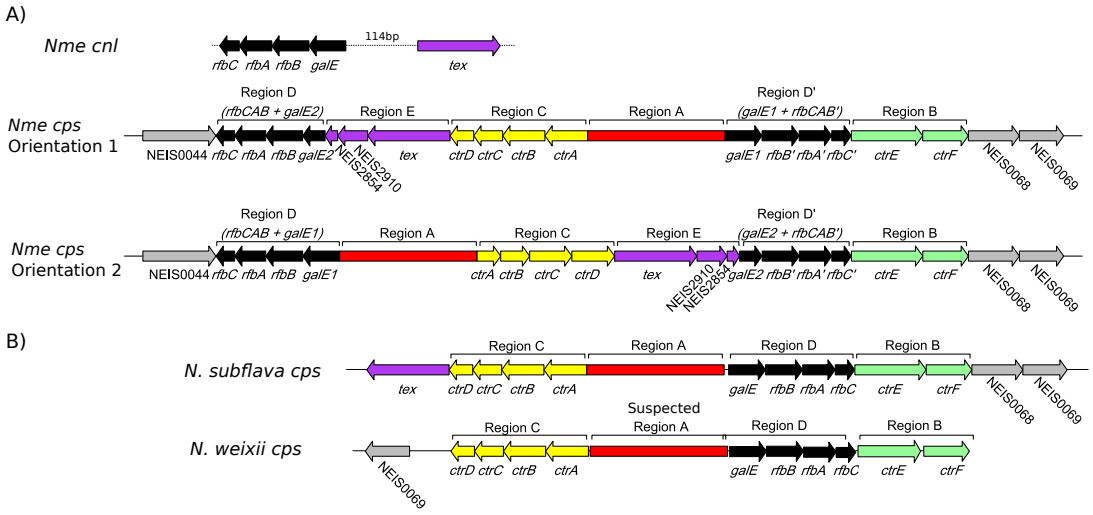

Figure 3.1: Comparisons of *N. meningitidis*, *N. subflava* and *N. weixii cps*. Organisation of genes within A) the two orientations of the meningococcal *cps* and B) the *N. subflava* and *N. weixii cps*.

of the putative transcriptional accessory protein *tex*, a modification methyltransferase, and a truncated adenine-specific methyltransferase. Flanking the 3' end of region B is an additional hypothetical gene designated NEIS0068 [111]. The only genes that are unique to encapsulated meningococci, among *N. meningitidis*, are those found within regions A, B and C, and NEIS0068. There is also only one copy of region D in capsule null *N. meningitidis*. Whilst regions A and C are contiguous, region B is separated from regions A and C by region D', or region D' and region E, depending on the orientation of the *cps* (Figure 3.1A).

There is dynamic inversion of genes within the capsule locus between *galE1*, and the truncated gene *galE2*, giving rise to two capsule orientations (Figure 3.1A) [111]. These inversions make it challenging to distinguish between region D and region D'. There is sufficiently high sequence identity between *rfbABC* and *rfbABC'* [78] that the automated genome annotator within BIGSdb is unable to distinguish between them, so they must be manually identified and extracted based on their context within the *cps*. *galE1* alleles are divergent from *galE2*, which is consistently truncated at the 5' end [111].

Bartley *et al.* proposed that the phylogenetic distribution and functionality of *galE1* and *galE2* could be explained by the process of an *en bloc* transfer of the entire capsule locus from a donor species into modern meningococcal clones, originally in Orientation 1 (Figure 3.1A) [111]. If the model presented by Bartley *et al.* is true, a number of other features would also be predicted. Firstly, region B and region C sequences would be expected to resemble the donor. Secondly, *rfbABC'* would be expected to more closely resemble the donor, whilst *rfbABC* would resemble capsule null *rfbABC*, as has been described for *galE1*, but in strains MC58, FAM18 and NMB, the regions were almost identical [111]. Nevertheless, it is possible that the intra-locus re-orientation events could result in total swapping of one version of region D for another, in certain lineages. Therefore, to address the aim of this chapter, *cps* sequences were investigated in *N. meningitidis* WGS from a range of clonal complexes and capsule groups.

3.2 Methods

3.2.1 Isolate collection

Meningococcal WGS data with good *cps* assembly were obtained from the Meningitis Research Foundation Meningococcus Genome Library (MRF-MGL) (consisting of UK disease-associated isolates), the meningococcal 107 global collection project (consisting mostly of disease-associated isolates) [54], and the UKMenCar4 project, a UK carriage dataset collected by the University of Oxford [233]. All of these datasets are available on PubMLST. Meningococcal genomes were chosen at random from the datasets to provide up to one clonal complex/capsule genogroup combination from both carriage and disease where available. WGS from additional public PubMLST isolates M01-240355 [234] and WUE2594 [235] were chosen to include cc213 and cc5, respectively. Additional *cps* sequence data from isolates with capsule genogroup E, L, W, X or Z were retrieved from GenBank [236], originating from characterisation

of meningococcal capsule serogroups [78]. WGS data from representative isolates of other *Neisseria* species were sourced from PubMLST, including the novel species *N. weixii* (strain 10022, GenBank accession number CP023429.1).

WGS from non-*Neisseria* species were sourced from GenBank and chosen for the presence of homologous region D and/or region B and C sequences, based on the result of BLASTn searches, using sequences from CA41967 as a query. Only species possessing homologous sequences for the full complement of genes in the region of interest were included. Representatives of the following species were included in analyses: *Actinobacillus succinogenes* (strain 130Z, GenBank accession number CP000746.1), *A. pleuropneumoniae* (AP76, CP001091.1), *Aggregatibacter actinomycetemcomitans* (D11S-1, CP001733.2), *Bibersteinia trehalosi* (USDA-ARS-USMARC-189, CP006955.1), *Vibrio vulnificus* (NBRC 15645, CP012881.1), *Glaesserella* sp. (15-184, CP023057.1), *Actinobacillus porcitonsillarum* (9953L55, CP029206.1), *Haemophilus influenzae* (18010, FQ312006.1), *Kingella kingae* (KW1, LN869922.1) and *Actinobacillus suis* (NCTC12996, LT906456.1). A full list of isolates can be found in Table A.2.

3.2.2 Annotation of capsule loci

The majority of the *Neisseria* genomes in PubMLST had previously been fully annotated manually for *cps* genes *rfbABC* (NEIS0045-7), *galE* (NEIS0048), *tex* (NEIS0059), the pseudo cytosine methyltransferases NEIS2854 and NEIS2910, *ctrABCDEF* (NEIS0055-8,66-67), and flanking genes NEIS0044, NEIS0068 and NEIS0069, where present. Genomes in which one or more of these genes had not been annotated were queried using the BLASTn-based scanning tool in PubMLST; if a relevant gene was identified, this was tagged in the WGS data and an appropriate allele designation set. *N. meningitidis* sequences from GenBank had a fully annotated *cps* [78]. Predicted orthologues of *cps* genes were identified in non-*Neisseria* species using BLASTn.

The meningococcal genomes possessing *cps* were investigated to determine whether the locus was in Orientation 1 or Orientation 2 (Figure 3.1A). *galE1* and *galE2* were distinguished according to the nomenclature used by Bartley *et al.* [111], in which the truncated form of *galE* within *cps* was designated *galE2*, and the full length form as *galE1*. If the capsule locus spanned more than one assembled contig, the orientation was assumed based on the co-localisation of the relevant genes and regions.

3.2.3 Neighbor-net analyses

Nucleotide sequence data for region C genes *ctrABCD*, or predicted orthologues, were extracted from meningococci, representative non-meningococcal *Neisseria*, and non-*Neisseria* species for which all genes were present. The same was done for region B genes *ctrEF*.

Nucleotide sequence data for *rfbABC+galE* or predicted orthologues were extracted from capsule null meningococci, representative non-meningococcal *Neisseria*, and non-*Neisseria* species for which predicted orthologues of all four genes were present. In meningococci *cps*, *rfbABC+galE2* and *rfbABC'+galE1* were extracted manually and analysed independently, regardless of the capsule orientation.

For each set of genes, amino acid sequences were deduced in MEGA X [208] and aligned using Muscle [237], correcting for frameshift mutations where applicable, and manually trimmed to give a final nucleotide sequence alignment. *ctrABCD*, *ctrEF* and *rfbABC* sequences were each concatenated, and each alignment was loaded into SplitsTree4 [61]. Phylogenetic split networks were generated using the neighbor-net algorithm [238]. Groups were identified based on a balance between maximising edge weighting, whilst minimising contradictory splits. Full length *galE* and *galE1* homologues (i.e. not including *galE2*) were also analysed separately.

3.2.4 Recombination analysis using BOOTSCANing

Based on the results of neighbor-net analyses, meningococcal isolates CA41967, Z2491, α707, WUE171 and 1.02397.V were chosen for further investigation of the whole capsule locus and its flanking regions, and compared to sequences from ST42119 (capsule null *N. meningitidis*), NJ9703 (*N. subflava*) and 10022 (*N. weixii*), with USDA-ARS-USMARC-188 (*Birbersteinia trehalosi*) included as an outgroup. *rfbCAB+galE2*, *ctrDCBA*, *galE1+rfbBAC*, *ctrEF* and NEIS0069 were analysed separately, since capsule null *N. meningitidis* does not contain *ctrDCBA* or *ctrEF*. Amino acid sequences were deduced in MEGA X, aligned using Muscle, correcting for frameshift mutations where applicable, and manually trimmed to give a final nucleotide sequence alignment. After alignment, sequences were orientated to be in the same direction as they would be relative to NEIS0044 in Orientation 1 of the meningococcal *cps* (Figure 3.1A), where NEIS0044 is in the forward orientation, then concatenated into the groups described above for separate analyses.

Each concatenated set of aligned sequences was loaded into the Recombination Detection Programme 4 (RDP4) [239]. Recombination was assessed in each *cps*$^+$ meningococcal isolate using manual BOOTSCANing [240], with ST42119 (capsule null *N. meningitidis*), NJ9703 (*N. subflava*), 10022 (*N. weixii*) and USDA-ARS-USMARC-188 (*B. trehalosi*) as reference sequences. Neighbor-joining trees were used with the Jukes-Cantor substitution model [241] and 100 bootstraps. Bootstrap support below 70% was disregarded. In order to minimise false breakpoints that may occur due to high sequence identity, appropriate window size was determined by testing CA40160, CA41628 and GL40098 (other capsule null *N. meningitidis*), and OX42005 (other *N. subflava*), which were not expected to have recombinant capsule sequences. Window size was set at 400 bp for *rfbABC+galE1/2*, 200 bp for regions B and C, and 250 bp for NEIS0069. Step size was set at 10% of window size.

3.2.5 Maximum-likelihood phylogenies

ctrEF sequences were extracted from all *Neisseria* isolates possessing predicted region B orthologues. *B. trehalosi* was included as an outgroup. The first 774 bp of *ctrE*, which were suspected to be recombinant, were removed. Amino acid sequences were deduced in MEGA X, aligned using Muscle, correcting for frameshift mutations where applicable, and manually trimmed to give a final nucleotide sequence alignment, then concatenated. A maximum-likelihood phylogeny was generated in PhyML (v3.1) [214] with 100 bootstraps using the GTR+I+G method, determined to be the best fit by jModelTest (v2.1.10) [215].

ctrD sequences were extracted from all *Neisseria* isolates possessing predicted region C orthologues. *B. trehalosi* was included as an outgroup. Sequences were aligned to generate a phylogenetic tree as described above.

3.3 Results

3.3.1 Gene annotations

The distribution of capsule export genes *ctrABCDEF*, region D genes *rfbABC+galE* and *tex* was consistent with that described in Chapter 2, and all 11 genes were annotated in the novel species *N. weixii*, which was also observed to contain homologues of the putative region A from the *N. animalis cps* (Figure 3.1B). Additionally, the 345 bp pseudo cytosine methyltransferase NEIS2854 was present within all cps^+ meningococci, as well as the capsule null isolate ST42119 (cc198); NEIS2854 is a truncated version of the 1008 bp locus NEIS2725, which was present in gonococcal isolates and one strain of *N. polysaccharea* (CCUG 4790) only. The pseudo cytosine methyltransferase NEIS2910 was found within all cps^+ meningococci, ST42119, all gonococci, and CCUG 4790 genomes. The hypothetical gene NEIS0068, which flanks region B of the capsule locus, was identified within genomes from

all *cps*$^+$ *N. meningitidis* and *N. subflava* isolates, but no other *Neisseria* species or capsule null meningococci.

3.3.2 Regions B and C of the meningococcal *cps* are mosaic

Phylogenetic analyses of each of meningococcal *cps* regions B and C, along with predicted orthologues from other *Neisseria* species and proteobacteria, were consistent with the presence of recombinant region C sequences in the meningococcal *cps*.

Neighbor-net analysis revealed a well-supported split that grouped region B sequences from meningococci and *N. subflava* together (Figure 3.2A). There was also some support for a contradictory split grouping region B sequences from 24 meningococci with *N. weixii*. BOOTSCANing in 200 bp windows of CA41967, one of the 24, was consistent with mosaic Region B sequences, with at least the first 774 bp of *ctrE* from CA41967 having support for *N. weixii* as the nearest neighbour grouping, before switching to *N. subflava* (Figure 3.3).

A split network generated by neighbor-net analysis of region C sequences contained several contradictory splits that either grouped meningococcal region C sequences with *N. subflava* or the novel species *N. weixii*, both with high support relative to the rest of the network. There was no well-supported split separating all three into a single group (Figure 3.2B). This was corroborated by BOOTSCANing in a subset of five isolates: all five *cps* sequences scanned had good bootstrap support for *N. subflava* as the nearest neighbour grouping for at least the first 440 bp of region C (at the 3' end of *ctrD*) (Figure 3.3). Across the rest of the region, there was greater bootstrap support for *N. weixii* as the nearest neighbour, although the signal was noisy. In Z2491, the signal for *N. subflava* extended for 2379 bp (comprising *ctrD*, *ctrC* and much of *ctrB*) before switching to *N. weixii*. This was consistent with a split which separated Z2491, as well as WUE2594, into a group with *N. subflava* (Figure 3.2B).

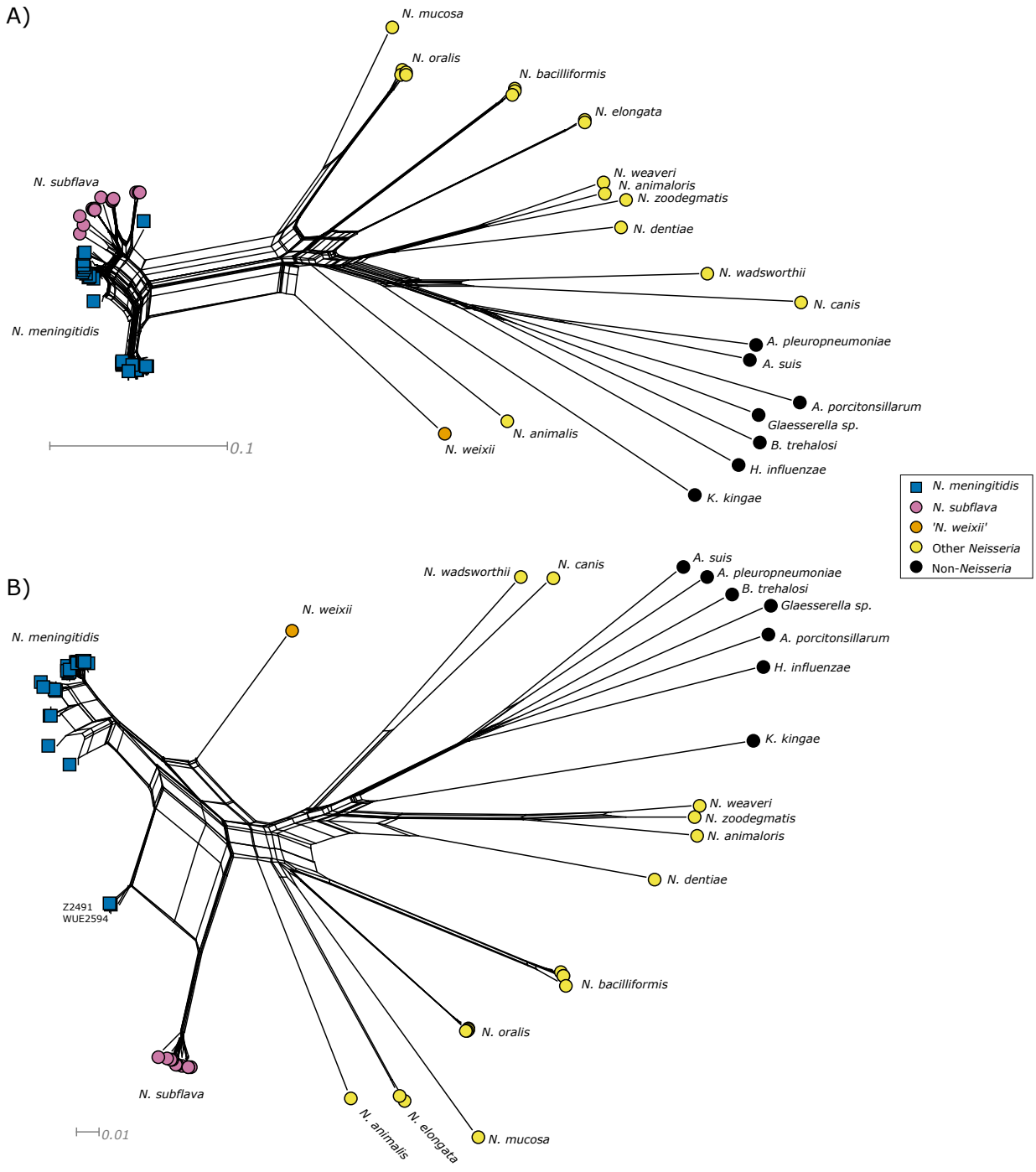

Figure 3.2: Split networks of region B and C sequences. Split networks generated using neighbor-net analysis of concatenated, aligned nucleotide sequences of A) region B and B) region C genes. Edges represent splits that support the separation of two groups in the network, with the length of the line representing the weight of the split. Increasing number of parallel edges represents contradictory splits.

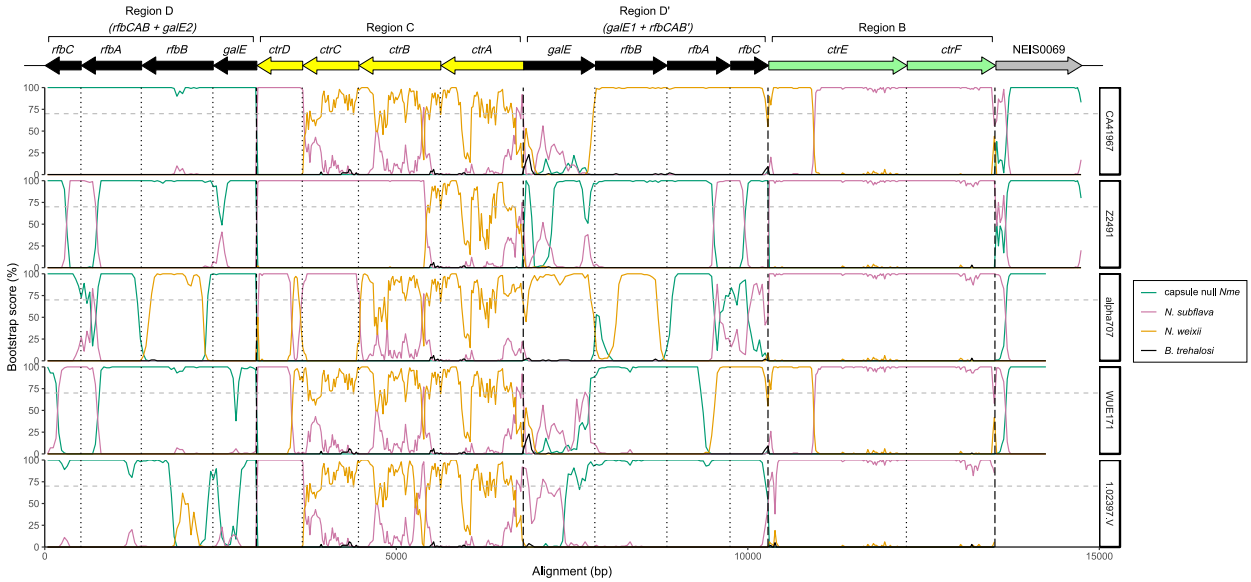

Figure 3.3: Recombination analysis of the meningococcal *cps*. BOOTSCANing of concatenated *cps* sequence alignments from CA41967, Z2491, α707, WUE171 and 1.02397.V, with ST42119 (capsule null *N. meningitidis*, green), NJ9703 (*N. subflava*, pink) and 10022 (*N. weixii*, yellow) included as potential parent sequences, and USDA-ARS-USMARC-189 (*B. trehalosi*, black) included as an outgroup. The y axis indicates the number of bootstrap trees out of 100 supporting a given isolate as the nearest neighbour to the query sequence in window sizes of 400 bp (*rfbCAB+galE1/2*, 200 bp for regions B and C, and 250 bp for NEIS0069. Step size set to 10% of window size. Vertical short dashed lines represent gene boundaries, and vertical long dashed lines represent separate analyses.

3.3.3 Meningococcal capsule export sequences are nested within homologous sequences in *N. subflava*

There is uncharacteristically high nucleotide sequence identity between *ctrD*, *ctrEF* and NEIS0068 in meningococci, and predicted orthologues in *N. subflava*, compared to flanking genes and the rest of the genome, even after accounting for differences that may be explained by the relative conservation of individual genes, determined by analysis relative to *N. oralis* (Table 3.1).

A maximum-likelihood phylogeny of region B, excluding the first 774 bp of *ctrE*, which was determined by BOOTSCANing to be potentially recombinant in some genomes, revealed that region B sequences from meningococci were nested within homologous sequences from *N. subflava* (Figure 3.4A). The diversity of these sequences was much lower in *N. meningitidis* (mean p-distance among isolates

Table 3.1: Nucleotide sequence identity between *N. subflava* and *N. meningitidis*. Comparison of nucleotide sequence identity of loci in the *cps* and flanking regions between *N. meningitidis* and *N. subflava*, relative to *N. oralis*.

Locus	Region	*N. subflava* vs. *Nme* ID(%)	*N. oralis* vs. *Nme* ID(%)	*N. subflava* vs. *N. oralis* ID(%)
NEIS0043	flanking	75.67	75.13	84.87
NEIS0044	flanking	78.36	74.68	74.44
NEIS0059	E	92.70	92.35	93.98
NEIS0058	C	97.24	81.57	81.41
NEIS0057	C	76.82	74.81	79.45
NEIS0056	C	72.78	70.96	76.50
NEIS0055	C	74.40	71.60	73.28
NEIS0066	B	84.97	76.01	78.37
NEIS0067	B	94.60	66.51	68.15
NEIS0068	flanking	93.70	-	-
NEIS0069	flanking	81.60	68.07	69.14

0.017) than *N. subflava* (mean p-distance among isolates 0.040). Similarly, suspected non-recombinant *ctrD* sequences from meningococci were nested within predicted homologous sequences belonging to *N. subflava* (Figure 3.4B).

3.3.4 Region D' of the *cps* locus is not a duplication of meningococcal region D

As described previously [111], sequenced meningococcal isolates possessed the *cps* locus in either Orientation 1 or Orientation 2 (Figure 3.1A). *galE1* was distinguished from *galE2* by the fact that the latter is consistently truncated at the 5' end.

Phylogenetic analyses of *rfbBAC'*, along with predicted orthologues from other *Neisseria* species and proteobacteria, were consistent with acquisition of *rfbBAC'* sequences by HGT from another *Neisseria* species (Figure 3.5A). In WGS from 23 isolates, neighbor-net network splits supported the the separation of *rfbBAC'* into a group with homologous sequences from the novel species *N. weixii*. This relationship was further supported by BOOTSCANing of concatenated *rfbBAC'* sequences from CA41967, which showed high bootstrap support for *N. weixii* as the nearest neighbour grouping across most of the region (Figure 3.3).

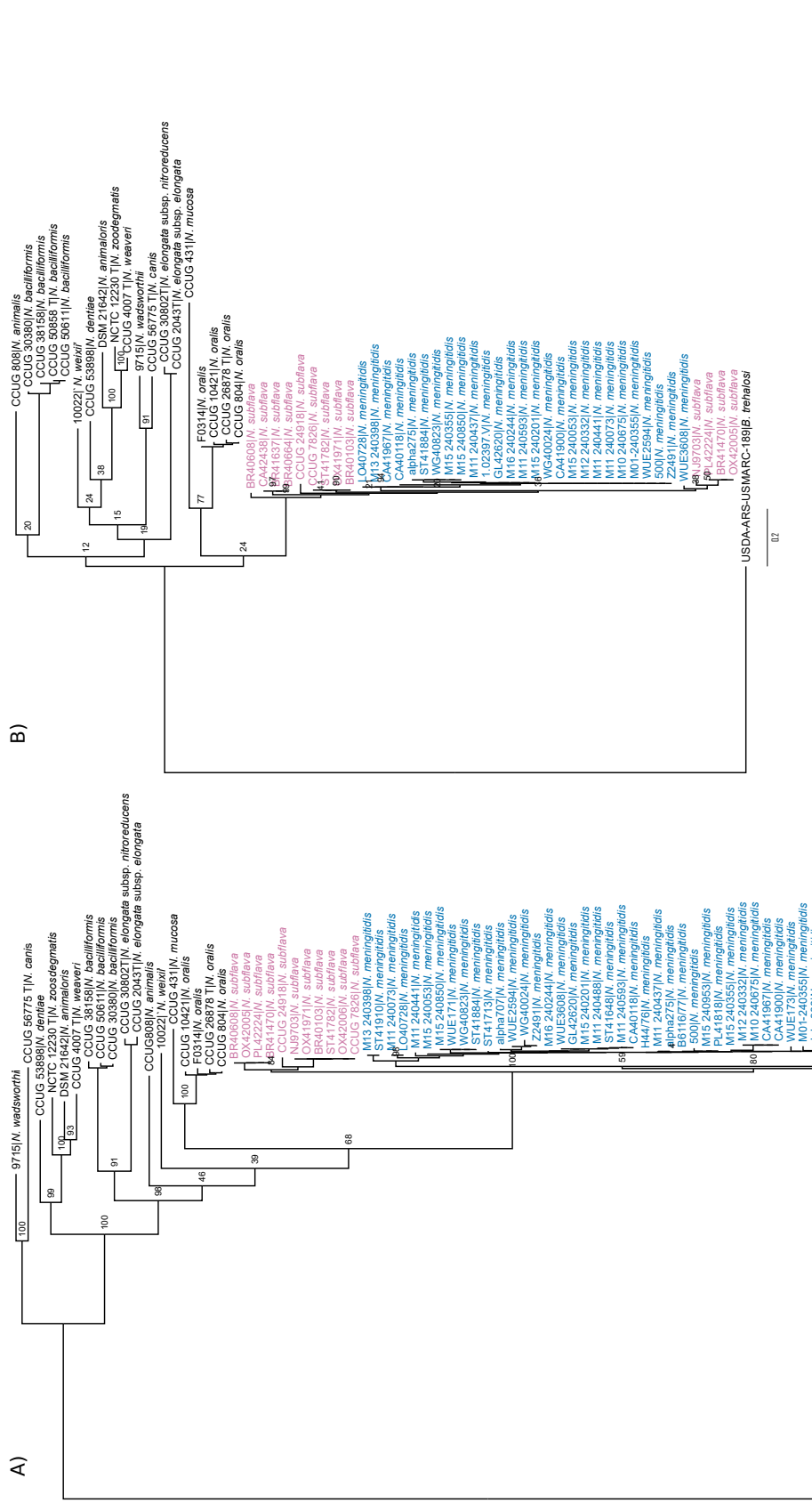

Figure 3.4: Phylogenies of meningococcal and *N. subflava* export sequences. Maximum-likelihood phylogenies using the GTR+I+G substitution model, generated from concatenated, aligned nucleotide sequences of A) region B genes *ctrEF* (NEIS0066 was trimmed at the 5' end to remove potentially recombinant sequences) and B) *ctrD*. In both phylogenies, *B. trehalosi* was included as an outgroup. *N. meningitidis* blue, *N. subflava* pink.

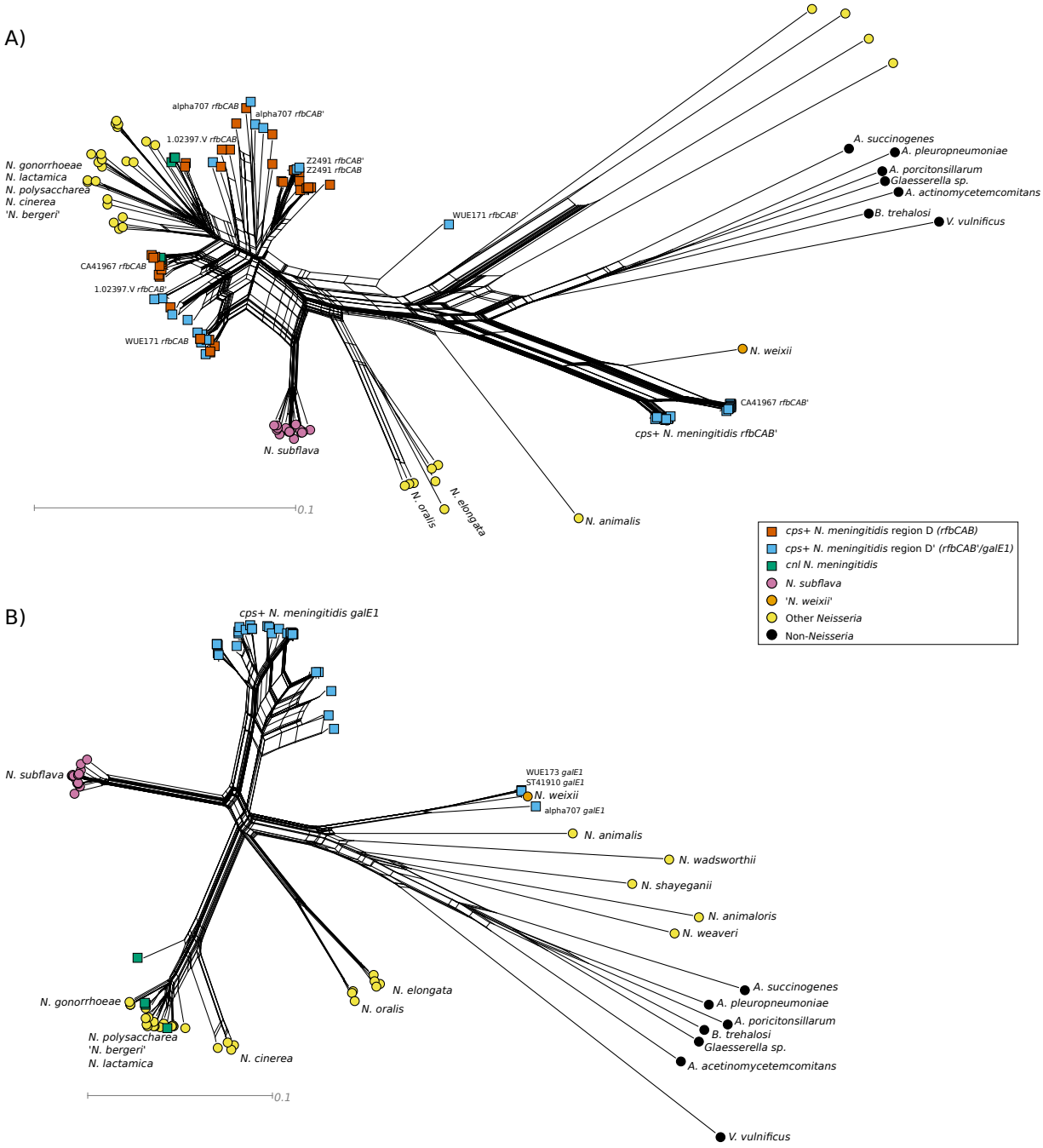

Figure 3.5: Split networks of region D sequences. Split networks generated using neighbor-net analyses of concatenated, aligned nucleotide sequences of A) *rfbABC* genes, with both *rfbABC* and *rfbABC'* extracted from *cps*$^{+}$ meningococci, and B) full length *galE* or *galE1* (truncated *galE2* alleles not included). Edges represent splits that support the separation of two groups in the network, with the length of the line representing the weight of the split. Increasing number of parallel edges represents contradictory splits.

There was a drop in bootstrap support for any reference across *galE1*, possibly indicating the absence of a good representative reference sequence. An appropriate sequenced reference could not be identified using a split network generated from aligned *galE*, *galE1* and predicted orthologues (Figure 3.5B). A split did support the separation of *galE1* sequences from α707 (capsule group E), WUE173 (capsule group Z) and ST41910 (capsule group Z) into a group with predicted orthologous sequences in *N. weixii*, consistent with BOOTSCANing analyses of α707 (Figure 3.3).

There were no highly weighted splits supporting the grouping of *rfbABC'* sequences from the remaining 16 isolates with any other species, and a high degree of reticulation was present within the split network, which can be indicative of recombination (Figure 3.5A). This was consistent with BOOTSCANing results in sequences from chosen isolates (Figure 3.3). *rfbBAC'+galE1* sequences from Z2491 contained tracts similar to both capsule null *N. meningitidis* and *N. subflava*. α707 and WUE171 contained *rfbBAC'* sequences similar to both *N. meningitidis* and *N. weixii*, with a drop off in bootstrap support across *galE1*. The 1.02397.V *rfbBAC'* sequences were similar to capsule null *N. meningitidis*, again with a drop-off in bootstrap support for any reference across *galE1*.

In contrast to *rfbBAC'*, the split network supported the grouping of all concatenated *rfbCAB* sequences with capsule null meningococci, but again there was a lot of reticulation (Figure 3.5A). Sequences from chosen isolates were investigated further using BOOTSCANing. BOOTSCAN results of *rfbABC+galE2* sequences from Z2491 and WUE171 were consistent with a recombination event involving *N. subflava* (Figure 3.3). In 1.02397.V, there was only bootstrap support for capsule null *N. meningitidis*, although there was a drop off in bootstrap support through parts of *rfbB* and *galE2*. According to the split network, α707 *rfbABC* sequences could be separated into a group with the capsule null sequences, but mosaic signals were identified in the sequences consistent with a recombination event involving a species more closely related to *N. weixii*. Sequences from CA41967 could be less

ambiguously separated into a group with capsule null *rfbABC* sequences, consistent with BOOTSCANs, which did not demonstrate recombination in this region.

3.4 Discussion

The origin of the meningococcal capsule is the subject of a debate that has been further obscured by the identification of putative capsule loci in NPN species in Chapter 2. In this chapter, evidence for recombination in the region B and C sequences, which are absent in capsule null meningococci, and region D sequences, which are only duplicated in encapsulated meningococci, has been examined. Sequence data have been further analysed to evaluate the cogency of hypotheses postulating that the meningococcal *cps* was acquired *de novo* by HGT.

Analyses of region B and C sequences are consistent with a donation of capsule export gene sequences from *N. subflava* to the meningococcal *cps*. Following the identification of recombinant sequence data (Figure 3.2, Figure 3.3), further analyses demonstrated that non-recombinant tracts within regions B and C were phylogenetically nested within homologous *N. subflava* sequences (Figure 3.4). This phylogenetic pattern would only be expected to occur if the true donor was a member of *N. subflava*, rather than another species closely related to *N. subflava*, and is consistent with the high sequence identity between homologues of these genes in *N. meningitidis* and *N. subflava* relative to *N. oralis*, compared to other *cps* and flanking genes (Table 3.1). HGT is more likely to occur between closely related species, since higher sequence identity facilitates homologous recombination. *N. subflava* is also widely carried by humans, and it has previously been suggested that strains of the close relative of *N. meningitidis*, *N. gonorrhoeae*, may have acquired *penA* genes associated with antimicrobial resistance from *N. subflava* through HGT [242]. The tissue tropisms of *N. meningitidis* and *N. subflava* are slightly different, with *N. subflava* isolated more commonly from the buccal cavity than the nasopharynx [16], but both are frequently

isolated from carriage studies using pharyngeal swabs [19, 158]. Therefore, HGT of *cps* sequences between *N. subflava* and *N. meningitidis* is biologically conceivable.

Sequence analyses also indicate that a second species donated sequences within region C of all meningococci analysed, and region B of several meningococcal isolates, resulting in highly mosaic loci (Figure 3.2, Figure 3.3). The novel species *N. weixii* was identified as a possible candidate, described as being isolated from the Tibetan Plateau Pika (*Ochotona curzoniae*) in Qinghai Province, China (GenBank accession CP023429.1) [21]. *N. weixii* contained sequences homologous to the putative region of A of *N. animalis*, a guinea-pig associated species [20], between its region C and region D homologues (Figure 3.1B), but not a full complement of *N. meningitidis* serogroup H or Z capsule synthesis genes. *N. animalis* is also the known species most closely related to *N. weixii*. Epidemiological interaction between *N. weixii* and *N. meningitidis* is unlikely, since the pika is a member of the *Lagomorpha*, found in alpine meadows. An as-yet-unidentified human-associated *Neisseria* species closely related to *N. weixii* could be the donor of mosaic region B and C sequences to *N. meningitidis*.

In many of the meningococcal genomes analysed, results were also consistent with HGT of duplicated region D sequences. In these genomes, *rfbABC'+galE1* sequences were divergent from *rfbABC+galE* sequences belonging to capsule null meningococci, either in whole or in part (Figure 3.5). As within regions B and C, BOOTSCANing analyses were consistent with either *N. subflava* or something related to *N. weixii* being the candidate donor of the divergent *rfbABC'* sequences (Figure 3.3). The *galE1* sequences of α707 (cc254, capsule group E), ST41910 (cc1157, capsule group Z) and WUE173 (capsule group Z) also grouped with *N. weixii* (Figure 3.5B). In all other meningococcal genomes, *galE1* was divergent from *N. subflava*, *N. weixii* and capsule null *N. meningitidis*, and the donor of this sequence may be another as yet unidentified *Neisseria* species (Figure 3.5B). The divergence between *galE1*, *galE2* and capsule null *galE* sequences, including capsule group E and Z outliers, corroborates with data from Bartley *et al.* [111]. They proposed that these divergences

are consistent with a model in which the meningococcal capsule was acquired *en bloc* in *N. meningitidis*. Whilst the data discussed so far indicates that the meningococcal *cps* is highly mosaic, having undergone HGT with as many as three other *Neisseria* species, they are not sufficient to support such a model of capsule acquisition.

The hypothesis that the meningococcal cps was acquired *de novo* by a previously capsule null meningococcal recipient, as a result of an HGT event, has been proposed several times [111, 199, 72, 83, 143]. The *de novo* acquisition of a capsule by a pathogen has been previously described in *H. influenzae* which, similarly to *N. meningitidis*, consists of variants both with and without a capsule. The *H. influenzae* capsule locus *cap* was proposed to have been donated by HGT from *H. sputorum*, although *H. sputorum* may actually be a member of another genus from the *Pasteurellacae* family [197]. A similar event in *N. meningitidis* would have had important consequences on the epidemiology of meningococcal disease, since the possession of a capsule is almost always necessary for IMD [23, 24].

The sequence identity between capsule export sequences in *N. meningitidis* and *P. multocida* formed the basis of a hypothesis invoking donation of the capsule by a member of the *Pasteurellacae* family, in the absence of further *Neisseria* WGS data at the time [72]. Data analysed in this chapter demonstrate that capsule export gene sequences are more closely related to homologous sequences from non-pathogenic *Neisseria* species (Figure 3.2), which raises the question as to whether the capsule was simply inherited by descent. Nevertheless, as discussed at the end of Chapter 2, the distribution of capsule genes among close relatives of *N. meningitidis* may have resulted from a loss of capsule in a common ancestor of these species, followed by reacquisition in *N. meningitidis*.

The validity of an *en bloc* acquisition model, as proposed by Bartley *et al.* [111], has been further tested using a genome dataset containing a wide diversity of meningococcal clonal complexes and capsule genogroups. BOOTSCANing analyses consistently supported *N. subflava* as the nearest neighbour grouping at both ends of

the capsule locus, and perhaps into NEIS0069 (Figure 3.3), and these sequences in *N. meningitidis* were nested within homologous sequences from *N. subflava* (Figure 3.4). The *en bloc* model also postulates that the donor capsule locus was arranged ←NEIS0059←Region C-Region A→Region D→NEIS0068→NEIS0069→, equivalent to *N. meningitidis* Orientation 1 [111], which is the same arrangement as characterised in *N. subflava* in Chapter 2 (Figure 3.1), and the flanking gene NEIS0068 has only been found in *cps*$^+$ meningococci and *N. subflava*. These observations are consistent with, but not proof of, an *en bloc* donation of capsule from *N. subflava*, as opposed to a member of another genus, to a capsule null meningococcal clone, with subsequent recombination events with at least two other *Neisseria* species.

This model does not account for the origins of region A. Neither *N. weixii* nor *N. subflava* possess putative region A sequences that are highly comparable to those found in meningococcal capsule serogroups, although they do share some homologous sequences. Recombination events that include region A capsule synthesis genes, which result in capsular serogroup switching, have been repeatedly observed within meningococcal populations [103, 98, 243]. The results presented here raise the question as to whether meningococcal serogroup diversity may have arisen through capsule switching with other species. The serogroup B capsule is structurally equivalent to that of *E. coli* K1, *M. haemolytica* A2 and *Moraxella nonliquifaciens* [76]. There is homology between the synthesis regions of these capsules [244], and also between other *N. meningitidis* capsule synthesis genes and homologues in other species [78], but amino acid sequence identity above 70% has not been reported. It is unlikely that meningococcal region A sequences would have diverged so much if this acquisition event happened after *N. meningitidis* diverged from other *Neisseria* species. Sequence data did not show evidence of non-*Neisseria* sequences having been donated to meningococcal region C or region D' (Figure 3.2, Figure 3.3). The grouping of *galE1* sequences within a split network indicates that an as-yet unsequenced *Neisseria* species may have donated sequences to *N. meningitidis*, in an

HGT event that may also have involved region A (Figure 3.5B). It may be informative to characterise putative region A homologues of any new species identified that are closely related to *N. weixii*, which may shed further light on this question.

Another unanswered question concerns the breakpoints of the *en bloc* model. Sequence analyses are consistent with homologous recombination within or adjacent to NEIS0069, but evidence for an illegitimate breakpoint between *galE2* and *tex*, as has been postulated [111], is yet to be described. This illegitimate breakpoint, proposed to have resulted in the truncation of *galE2*, is incompatible with the presence of pseudo cytosine methyltransferases downstream of *tex*, which are not present in *N. subflava*. There is no evidence for homologous recombination, since all capsule null isolates studied possessed *tex* in the opposite orientation to that seen in Orientation 1 of encapsulated meningococci. An attempt was made to incorporate *tex* into recombination models, but the conservation of *tex* sequences among *Neisseria* (Table 3.1) inhibited the detection of meaningful signals. Previous studies of *tex* have been inconclusive [245, 199]. Given that the pseudo cytosine methyltransferase were found in an example of a capsule null isolate, and complete copies of both methyltransferases are present in *N. gonorrhoeae*, further exploration of region E may provide additional evidence with respect to the accuracy of the *en bloc* acquisition model.

An alternative explanation to *de novo* acquisition models is that a duplication of meningococcal region D, and perhaps the whole capsule locus, previously existed in the recipient organism, and HGT has only resulted in mosaicism within these native sequences. This alternative explanation could account for the fact that some isolates still contain sequences resembling capsule null meningococci in region D' (Figure 3.3, Figure 3.5). On the other hand, this may be explained by dynamic inversions of the capsule locus (Figure 3.1A) [111]. If inversions involve multiple recombination events with different break points, sequences within the two regions could become unlinked, making it difficult to trace their evolutionary history, and erasing evidence of an acquisition event within region D' sequences. This problem

may be further exacerbated by capsule switching, wherein region A sequences, as well as flanking *cps* sequences, are exchanged horizontally among meningococci [103, 97]; the breakpoints of these events are inconsistent. With these issues, by the nature of the question [246], and the relatively small size of WGS datasets compared to global *Neisseria* populations through time, it would be difficult to prove beyond doubt that the meningococcal capsule was acquired *de novo* by a capsule null clone, unless a meningococcal isolate were identified with a complete *N. subflava* capsule locus, requiring an absence of further recombination events. This illustrates a major challenge of analysing retrospective evolutionary events using contemporary WGS data.

This challenge is made further complicated by the high rate of recombination across the genome among meningococci, which impedes within species comparative evolution. In an otherwise highly clonal population, encapsulated meningococci resulting from HGT would be expected to belong to a single clade. An attempt has been made to map capsule acquisition onto a *N. meningitidis* phylogenetic tree based on core genes, which reveals a divergence between uncencapsulated clonal complexes cc198 and cc1136 and other meningococcal clonal complexes [199]. This could be interpreted as consistent with a capsule acquisition event in a common ancestor of the latter clade, although cc53 (also unencapsulated) is nested within encapsulated isolates, indicating capsule loss in this model. Nevertheless, as discussed in Introduction §1.2.3, it may not be reasonable to try and resolve clonal complex relationships using bifurcating tree models. Another study, which used a neighbor-net model based on genome rearrangement distances to resolve recombination effects, indicated that the strain α14 (cc53-*cnl*) could be separated from encapsulated meningococci to group with *N. gonorrhoeae* and *N. lactamica*, although the weighting of this split was small [193]. This interpretation has been disputed as an over-interpretation of a marginal split weighting, with a process of infectious genetic exchange argued to be more likely [56], although this argument

cited the existence of cc53-C isolates, which although identified in serogroup testing, have never been identified by more reliable capsule genogrouping. Unravelling the complete evolutionary process that led to the modern-day *cps* may not be possible using only contemporary meningococcal genomes, if evolutionary processes within branches has been lost through recombination.

3.4.1 Conclusions

WGS data are consistent with a model whereby the meningococcal capsule locus was acquired by a capsule null meningococcal clone *en bloc* in an HGT event from a single donor, most likely *N. subflava*. Subsequent homologous recombination events with at least two other species resulted in a highly mosaic locus. Nevertheless, these data are insufficient to rule out an alternative model, in which native meningococcal capsule existed prior to undergoing HGT with *N. subflava* and other species. It is also possible that serogroup diversity of meningococcal populations increased as a result of cross-species HGT events. Characterisation of putative capsule genes of newly isolated *Neisseria* species, particularly those isolated from humans, may provide new insights into the complex evolutionary history of the meningococcal capsule locus.

4

Exploring E to B capsule switching associated with invasive meningococcal disease

This work was presented as a poster at the ASM (American Society for Microbiology) Microbe conference, Atlanta, GA, 2018. Work presented here was carried out by the candidate.

4.1 Introduction

In the previous chapter, a model for capsule acquisition was discussed, involving HGT events between *N. meningitidis* and other *Neisseria* species. This event may have been responsible for a change in the epidemiology of the meningococcus, resulting in IMD in human populations. HGT of *cps* sequences within *N. meningitidis* populations is also an important concept in meningococcal epidemiology. Specifically, the exchange of *cps* region A sequences between meningococcal isolates from different serogroups results in capsule switching. Capsule switching has reported to have taken place from A to X, B to C, C to B, W to B, Y to W, Y to B and C to W [97, 98, 99, 100, 101, 102, 103]. These capsule groups are all associated with IMD.

Another typing characteristic associated with meningococcal disease, in addition to capsule group, is the clonal complex. The clonal complexes represent groups of meningococci that share similar allelic profiles for the seven MLST house-keeping genes [50, 51]. Some clonal complexes are associated with certain capsular groups, for instance cc23 and cc167 meningococci typically express a serogroup Y capsule [96], although these associations can be broken up by capsule switching. They are also associated with other genotypic and phenotypic profiles, for example clonal complexes cc41/44, cc11, cc32, cc8 and cc269 are most frequently isolated from IMD in Europe [53]. Unlike cc41/44, whose high incidence in IMD may be partially explained by high carriage rates, cc11 is over-represented in IMD compared to carriage collections [247, 248, 249]. It is not entirely clear whether this is because cc11 meningococci are intrinsically more invasive than meningococci from other clonal complexes, or because they are highly transmissible, but rapidly cleared from carriage [247]. It is also unclear to what extent clonal complex associations with disease are misrepresented because of the association of certain capsule groups with clonal complex. In this chapter, capsule switching from E to B within clonal complexes that are not typically associated with disease are investigated, and the association between capsule, clonal complex and disease is explored.

The serogroup E meningococcal capsule is not associated with IMD [247]. There are occasional reports of IMD cases from which serogroup E meningococci were isolated, but these are typically associated with complement-deficient or otherwise immunocompromised patients [130, 250, 251]. The serogroup E capsule structure consists of alternating 2-acetamido-2-deoxy-D-galactose and 3-deoxy-D-manno-octulosonic acid residues [109], and in bactericidal assays, serogroup E meningococci have been demonstrated to have higher sensitivity to human serum than IMD-associated capsule serogroups [134, 130]. In contrast, serogroup B meningococci are frequently isolated from cases of IMD, and at the time of writing are responsible for the highest burden of IMD cases in Europe [252]. The serogroup B capsule structure consists

of α(2->8)-linked sialic acid polymers [81]. Sialic acid facilitates immune evasion in many bacteria [253], and genetic deletion of the capsule greatly increases the serum sensitivity of serogroup B meningococci [254].

The contrasting disease-associated behaviours of serogroup E and B meningococci are reflected in 21st Century UK carriage and IMD isolate collections, collected for the UKMenCar4 project [233] and the MRF-MGL (`https://pubmlst.org/bigsdb?db=pubmlst_neisseria_mrfgenomes`) respectively (Figure 4.1). In the three epidemiological years 07/2013-06/2016, the MRF-MGL sequenced 1590 meningococcal genomes isolated from IMD cases in the UK, of which three (0.19%) were capsule group E. In the same period, 761 (48%) capsule group B genomes were deposited. In comparison, the UKMenCar4 project, which sequenced 1425 carried meningococcal isolates from UK teenagers 2015-2016, isolated 180 (13%) capsule group E isolates and 346 (24%) capsule group B isolates. Group E isolates most frequently belonged to cc60 (34/180, 19%) or cc1157 (129/180, 72%). Within cc60 and cc1157, E was the predominant capsule group at 100% and 80% of isolates respectively. A further 6.3% of cc1157 isolates had group B capsules. 10 cases of cc60-B and 10 cases of cc1157-B IMD were isolated in the UK 07/2013–06/2016. The first aim of this chapter is to determine whether disease-associated cc1157-B and cc60-B isolates emerged as a result of capsule switching.

At the time of writing, there is no vaccine available that targets the capsular polysaccharide of serogroup B meningococci, as is the case for serogroups A, C, W and Y. Two so-called "MenB" vaccines have been developed that target sub-capsular surface proteins, which have been designed to cover the most prevalent IMD-associated serogroup B meningococci, but do not specifically target the serogroup B capsule [255, 256]. Bexsero® is an OMV based vaccine containing the protein antigens fHbp (variant 1), NadA (variant 2/3), NHBA and PorA [121]. Trumenba® combines two families of lipidated fHbp [122]. There is expected to be a degree of cross-reactivity against meningococci that do not have exact matches to the vaccine antigens fHbp and

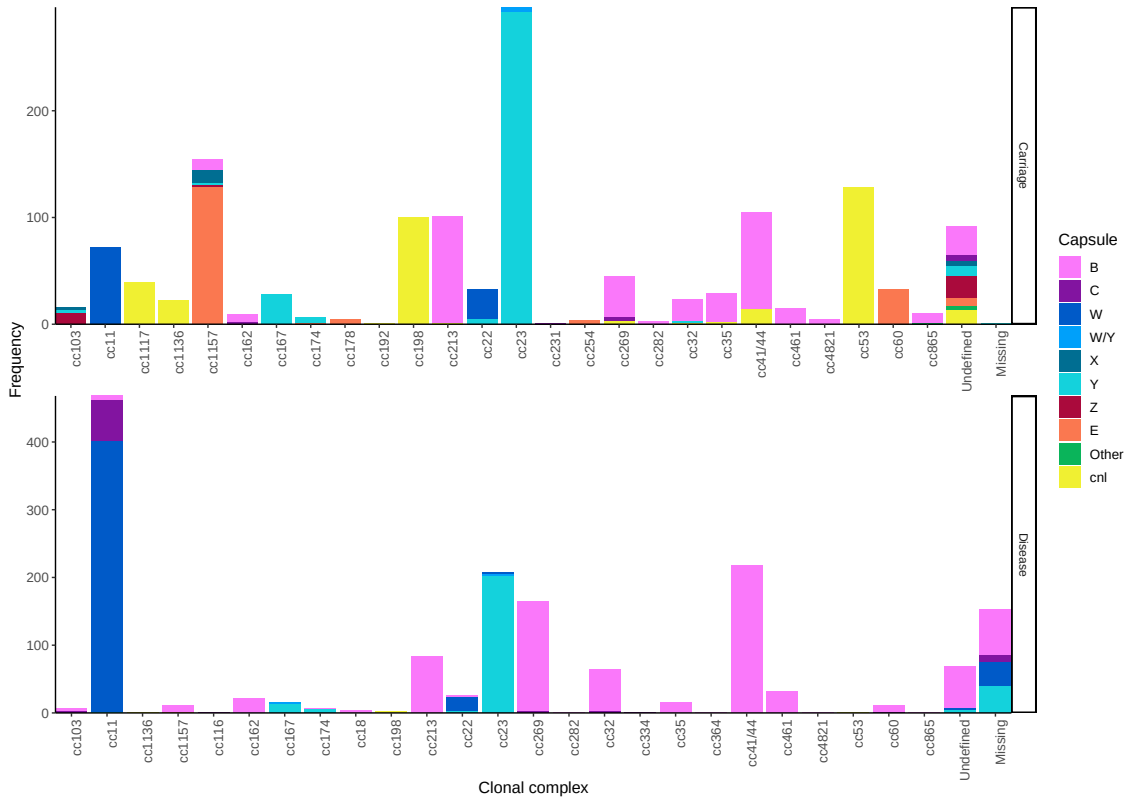

Figure 4.1: Comparison of carriage and disease distributions. Distribution of clonal complexes and capsule groups within UKMenCar4 carriage dataset collected 2014-2015 and MRF-MGL disease dataset collected in epidemiological years 07/2013-06/2014–07/2015-06/2016. Capsule group assigned on the basis of genogroup. "Other" capsule group includes groups L and H. Isolates with an undefined clonal complex have been assigned a sequence type that has not been shown to cluster with any existing clonal complex. One or more MLST loci assembled too poorly to assign an allele number in isolates with missing clonal complex data.

NadA [123, 124, 125, 122]. Since Bexsero® and Trumenba® do not target any factor specific to serogroup B meningococci, they may also have cross-reactivity with other serogroups [126]. Predictions about the cross-reactivity of fHbp and NadA from amino acid variant profiles have been made using data from bactericidal antibody assays and the Meningococcal Antigenic Typing System (MATS) [125, 123, 122, 257, 258, 259]. The second aim of this chapter is to assess whether the “MenB” vaccines are predicted to protect against emerging cc1157-B and cc60-B meningococci, and at the same time potentially have an impact on carried capsule group E meningococci.

Clonal complexes cc1157 and cc60 are not considered to be hyperinvasive lin-

eages, since they are isolated from IMD comparatively infrequently compared to cc41/44, cc32 and cc11 [260]. Nevertheless, meningococci from these clonal complexes predominantly possess a carriage-associated group E capsule. The overall numbers are smaller, but the observation of 10 cc60-B meningococci isolated from IMD, despite the fact that no cc60-B meningococci were collected in UKMenCar4, may indicate an over-representation of cc60-B in disease similar to cc11 (Figure 4.1). This could be a consequence of a polygenic profile favouring invasive disease, which is usually masked by a non-invasive capsule.

As many as 177 genes have been proposed to contribute to the virulence phenotype of *N. meningitidis*, many of which are also found in NPN species [64]. An association between specific variants and disease has been established for a much smaller subset, including genes involved in iron acquisition and serum survival, and the type IV pilus and MDA bacteriophage: HmbR is involved in haemoglobin acquisition during invasion [261], and the accessory gene *hmbR* is over-represented among disease-associated meningococci and the hyperinvasive lineages [66]; fHbp negatively regulates the alternative complement pathway [262], and of the three fHbp variants, variant 1 is associated with disease whilst variants 2 and 3 are associated with carriage [67]; the type IV pilus is involved in colonisation and disease, and class II of the pilus sub-unit PilE is associated with invasive meningococcal lineages [68, 69]; and finally, presence of the so-called MDA bacteriophage, encoded by an 8 kb genetic island, is also associated with meningococcal disease [70, 71]. The third aim of this chapter is to investigate factors within the cc60 genome that might be consistent with a tendency towards IMD.

4.2 Methods

4.2.1 Isolate collection

WGS data from meningococcal isolates were obtained from the PubMLST database. For exploration of serogroup-clonal complex associations, carriage isolates were sourced from the UKMenCar4 carriage study collected from 2014-2015, and disease isolates were sourced from the MRF-MGL, epidemiological years 07/2013-06/2014–07/2015-06/2016. This is a reasonably balanced dataset containing 1590 disease- and 1425 carriage-associated isolate genomes. For WGS analyses, the MRF-MGL dataset was expanded back to epidemiological year 07/2010-06/2011. A full list of isolates used in WGS analyses can be found in Table A.3. Isolate WGS data were manually checked to confirm the capsular genogroup, which was the definition used for capsule group assignment.

4.2.2 Allelic and protein variant profiling

The BIGSdb genome comparator tool was used to generate a 1605 locus *N. meningitidis* cgMLST v.1 scheme profile [54] for the cc1157 and cc60 datasets. The resulting distance matrices, based on allelic differences, were used to generate split networks based on neighbor-net analyses [238] in SplitsTree4 [61]. Groups were identified based on a balance between maximising edge weighting, whilst minimising contradictory splits.

Deduced amino acid sequence variants of the Bexsero® Antigen Sequence Typing (BAST) loci [124] were extracted from cc60 and cc1157 isolates, and the PubMLST vaccine coverage tool was used to assign potential vaccine coverage. The genome comparator tool was used to determine whether the iron acquisition genes *hmbR* and *hpuAB*, and eight MDA phage genes (NEIS0031, NEIS0030, NEIS0029, NEIS0028, NEIS0027, NEIS0025, NEIS0024 and NEIS0023) were present in cc60 and cc1157 isolate genomes.

4.2.3 Allele sequence extraction and alignment

cgMLST and *cps* allele sequences were extracted using the genome comparator tool, selecting to concatenate and align using MAFFT [213]. This approach enables sequences to be extracted even if they have not been previously annotated in an isolate genome. Alignments were checked manually in MEGA X [263], and sites with gaps were removed using TrimAl [264]. Allele sequences for *fHbp* and *pilE* were extracted. Deduced amino acid sequences were aligned using Muscle [237] in MEGA X, correcting for frameshifts.

4.2.4 Phylogenetic analyses

Phylogenetic trees were generated from nucleotide sequence alignments of *cps* and cgMLST genes using the phangorn R package [265]. Trees were generated using the neighbor-joining method, with the Jukes-Cantor substitution model [241], and rooted at the midpoint. For each capsule group of the two clonal complex datasets, tanglegrams comparing *cps* and cgMLST sequence phylogenies were generated using the cophylo function in the R package phytools [266]. Neighbor-net analysis in SplitsTree4 was used to generate a split network from region C nucleotide sequence alignments. In both tanglegrams and the region C split network, carriage isolates were used as reference clonal complexes, since carriage populations represent the typical life-cycle of *N. meningitidis*.

4.2.5 Random Forest

The genome comparator tool was used to generate an allele table of all NEIS loci in the PubMLST sequence database against the cc60 genomes. Loci that were absent in all isolates, or incomplete in greater than 10% of isolates, were removed from the analysis. The resulting allele table of 2114 genes was run through the randomForest algorithm in R, with capsule group as a response variable, and NEIS loci as predictor

variables. Random forest determines the importance of predictor variables in the classification of response variables [267].

The performance of the model was measured using out-of-bag (OOB) error rate. The number of trees in the model and the value of mtry (the number of predictors sampled for splitting at each node), were set to optimise OOB error rate based on the median of five runs for each tested value (Figure 4.2). The optimal number of trees was 3000, and the optimal value of mtry was $5\sqrt{p}$, where p is the number of alleles. The allele table was permutated to minimise biases arising as a result of the numerical allele designations. Based on the Pearson correlation coefficient between the values of two runs, 50 permutations were sufficient to eliminate bias (Figure 4.2).

Predictors were sampled with replacement. Mean decrease in gini (MDG) was used to assess the average importance of each NEIS locus in the classification of capsule group, with each NEIS locus being assigned the mean MDG of the 50 permutations.

4.2.6 Statistics

The association of between capsule group and clonal complex and disease was investigated using a Firth's bias-reduced logistic regression [268] to generate an odds ratio for disease, with cc1157 as a reference for clonal complexes and group B as a reference for capsule groups. Isolates with missing or undefined clonal complexes were not included in the analysis.

4.3 Results

4.3.1 cc60 and cc1157 capsule group B *cps* sequences are not clonal

Neighbor-net phylogenies were generated from cgMLST allele profiles to determine whether isolate genomes could be separated into groups according to capsule group. cc60 genomes were clustered into two distinct groups, both of which contained group B and group E isolates (Figure 4.3A). One of the groups was dominated by capsule

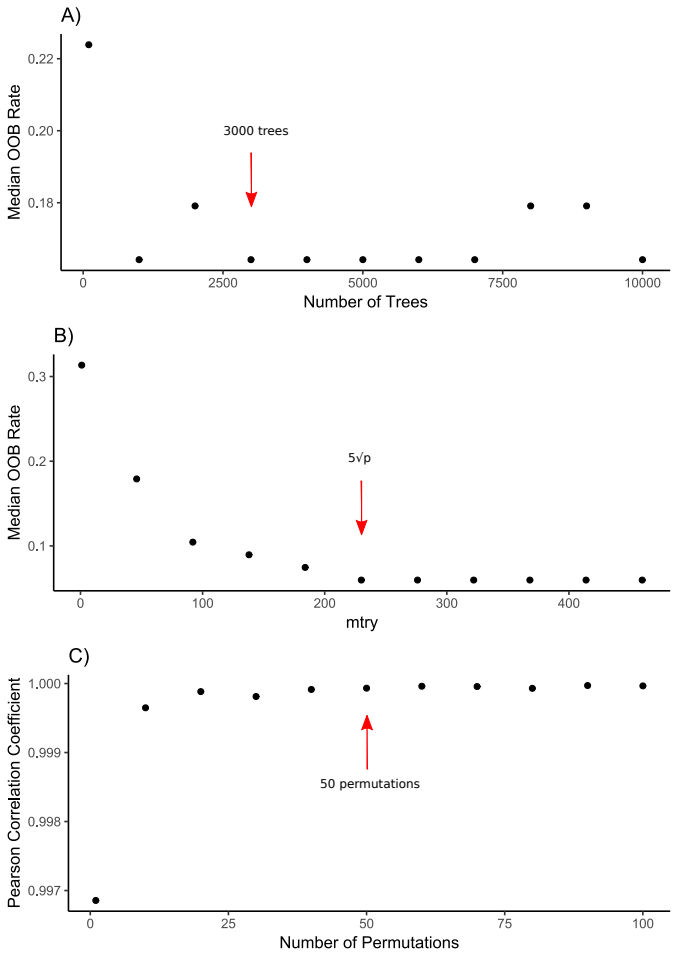

Figure 4.2: Optimisation of random forest for the classification of cc60 capsule group by NEIS loci. A) number of trees and B) value of mtry based on the median OOB error rate of five runs. C) Optimisation of number of allele table permutations based on the correlation of output mean MDG values between two runs.

B isolates (14/17), but edges that split singletons tended to have much greater weight than those that grouped them together.

Edges did not support the separation of cc1157 isolates into distinct groups corresponding with capsule group (Figure 4.3B). There was support for several groups of five or more group E isolates; these were spread across the network. Group E was by far the dominant capsule group, comprising 133 of the 182 isolates. Group B isolates were spread across the network, clustering into singletons or pairs. There was support for one group containing five capsule group X isolates, but additional group X isolates were also found across the cc1157 network.

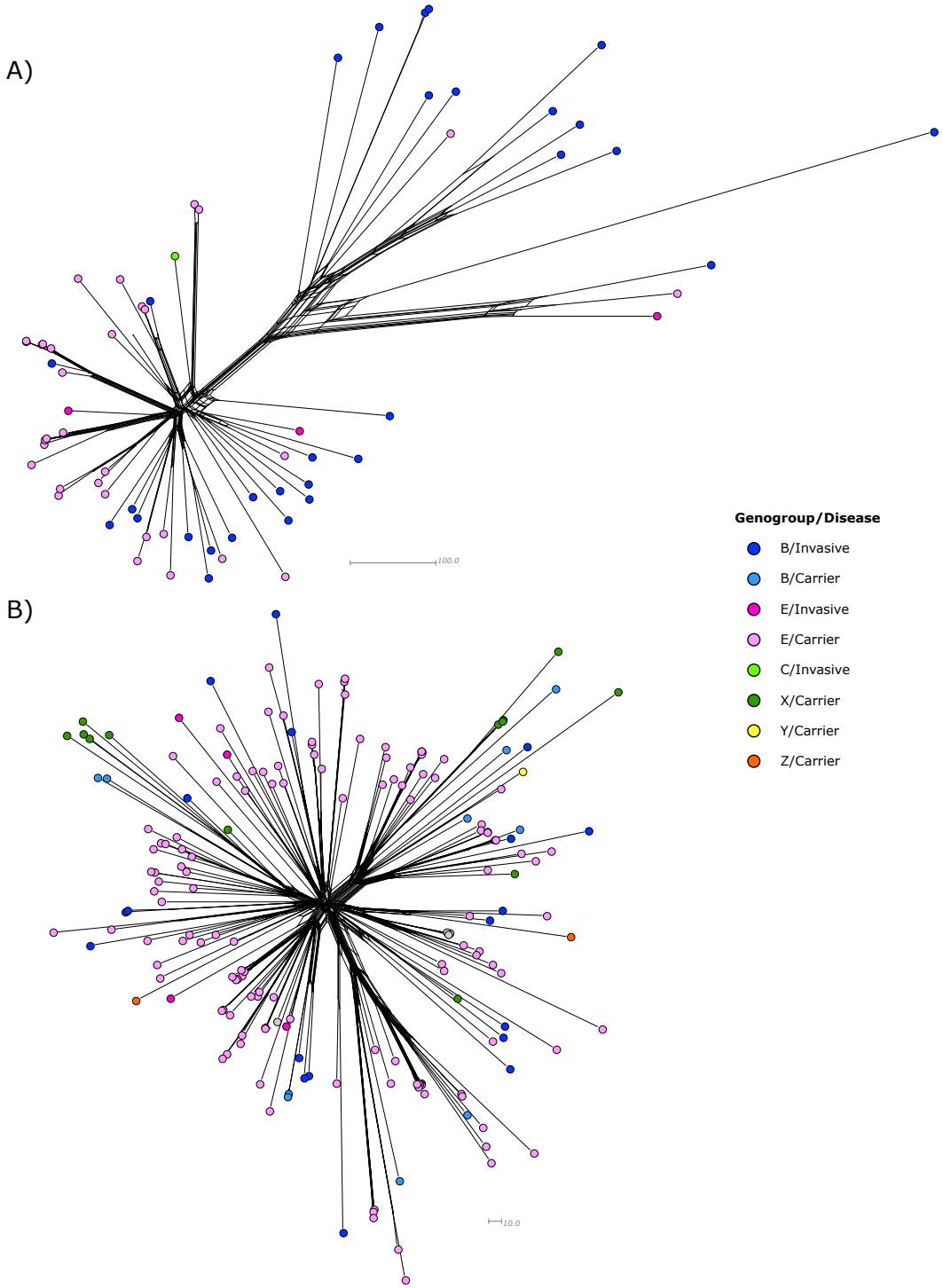

Figure 4.3: Distribution of capsule groups within cc60 and cc1157. Split networks generated using the neighbor-net algorithm, based on cgMLST scheme allele profiles for A) cc60 isolates and B) cc1157 isolates.

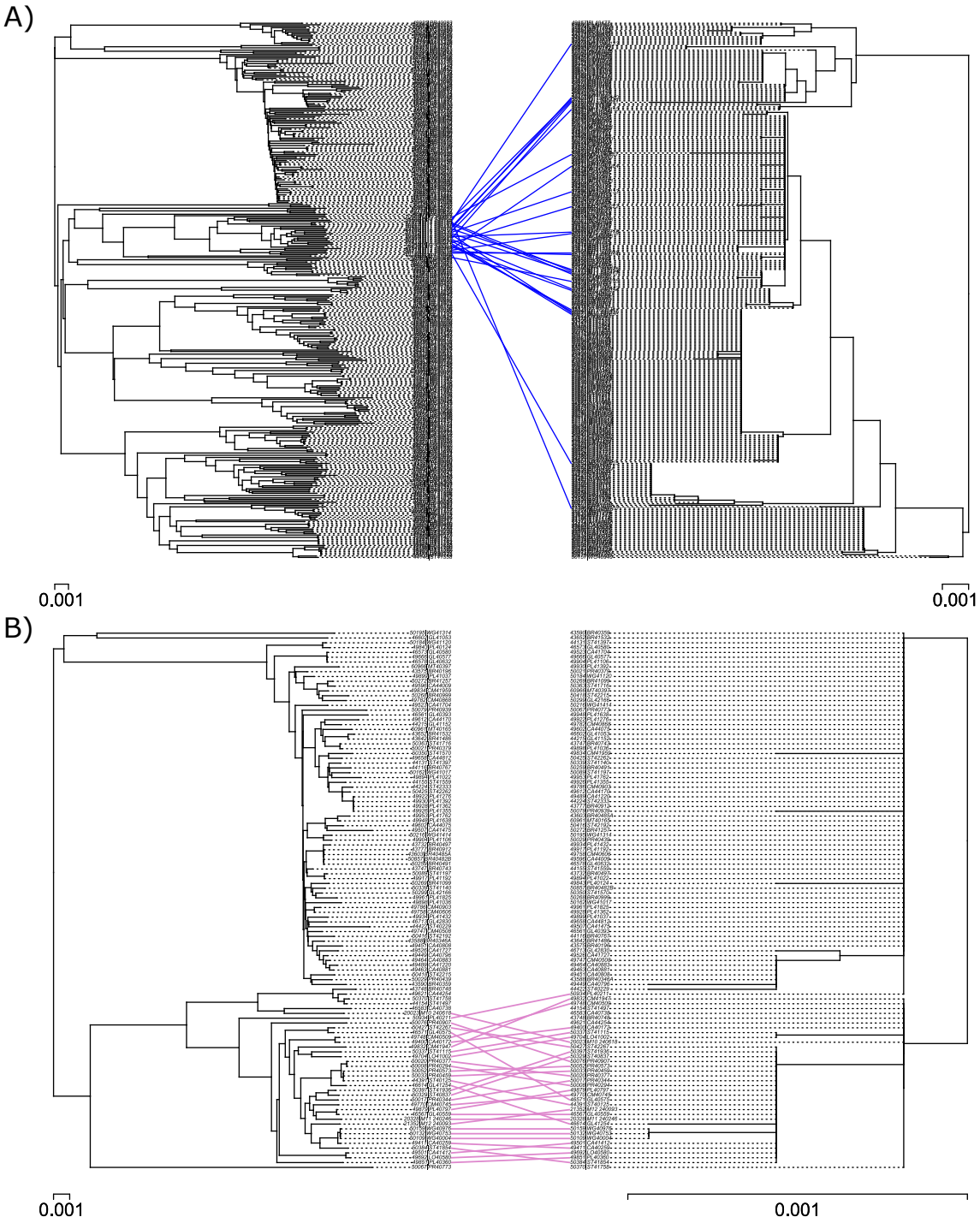

Figure 4.4: Congruence between *cps* region A and the core genome (cc60). Tanglegrams depicting links between cc60 isolates in a cgMLST sequence phylogeny (left) and *cps* region A sequence phylogeny (right) among clonal complexes with a A) B capsule group and B) E capsule group. Phylogenies generated from concatenated aligned nucleotide sequences using the neighbour-joining method with the Jukes-Cantor substitution model.

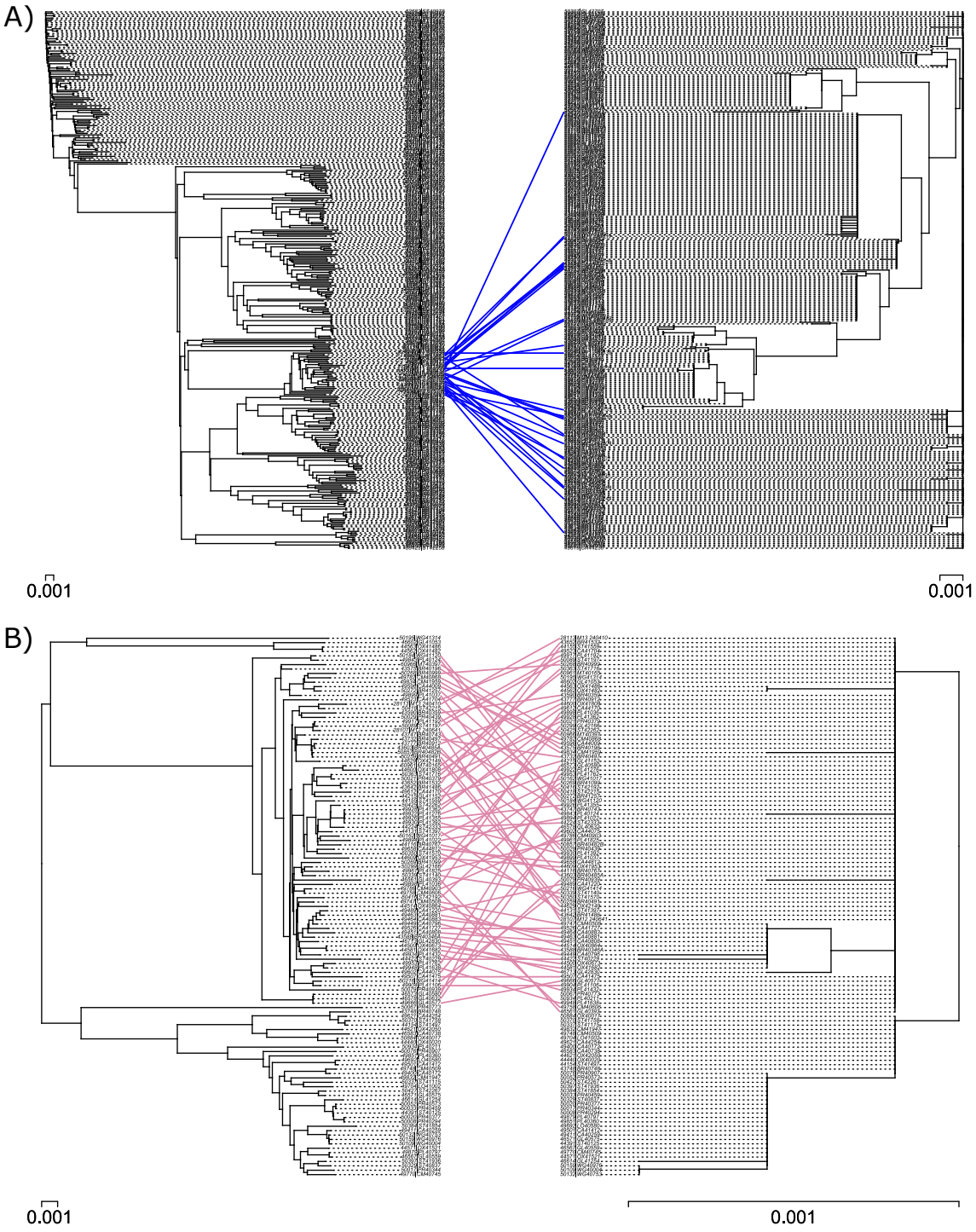

Figure 4.5: Congruence between *cps* region A and the core genome (cc1157). Tanglegrams depicting links between cc1157 isolates in a cgMLST sequence phylogeny (left) and *cps* region A sequence phylogeny (right) among clonal complexes with an A) B capsule group and B) E capsule group. Phylogenies generated form concatenated aligned nucleotide sequences using the neighbour-joining method with the Jukes-Cantor substitution model.

In order to investigate whether *cps* region A sequences within each clonal complex were clonal, tanglegrams were used to assess the congruence between cgMLST sequences and region A sequences, in the context of other clonal complexes with the same capsule group sampled from carriage. Due to the poor assembly of the group E *cps*, 4/32 cc60-E genomes and 53/133 cc1157-E genomes could not be included in the analysis. Similarly, 4/36 cc60-B genomes and 1/29 cc1157-B isolates could not be included due to incomplete region A assembly.

In the cc60 dataset, the capsule group E region A was relatively highly conserved, with a mean overall p-distance across the whole alignment of 0.00027 compared to 0.0051 across the capsule group B region A alignment. Group E tanglegrams revealed high congruence within the cc60 clonal complex between region A sequences and cgMLST sequences (Figure 4.4A). In contrast, the tips of the group B cc60 cgMLST clade did not correspond to a matching clade in the region A phylogeny (Figure 4.4B).

In the cc1157 dataset, the capsule group E region A was also highly conserved, with a mean overall p-distance across the whole alignment of just 0.00027 compared to 0.0041 across the capsule group B region A alignment. There were very few lines linking group E cc1157 cgMLST tips to *cps* tips in clades that would be expected to contain isolates from different clonal complexes, despite the low sequence diversity of the region A sequences (Figure 4.5A). In contrast, the tips of the group B cc1157 cgMLST clade did not correspond to a matching clade in the region A phylogeny (Figure 4.5B).

A split network of nucleotide sequence alignments of region C alleles from both original datasets, reduced to a maximum of three representatives per non-cc1157 or non-cc60 clonal complex to aid computation, was generated to determine whether sequences from each capsule group grouped together within their clonal complex. 28/133 cc1157-E and 1/36 cc60-E isolates were excluded due to incomplete assemblies. All cc60-E region C sequences could be separated into a single group (Figure 4.6A). cc60-B region C sequences were distributed across the network, grouping with region

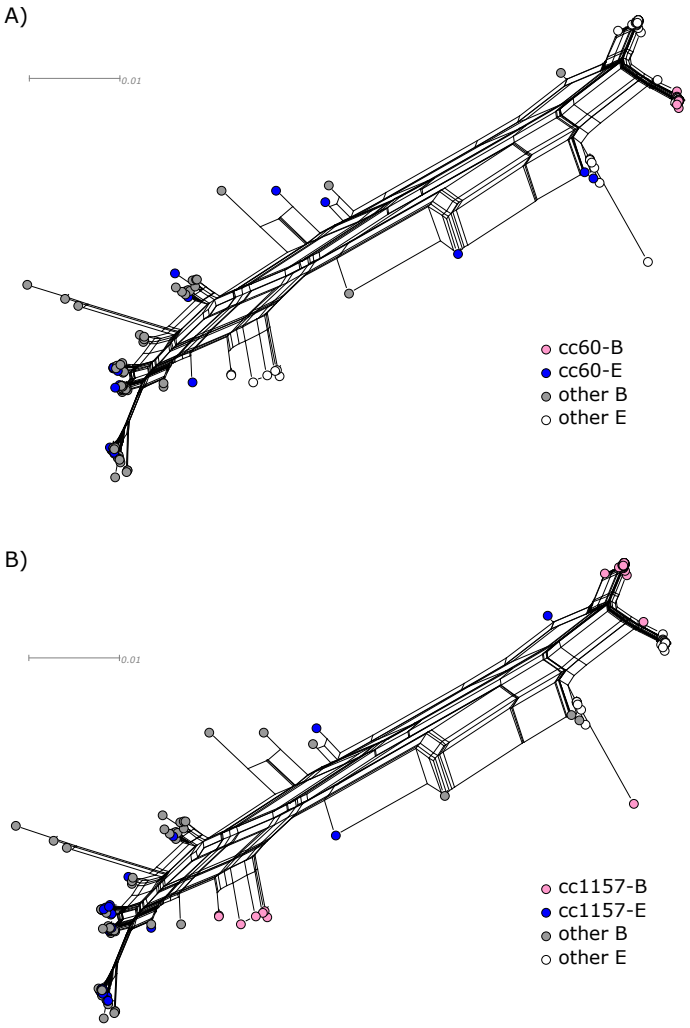

Figure 4.6: Split network of region C sequences. Neighbor-net of concatenated aligned region C sequences from cc60 and cc1157 B (blue) and E (pink) isolate genomes, and B (grey) and E (white) genomes from other clonal complexes. Image is duplicated, highlighting A) cc60 sequences and B) cc1157 sequences.

C sequences from various different capsule group B-associated clonal complexes. With a few exceptions, most cc1157-E region C sequences could be separated into a single group (Figure 4.6B). cc1157-B region C sequences were distributed across the network, grouping with region C sequences from various different capsule group B-associated clonal complexes.

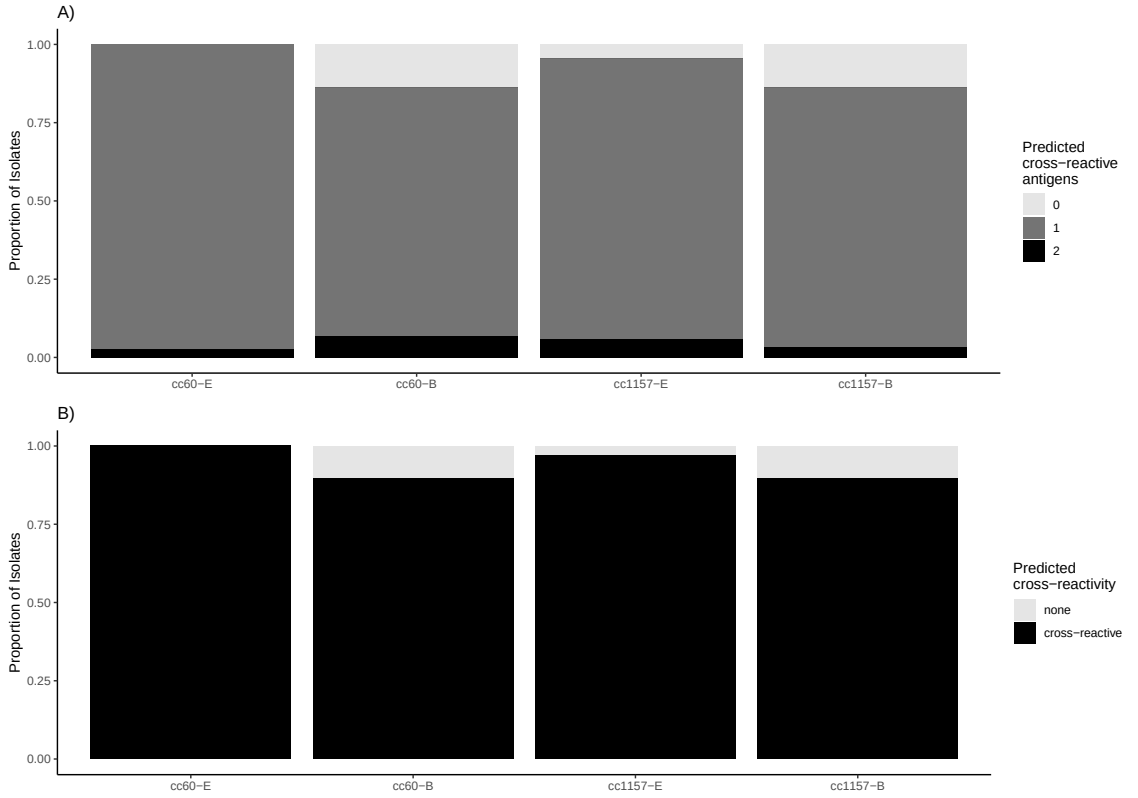

Figure 4.7: Predicted cross-reactivity of "MenB" to cc1157 and cc60 Cross-reactivity based on deduced protein variants and results of bactericidal antibody assays and MATS. A) Proportion of isolates predicted to contain zero, one or two cross reactive (or exact) matches to Bexsero® antigens (fHbp, NadA, NHBA and PorA). No isolates had more than two matches. B) Proportion of isolates with predicted cross-reactivity to Trumenba® .

4.3.2 Both capsule group E and B isolates may be covered by "MenB" vaccines

The vaccine coverage tool in PubMLST was used to predict the potential coverage of cc60 and cc1157 isolates by "MenB" vaccines, and cross-reactive antigens were identified based on variant family. According to the BAST variant prediction tool, 25/29 cc60-B isolates contained predicted cross-reactive matches for Bexsero® fHbp (Figure 4.7A). Of these, one isolate also contained an exact match for PorA variant region 2 (VR2), and one isolate contained an exact match for NadA. 37/37 cc60-E isolates contained predicted cross-reactive matches for fHbp. One isolate contained an additional predicted cross-reactive match for NadA. Similarly, 25/29 cc1157-B

isolates contained predicted cross-reactive matches to Bexsero® fHbp, one of which also contained an exact match for PorA (variant region 2) (Figure 4.7A). 127/133 cc1157-E isolates contained predicted cross-reactive matches for fHbp. Of these, five isolates also contained an additional predicted cross-reactive match for NadA, and three isolates contained exact matches for PorA VR2.

According to the Trumenba® variant prediction tool, 26/29 cc60-B isolates and 37/37 cc60-E isolates contained fHbp variants that are predicted to be partially cross-reactive (Figure 4.7B). Similarly, 26/29 cc1157-B and 129/133 cc1157-E isolates contained fHbp variants that are predicted to be partially cross-reactive (Figure 4.7B).

4.3.3 cc60 may be over-represented in disease

A Firth's bias-corrected logistic model assessing the association of clonal complex and capsule group with disease showed that capsule groups X, W, W/Y, Y, E, Z and *cnl* were significantly less likely to be associated with disease than group B (Table 4.1). The model also showed that cc60, cc11, cc269, cc23, cc22, cc103, cc167 and cc174 were significantly more likely to be associated with disease than cc1157.

4.3.4 Virulence profile of cc60

A Random Forest analysis was used to identify any other genes that may have been exchanged during a capsule switching event in cc60, based on the classification of capsule group by NEIS loci, based on alleles. Region A loci were expected to have high scores since, by definition, *cseABCDEFG* should be present in all capsule group E isolates, but not group B isolates and *vice versa* for *cssABC+csb+ctrG*. No other NEIS loci scored as highly as region A loci apart from region C genes and *tex*, which had almost four-fold higher scores than other loci (Figure 4.8).

Isolate genomes were investigated to determine whether disease-associated accessory genes or core gene variants were present in cc60 and cc1157 genomes (Figure 4.9). All cc60 isolate genomes contained *hpuAB*, but 7/67 (10%) lacked *hmbR*, One

Table 4.1: Association of clonal complex and capsule group with disease. Results of a Firth's bias-corrected logistic regression to generate an odds ratio for disease with 95% confidence intervals, with cc1157 as a reference for clonal complexes and group B as a reference for capsule groups, using UKMenCar4 as a carriage dataset (n=1285) and MRF-MGL epidemiological years 07/2013-06/2016 as a disease dataset (n=1365). Isolates without missing or undefined clonal complexes included only. ORs significantly different from 1 in bold. Value of α 0.05.

Variable	OR	lower	upper	*p*
W	**0.0149**	0.000952	0.103	<.001
W/Y	**0.0102**	0.000568	0.117	<.001
X	**0.00864**	0.0000422	0.114	<.001
cnl	**0.00863**	0.0000677	0.0631	<.001
Y	**0.00663**	0.000482	0.0492	<.001
E	**0.00491**	0.000742	0.0202	<.001
Z	**0.00204**	0.00000938	0.0466	<.001
C	**0.339**	0.116	1.02	.055
Other	2.11	0.0121	71.3	.698
cc11	**281**	35.0	4626	<.001
cc174	**117**	11.5	1673	<.001
cc23	**78.1**	9.17	1132	<.001
cc167	**60.2**	6.46	922.5	<.001
cc22	**41.2**	4.93	675	<.001
cc103	**16.0**	1.57	428	.016
cc60	**6.52**	1.39	43.3	.016
cc269	**3.04**	1.21	7.49	.019
cc865	**0.118**	0.0113	0.653	.012
cc192	28.9	0.127	6817	.172
cc254	17.0	0.110	331	.191
cc178	13.9	0.0905	262	.216
cc18	6.74	0.601	935	.136
cc116	6.64	0.268	1087	.249
cc334	6.64	0.268	1087	.249
cc1136	5.79	0.260	906	.268
cc32	2.51	0.935	6.66	.067
cc364	2.25	0.107	343	.614
cc198	2.16	0.144	317	.614
cc162	1.92	0.620	6.10	.257
cc41/44	1.81	0.740	4.32	.189
cc461	1.57	0.554	4.44	.392
cc1117	1.10	0.00558	216	.964
cc53	1.01	0.0465	157	.994
cc231	0.738	0.00450	18.2	.861
cc213	0.622	0.251	1.51	.292
cc35	0.422	0.146	1.19	.103
cc282	0.321	0.0276	2.34	.267
cc4821	0.204	0.0186	1.27	.090

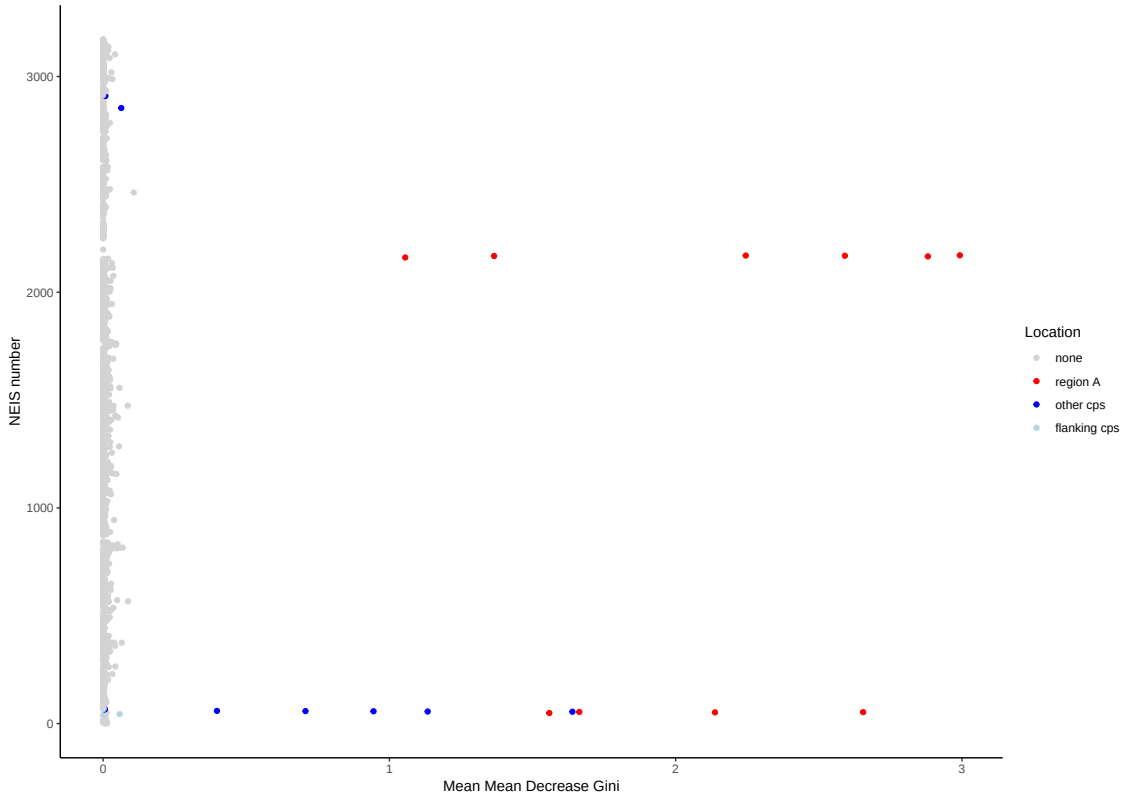

Figure 4.8: Classification of cc60 capsule groups by NEIS loci. Classification of capsule serogroups in 67 isolate genomes based on 2114 NEIS loci as predictors, using the random forest algorithm. Loci scored using the mean of mean decrease gini across 50 permutations. Region A loci in red, other *cps* loci in dark blue, loci flanking *cps* in light blue and all other loci grey. Random forest ran with 3000 trees, with mtry set to 230.

cc1157 isolate genome lacked *hpuA*, and one lacked *hpuAB*, but both contained *hmbR* (Figure 4.9A). 19/182 (10%) of cc1157 isolate genomes lacked *hmbR*. All of the MDA phage genes NEIS0031, NEIS0030, NEIS0029, NEIS0028, NEIS0027, NEIS0025, NEIS0024 and NEIS0023 were found in 43/67 (64%) of cc60 isolate genomes (Figure 4.9B). One additional isolate genome contained NEIS0025 only. At least five of the bacteriophage genes were present in 177/182 (97%) of cc1157 isolate genomes, although five isolates were missing up to three of the genes.

Sequence analyses of the gene encoding factor H binding protein, *fHbp*, and the pilin gene, *pilE*, were carried out to determine the variant or class respectively. MC58 was used as a reference for *fHbp* variant 1, 961-5945 for variant 2, and M1239 for variant 3. 3/67 (4%) of cc60 isolate genomes and 7/182 (4%) of cc1157 isolate

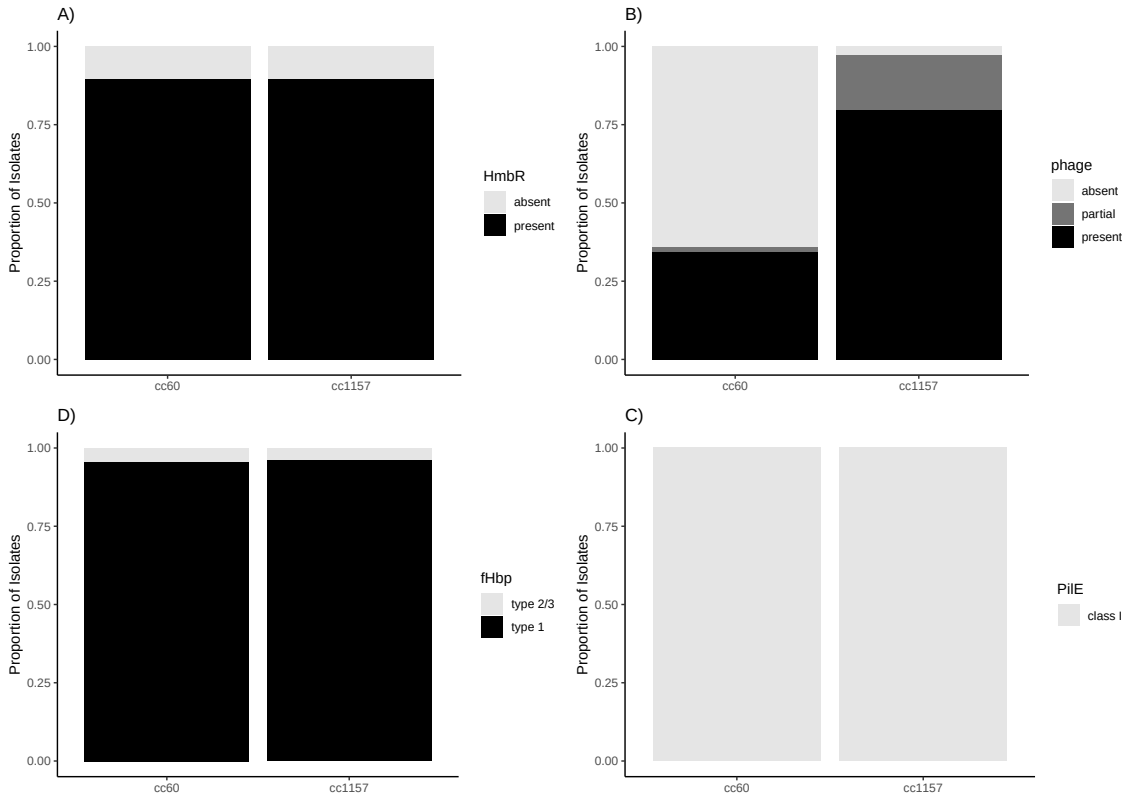

Figure 4.9: Presence of virulence factors in cc60 and cc1157. Presence of A) HmbR, B) MDA phage, C) type 1 or 2/3 fHbp and D) class I or II PilE in cc60 and cc1157 isolate genomes.

genomes contained variant 2 or variant 3 (Figure 4.9C). MC58 and Z2491 were used as references for *pilE* class I, and FAM18 and WUE2594 for class II. All isolates from both cc60 and cc1157 contained class I *pilE* (Figure 4.9D).

4.4 Discussion

Meningococcal isolates belonging to cc1157 or cc60, and expressing a group B capsule, have been responsible for several cases of UK IMD between 2013 and 2016. In carriage studies, these clonal complexes are usually dominated by group E isolates. cc60-B is hardly ever isolated from carriage studies, yet at least one IMD case has been recorded annually in the UK since 2010. In this chapter, the origin of these group B isolates and the potential for coverage by "MenB" vaccines was examined. Possible explanations for the scarcity of cc60-B isolates in carriage were explored, with reference to the

association between capsule group, clonal complex and disease.

Phylogenetic analyses suggest that isolates expressing serogroup B capsules within cc60 and cc1157 have emerged as a result of capsule switching from E to B. The low congruence between the capsule synthesis sequences of group B isolates and their corresponding cgMLST sequences (Figure 4.4A, Figure 4.5A) indicates that the evolutionary relationships of these sequences differs from the average relationship of sequences across the genome. Whereas cc1157 and cc60 core genome sequences form a single clade, corresponding to the clustering of house-keeping genes that define the clonal complexes, capsule group B region A sequences are clustered with *cps* sequences from several different clonal complexes. These relationships suggest that the capsule synthesis sequences have not been acquired by descent, but as a result of several independent HGT events, involving exchange between cc1157 or cc60 and the *cps* of different clonal complexes. This low congruence was not seen between capsule synthesis genes and core genes in group E isolates, consistent with capsule switching from E to B in both clonal complexes (Figure 4.4B, Figure 4.5B).

This analysis may not provide the full picture, since many group E genomes had to be excluded from the tanglegram analysis due to poor assembly of the *cps*. The poor assembly the of the region also meant that it was not possible to analyse the process of recombination in detail as has been described for other capsule switching events [103]. The reason for this poor assembly is unclear, although results seem to be improving with the newer sequencing technologies such as Illumina NovaSeq (H. Bratcher, personal communication). Nevertheless, it was possible to corroborate results from tanglegrams through neighbor-net analysis of region C sequences, for which more genomes were sufficiently well assembled (Figure 4.6). Region C sequences from capsule group B genomes grouped with several different clonal complexes, indicating multiple origins and non-clonal evolution, whereas region C sequences from capsule group E genomes were separated into just one to three groups. Region C sequences are frequently transferred during capsule switching

events as they are adjacent to region A [103, 97], so this is also consistent with capsule switching from E to B in cc60 and cc1157.

There is no indication that either cc60-B or cc1157-B meningococci have persisted and proliferated. The split networks generated from cgMLST allelic profiles (Figure 4.3) do not contain splits indicating the separation of distinct groups of group B profiles, which would be expected in the emergence of persistent cc60-B or cc1157-B lineages. This contrasts with the cc11-W "Hajj" strain, which resulted from C to W capsule switching, and has expanded and persisted [104]. Instead, cc60-B and cc1157-B cases of IMD appear to have arisen as a result of sporadic capsule switching.

Incidence of cc60-B and cc1157-B disease may be reduced in areas where Trumenba® or Bexsero® have been introduced, since antigen typing reveals that the majority of cc60 and cc1157 isolates may have some cross-reactivity with both vaccines (Figure 4.7). In the majority of cases, predicted Bexsero® cross-reactivity was dependent on a single predicted cross-reactive match to fHbp. Predictions were based on the deduction of peptide variants based on genomic sequences, and the cross-reactivity of these variants in bactericidal antibody assays and MATS [125, 123, 122, 257, 258, 259], and may not be reflected in a real-world setting [124]. Early indications suggest Bexsero® may be having efficacy in real-world settings, but with high statistical uncertainty to date [269]. It is already a challenge to quantify the efficacy of the "MenB" vaccines in general, since incidence of serogroup B IMD is low and naturally declining. With cc60-B and cc1157-B cases even lower, quantification may not be possible, but efficacy could be inferred from the general efficacy against partially cross-reactive strains.

Any effect that the vaccines have on reducing cc60-B and cc1157-B disease would benefit public health, but the cross-reactivity estimates also include cc60-E and cc1157-E isolates. cc1157-E in particular constitutes a large proportion of the carried meningococcal population, and the full epidemiological consequences and impacts on the dynamics of the nasopharyngeal microbiota from targeting these populations is

not fully understood [270]. On the other hand, targeting any carried meningococci that are a reservoir of disease will help to control IMD, since carriage is a key source of transmission [36], and was important for the success of the MenC campaign in the early 2000s [168]. Speculation about the risks of "MenB" vaccines reducing carriage of non disease-associated meningococci may be purely academic, since initial investigations indicate that Bexsero® does not have an impact on carriage [271].

In the absence of an effective vaccine against serogroup B, it is to be expected that capsule switching from group E to group B would increase the prevalence of cc1157 and cc60 cases of IMD, since expression of a serogroup B capsule is associated with IMD. Nevertheless, since carriage is an important source of transmission and circulation of meningococci, it is unusual that cc60-B meningococci were not also isolated in UKMenCar4, given the modest incidence of cc60-B isolated from IMD. Complementary carriage and disease datasets have previously been used to generate an odds ratio of disease for given capsule groups or clonal complexes in both *N. meningitidis* and other bacterial species [247, 272, 273]. Typically, cc11, cc41/44, cc269 and cc32 are over-represented in cases of IMD. These models do not necessarily account for the effect of the dominant capsule group in a clonal complex, which is important, because half the known meningococcal serogroups are not associated with disease. This is particularly relevant in cc60, since there are relatively high rates of cc60-E carriage. Incorporating both capsule group and clonal complex into a logistic regression requires a Firth's bias reduction [268], to correct for the fact that by a combination of meningococcal epidemiology and under-sampling of rare genotypes, many cells in a contingency table will tend toward zero. As expected, odds ratios indicated that capsule groups E [95% CI, 0.000742 to 0.0202], Z [95% CI, 0.00204 to 0.0466] and *cnl* [95% CI, 0.00863 to 0.0631] were less likely to be associated with disease than group B, lending confidence to the model (Table 4.1). cc11 [95% CI, 35.0 to 4640], which consistently comes up as being strongly associated with disease [247, 249, 248], was also over-represented in disease relative to the reference, cc1157.

Clonal complexes cc32 [95% CI 0.935, to 6.67] and cc41/44 [95% CI, 0.740 to 4.32], which are less strongly associated with disease [247], were not. With capsule group included as a factor, the model indicated that cc60 [95% CI, 1.391 to 43.3] was significantly over-represented in disease compared to cc1157 (Table 4.1) .

There are a few caveats to the result of the model that must be considered. Firstly, the p-value for association of cc60 with disease relative to cc1157 was only slightly below the bound of the α value of 0.05 (Table 4.1) . Since incidence of cc60-B disease is still relatively low, confidence in the analysis would increase with larger datasets of both carriage and disease isolates. Precise matching between disease datasets and control datasets would further increase confidence in these conclusions, but is largely precluded by the high expense and logistical requirements of carrying out a sufficiently large matched carriage study. The analysis also only gives an estimate relative to cc1157. Nevertheless, in support of this analysis, only four cc60-B carriage isolate genomes have ever been publicly deposited in PubMLST, consisting of two isolates from a university carriage study in the USA, and two isolates from a close contact of an IMD case in Greece, consistent with generally very low carriage rates. Replication of these immeasurably low rates of cc60-E carriage in future large carriage studies would increase confidence in the validity of this distribution. Bearing in mind these caveats, the data do point toward the possibility that cc60-B is over-represented in disease.

There are two possible explanations for why a clonal complex may be over-represented in disease. The first relates to the carriage dynamics of an organism. This theory is based on the premise that IMD takes place within the first few of weeks of acquisition of a meningococcus [274, 275]. If a strain is transmitted quickly, but also cleared quickly, it will spread to more hosts in a population, resulting in more cases of IMD, but will not necessarily be present at high frequencies in a point of prevalence carriage study [247]. Alternatively, the strain may have an unusually virulent phenotype, which increases the incidence of disease per host infected. Theoretically, both scenarios are equally valid [247]. In the case of cc60,

it is more intuitive that acquisition of a disease-associated capsule would facilitate a highly virulent phenotype of *N. meningitidis* to cause disease, than change its transmission dynamics, although a better understanding of the capsule's function in asymptomatic colonisation might inform this further.

No other major HGT events in genes with a function in virulence or colonisation could be identified in the genomes of isolates that had switched from capsule group E to group B. According to random forest analysis, the most important allelic differences that could be used to predict the capsule group of cc60 isolates, besides region A itself, occurred in *cps* genes flanking region A (Figure 4.8). This can be expected in a capsule switching event, since recombination can take place in the genes flanking the capsule synthesis region [103, 97], and corroborates with the neighbor-net analysis of region C sequences (Figure 4.6). This analysis may not pick up smaller regions of recombination or events that happen in only some representatives, but it should pick up any major differences in accessory genes. In this case, no other major differences between carriage-associated and disease-associated cc60 meningococci were identified.

The whole clonal complex was investigated to determine whether it had a virulence phenotype that might further increase the risk of IMD when a serogroup B capsule is expressed, and compared with cc1157 (Figure 4.9). Of the most well investigated factors associated with virulence, almost all isolates possessed HmbR, and variant 1 of fHbp. The MDA phage was also common, but was not present in a majority of isolates. Although these factors are associated with virulence, they have also been found in many carriage-associated isolates. cc1157 profiles were very similar, in fact a higher proportion of cc1157 genomes contained the MDA phage. Therefore, at least with respect to the genes analysed, cc60 isolates generally possess accessory genes associated with virulence, and a virulence-associated variant of fHbp, but cannot be said to have a particularly unusual, potentially highly virulent profile. The challenge for understanding hyperinvasiveness in *N. meningitidis* lies in the likely polygenic

link between genotype and phenotype. Recognising other clonal complexes that may be over-represented in disease, albeit rarely, may help to overcome this challenge.

4.4.1 Conclusions

Independent capsule switching events from E to B within cc60 and cc1157 have resulted in a modest number of cases of IMD in the UK. Future cases may be prevented due to the implementation of the Bexsero® vaccine in UK infants, if the efficacy of partially cross-reactive matches is as high as has been claimed. This would be good news for public health, since initial analyses suggest that cc60-B may be over-represented in disease, which could be indicative of rapid transmission or a highly virulent phenotype. Either of these traits would be a cause for concern if clonal expansion of a cc60-B lineage were to occur. Capsule switching has important impacts on the epidemiology of *N. meningitidis*, and this exploratory investigation is the first reported case of switching within a clonal complex from a non disease-associated capsule to one that is associated with disease.

5
General Discussion

From a biomedical perspective, the meningococcal capsule locus, *cps*, is one of the most important accessory loci in the *N. meningitidis* gene pool. In the absence of the capsule synthesis genes required for the expression of one of capsule serogroups A, B, C, W, X or Y, IMD would be a extremely rare, albeit serious disease, most commonly isolated from immunocompromised patients [23, 24]. As a consequence, the meningococcal capsule has been widely studied as a virulence factor [205], cited as a classic example of a factor that tips the balance between commensalism and disease, and has successfully been exploited in the development of conjugate meningococcal vaccines [116]. Nevertheless, like most human-associated *Neisseria* species, *N. meningitidis* is principally a commensal organism [36], and capsules are present in numerous non-pathogenic Gram-negative bacteria [147, 145, 146, 203]. Following the sequencing of several NPN species, the opportunity has arisen to use genomic approaches to confirm or refute the assumed acapsulate status of these species. This has important implications for the evolutionary history of the meningococcal capsule, a subject which is has been largely speculative [72, 83, 143]. In this thesis, genomic analyses of the polysaccharide within the genus *Neisseria* have been undertaken, including characterisation of capsule genes in NPN species,

re-examination of meningococcal capsule acquisition hypotheses, and exploration of a contemporary example of capsule switching.

In spite of its reputation as a major virulence factor [205], and a detailed understanding of its function during IMD, the polysaccharide capsule had been incorrectly assumed to be unique, among the *Neisseria*, to *N. meningitidis* [64]. Chapter 2 refuted this belief through the identification and characterisation of putative capsule loci in 13 of 20 NPN species. This finding relied on several genomic analyses. Firstly, orthologues of the meningococcal capsule export genes *ctrABCDEF*, which are highly conserved among *N. meningitidis* and other Gram-negative bacteria, were identified in the NPN species. In the majority of cases, these genes were organised similarly to the meningococcal *cps* and the capsule locus of other species, commonly with *ctrABCD* and *ctrEF* flanking a complement of putative capsule synthesis genes. All putative capsule synthesis regions at least partially resembled previously characterised capsule synthesis regions in other bacteria, including *N. meningitidis*, *Actinobacillus* and *Haemophilus*, and some of the structural predictions inferred from these comparisons were confirmed by preliminary ^{1}H NMR analyses of extracted polysaccharides. Further confirmation could be acquired through: microscopic methods, including India ink and alcian blue staining [221]; further ^{1}H and ^{13}C NMR structural analyses; and functional analyses involving the generation of mutants. Nevertheless, given the wealth of capsule examples in other bacteria available for comparison, these genomic data, which have enabled high-throughput genetic characterisation, present a strong indication that capsules exist in many *Neisseria* species, in addition to *N. meningitidis*.

Many other virulence factors have been previously identified in NPN species [65, 64], and it has been suggested that only a few genes can truly be considered virulence factors, as opposed to "host adaptation" or "niche" factors [64, 224]. In light of the findings in Chapter 2, the term virulence factor is also an inaccurate, albeit biomedically convenient, descriptor of the polysaccharide capsule. The term virulence

could perhaps be more readily applied to the region A loci of disease-associated serogroups, equivalents of which have not been identified in NPN species. To further explore this, it would be interesting to determine the serum sensitivity of the novel NPN capsules compared to both disease-associated and non disease-associated meningococcal capsules. A complementary question that warrants investigation is whether switching out native NPN capsules for disease-associated meningococcal capsules increases virulence potential, a question that might be best answered through using novel *Neisseria* animal models including the zebrafish, and the natural commensal inhabitant of the wild house mouse, *N. musculi* [173, 221, 276]. Nevertheless, even among disease-associated serogroups, carriage is the principal state of the *N. meningitidis* life-cycle.

The presence of capsules in NPN species is consistent with the biology of *N. meningitidis* and the meningococcal capsule: the capsule is necessary for IMD, but by no means sufficient, and it is unlikely that *N. meningitidis* expresses a capsule primarily for the purpose of survival during IMD, since it exists primarily as a commensal organism, [141, 143, 245]. This is most evident by the fact that there are a further six capsule groups (E, H, I, K, L, and Z) that are not associated with IMD [78]. Capsules have also been found in many other non-pathogenic bacterial species, including members of the mitis group streptococci, which contains the pathogen *S. pneumoniae* [147, 145, 146, 203]. The function of the meningococcal capsule, besides its role in immune evasion during rare cases of IMD, remains obscure. A major obstacle to further investigating its function is a limited knowledge about both asymptomatic colonisation of the nasopharynx and transmission, which makes it difficult to design functional experiments.

New models of colonisation and transmission may be helpful in answering these questions. Understanding the degree of exposure to mucosal immunity, and the benefits of a capsule in this scenario, would be particularly pertinent. Specific analyses of associations between previous infection and smoking, and encapsulation

status of human carriage isolates, may further complement such experiments. On the other hand, it is possible that the capsule does not have a very important impact on meningococcal survival, explaining the presence of several unencapsulated *Neisseria* species, including certain strains of *N. meningitidis*. IMD is relatively rare compared to carriage rates, with 0.5-6 cases per 100,000 population in industrialised countries [115]. Therefore, although disease is considered to be evolutionarily disadvantageous for *N. meningitidis*, the trade-off between the beneficial functions of a capsule and the risk of its enabling disease need not be great. Variation in distribution of the capsule may be largely determined by chance, potentially driven by the opportunity for horizontal exchange of capsule genes.

The presence of capsule genes in NPN species raises the possibility that *N. meningitidis* acquired its capsule by descent, contrary to established hypotheses that posit that *cps* was acquired *de novo* by HGT [111, 199, 72, 143, 193]. Nevertheless, the capsule genes are not universally distributed through the *Neisseria* genus, and are absent in all five of the species most closely related to *N. meningitidis*. Therefore, the loss of capsule in a common ancestor of these species, followed by a re-acquisition in *N. meningitidis* by HGT, is certainly conceivable. In this thesis, sequence analyses have been used to build on previous models postulating how such an acquisition event may have occurred.

It has been frequently argued that the low G+C content of region A of the meningococcal *cps* is evidence of an HGT event [78, 83, 193]. Nevertheless, in addition to the similarly low G+C content in the capsule synthesis regions of other Gram-negative bacteria [196, 197, 198], G+C content was also low in the majority of the putative region A sequences identified in Chapter 2. This is an interesting observation, which may be a starting point for investigating the common ancestral history of capsule synthesis genes among the Gram-negative bacteria, but it is not unique to *N. meningitidis*, and is not a convincing piece of evidence in support of an HGT event after the divergence of *N. meningitidis*. Whilst G+C content can

be a helpful "parametric" method of identifying sites of HGT, its scope is limited, and more detail can be obtained from phylogenetic methods [277].

Sequence analyses in Chapter 3 indicate that the meningococcal *cps* is highly mosaic, and has undergone HGT with two or more *Neisseria* species. Specifically, phylogenetic analyses are consistent with the donation of capsule export gene sequences from *N. subflava* to *N. meningitidis*. The breakpoints of these recombinant sequences, relative to the overall organisation of the meningococcal *cps* and the putative capsule locus of *N. subflava*, are consistent with a model postulating that the capsule was acquired *en bloc* by a formerly capsule null organism in an HGT event [111]. Nevertheless, it was not possible to prove definitively that a capsule did not exist in *N. meningitidis* prior to recombination with *N. subflava*, nor to identify potential donors of the meningococcal region A.

There are a number of next steps to follow on from this work. First, a more detailed phylogenetic analysis of region E, the potential second site of *cps* capture in the *en bloc* model, should be performed. Second, reconciliation between the distribution of capsule sequences among meningococci, and evolutionary models of meningococcal diversification and strain structure are required [193, 56, 199]. Finally, carriage studies should continue to value the isolation and sequencing of NPN species, which may further inform debate. The publication of the capsule sequences of two new *Neisseria* species reported at the International Pathogenic *Neisseria* Conference 2018 will be read with great interest [278]. Resolving the evolutionary history of the capsule provides insight into the emergence of meningococcal virulence.

From the perspective of the meningococcus, the possession of the capsule results in a potentially minor phenotypic variance, which may have minimal impacts on its preferred niche. From a biomedical perspective, if the capsule was acquired *de novo* by HGT, this would constitute a major event with a profound impact on the epidemiology of IMD [72]. It is tempting to map the acquisition of the meningococcal capsule to the early 1800s, when descriptions of IMD epidemics were first recorded

[26, 72]. Unfortunately, *N. meningitidis* mutates too slowly to attempt to quantify the time-frame of a potential capsule acquisition event: *fetA* and *porA* alleles have been shown to persist over a time-frame of decades [279], and the oldest sequenced isolate in PubMLST (63049, isolated 1915) contains *csaC*, *csaD*, *ctrC* and *ctrD* alleles that have been sequenced from meningococci isolated since 2010. With the improvement of NGS techniques and ancient DNA recovery, paleogenomics, particularly of dental calculus, may be the only means of historical investigation of meningococcal evolution [280].

More contemporary evolution of the meningococcal capsule locus also has an impact on the epidemiology of IMD. An example explored in Chapter 4 of this thesis is the switching of capsule serogroup from the non disease-associated group E to the disease associated group B in clonal complexes cc60 and cc1157. IMD is rarely associated with these clonal complexes, since cc60 and cc1157 isolates predominantly possess a group E capsule, which does not confer the same immune protection to meningococci as disease-associated capsules [247]. By exchanging region A genes with serogroup B isolates from other clonal complexes, cc60 and cc1157 capsule switching has resulted in modest incidence of IMD in the UK, although future cases may be limited due to cross-reactivity with the protein antigens contained in "MenB" vaccines [124]. These two capsule switching events are good epidemiological examples demonstrating the association of certain capsule groups with virulence.

cc60-B may be a particularly interesting case study since, despite modest annual cases of cc60-B disease occurring as a result of capsule switching, no cc60-B isolates were isolated from UKMenCar4, and WGS data from only four carriage cc60-B isolates have ever been publicly deposited in PubMLST. Therefore, in common with a number of hyperinvasive clonal complexes, including cc11 [247, 248, 249], cc60 may be over-represented in disease. Further large-scale carriage studies will need to be analysed to confirm whether over-representation of cc60 in disease is replicated in other settings, but it points to the importance of the relationship between capsule serogroup expression and hyperinvasiveness in the context of the rest

of the genomic profile. Distinguishing between the effects of the predominant capsule group on the association between clonal complex and disease is necessary to identify the polygenic profiles that either facilitate rapid transmission, or a hyperinvasive and/or highly virulent phenotype.

In conclusion, this thesis has demonstrated that the polysaccharide capsule is not unique to *N. meningitidis*, among the *Neisseria*, although six serogroups are uniquely associated with IMD, the synthesis genes of which have not been identified elsewhere within the genus. Recombination between the capsule sequences of *N. meningitidis* and NPN species has occurred on more than one occasion, shaping the evolutionary history of this biomedically important locus. It is possible that *N. subflava* may have served as the donor of a *de novo* capsule acquisition event in *N. meningitidis*, conferring on the bacterium the potential to cause IMD. Further intra-species recombination within the *cps* continues to impact on the dynamics of IMD today. Analysis of the capsule is essential to form an understanding of the biology of *N. meningitidis*, and factors that tip the balance away from commensalism toward disease. Future analyses must be grounded in the biological context of the capsule, a factor whose primary function is surely not to cause IMD, and which only in combination with other gene products, in very specific circumstances, facilitates survival of meningococci in the bloodstream and the onset of life-threatening disease.

References

[1] G Liu, C M Tang, and R M Exley. Non-pathogenic *Neisseria*: members of an abundant, multi-habitat, diverse genus. *Microbiology*, 161(7):1297–1312, 2015.

[2] N J Weyand, M Mancheong, M Phifer-Rixey, N A Taku, M A Rendón, A M Hockenberry, W J Kim, A B Agellon, N Biais, T A Suzuki, L G Sait, O B Harrison, M W Nachman, M C J Maiden, and M So. Isolation and characterization of a new species of *Neisseria*, *Neisseria musculi*, from the wild house mouse. *International Journal of Systematic and Evolutionary Microbiology*, 66(9):3585–3593, 2016.

[3] P H A Sneath and S J Barrett. A new species of *Neisseria* from the dental plaque of the domestic cow, *Neisseria dentiae* sp. nov. *Letters in Applied Microbiology*, 23:355–358, 1996.

[4] S J Barrett, L K Schlater, R J Montali, and P H A Sneath. A new species of *Neisseria* from iguanid lizards, *Neisseria iguanae* sp. nov. *Letters in Applied Microbiology*, 18(4):200–202, 1994.

[5] C Wang, J Luo, H Wang, S Amer, H Ding, Y Dong, and H He. Isolation and characterization of a *Neisseria* sp. from the captive wild goose (*Anser anser*). *Integrative Zoology*, 11(2):128–133, 2016.

[6] N E Rosenstein, B A Perkins, D S Stephens, T Popovic, and J M Hughes. Meningococal disease. *The New England Journal of Medicine*, 344(18):1378–1388, 2001.

[7] L Newman, J Rowley, S V Hoorn, N S Wijesooriya, M Unemo, Nicola Low, G Stevens, S Gottlieb, J Kiarie, and M Temmerman. Global estimates of the prevalence and incidence of four curable sexually transmitted infections in 2012 based on systematic review and global reporting. *PLoS ONE*, 10(12):e0143304, 2015.

[8] A P Zavascki, L Fritscher, S Superti, C Dias, L Kroth, M A Traesel, I C F Antonello, and D Saitovitch. First case report of *Neisseria lactamica* causing cavitary lung disease in an adult organ transplant recipient. *Journal of Clinical Microbiology*, 44(7):2666–2668, 2006.

[9] L J Haddow, C Mulgrew, A Ansari, J Miell, G Jackson, H Malnick, and G Gopal Rao. *Neisseria elongata* endocarditis: case report and literature review. *Clinical Microbiology and Infection*, 9(5):426–430, 2003.

[10] D Wakui, G Nagashima, Y Otsuka, T Takada, T Ueda, Y I Tanaka, and T Hashimoto. A case of meningitis due to *Neisseria subflava* after ventriculostomy. *Journal of Infection and Chemotherapy*, 18(1):115–118, 2012.

[11] X Y Han, T Hong, and E Falsen. *Neisseria bacilliformis* sp. nov. isolated from human infections. *Journal of Clinical Microbiology*, 44(2):474–479, 2006.

[12] J S Bennett, K A Jolley, and M C J Maiden. Genome sequence analyses show that *Neisseria oralis* is the same species as '*Neisseria mucosa* var. *heidelbergensis*'. *International Journal of Systematic and Evolutionary Microbiology*, 63(10):3920–6, 2013.

[13] J S Bennett, K A Jolley, S G Earle, C Corton, S D Bentley, J Parkhill, and M C J Maiden. A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*. *Microbiology*, 158(6):1570–80, 2012.

[14] K A Jolley, C M Bliss, J S Bennett, H B Bratcher, C Brehony, F M Colles, H Wimalarathna, O B Harrison, S K Sheppard, A J Cody, and M C J Maiden. Ribosomal multilocus sequence typing: universal characterization of bacteria from domain to strain. *Microbiology*, 158:1005–1015, 2012.

[15] K Diallo. *The molecular epidemiology and ecology of Neisseria species in the African meningitis belt*. PhD thesis, University of Oxford, 2017.

[16] C Donati, M Zolfo, D Albanese, D Tin Truong, F Asnicar, V Iebba, D Cavalieri, O Jousson, C De Filippo, C Huttenhower, and N Segata. Uncovering oral *Neisseria* tropism and persistence using metagenomic sequencing. *Nature Microbiology*, 1:16070, 2016.

[17] K Bovre and E Holten. *Neisseria elongata* sp. nov., a rod-shaped member of the genus *Neisseria*. Re-evaluation of cell shape as a criterion in classification. *Journal of General Microbiology*, 60:67–75, 1970.

[18] B M Andersen, R S Weyant, A G Steigerwalt, C W Moss, D G Hollis, R E Weaver, D Ashford, and D J Brenner. Characterization of *Neisseria elongata* subsp. *glycolytica* isolates obtained from human wound specimens and blood cultures. *Journal of Clinical Microbiology*, 33(1):76–78, 1995.

[19] K Diallo, C Trotter, Y Timbine, B Tamboura, S O Sow, B Issaka, I D Dano, J M Collard, M Dieng, A Diallo, A Mihret, O A Ali, A Aseffa, S L Quaye, A Bugri, I Osei, K Gamougam, L Mbainadji, D M Daugla, G Gadzama, Z B Sambo, B A Omotara, J S Bennett, L S Rebbetts, E R Watkins, M Nascimento, A Woukeu, O Manigart, R Borrow, J M Stuart, B M Greenwood, and M C J Maiden. Pharyngeal carriage of *Neisseria* species in the African meningitis belt. *Journal of Infection*, 72(6):667–677, 2016.

[20] U Berger. *Neisseria animalis* nov. spec. *Journal of Hygiene and Infectious Diseases*, 147(2):158–161, 1960.

[21] G Zhang, J Yang, L Xin-He, D Jin, J Pu, X Bai, Y Xiong, Z Ren, X Luo, Y Huang, S Lu, and J Xu. *Neisseria weixii* sp. nov., isolated from rectal contents of Tibetan Plateau pika (*Ochotona curzoniae*). *International Journal of Systematic and Evolutionary Microbiology*, 69(8), 2019.

[22] D S Stephens, B Greenwood, and P Brandtzaeg. Epidemic meningitis, meningococcaemia, and *Neisseria meningitidis*. *The Lancet*, 369(9580):2196–2210, 2007.

[23] M Coureuil, O Join-Lambert, H Lécuyer, S Bourdoulous, S Marullo, and X Nassif. Pathogenesis of meningococcemia. *Cold Spring Harbor Perspectives in Medicine*, 3(6):a012393, 2013.

[24] M Van Deuren, P Brandtzaeg, and J W M Van Der Meer. Update on meningococcal disease with emphasis on pathogenesis and clinical management. *Clinical Microbiology Reviews*, 13(1):144–166, 2000.

[25] C B Vicentini, S Manfredini, M Maritati, M Di Nuzzo, and C Contini. Gonorrhea, a current disease with ancient roots: from the remedies of the past to future perspectives. *Le Infezioni in Medicina*, 27(2):212–221, 2019.

[26] G Vieusseux. Memoir on the illness that reigned in Geneva in the spring of 1805 [Memoire sur la maladie qui regnéà Geneve au printemps de 1805]. *Journal de Médecine, de Chirurgie et de Pharmacologie*, 11:50–53, 1806.

[27] D A Caugant. Population genetics and molecular epidemiology of *Neisseria meningitidis*. *APMIS*, 106:505–525, 1998.

[28] M G Krauland, J C D Hotopp, D R Riley, S C Daugherty, J W Marsh, N E Messonnier, L W Mayer, H Tettelin, and L H Harrison. Whole genome sequencing to investigate the emergence of clonal complex 23 *Neisseria meningitidis* serogroup Y disease in the United States. *PLoS ONE*, 7(4):e35699, 2012.

[29] M Bröker, S Bukovski, D Culic, S Jacobsson, M Koliou, M Kuusi, M J Simões, A Skoczynska, M Toropainen, M-K Taha, and G Tzanakaki. Meningococcal serogroup Y emergence in Europe. *Human Vaccines & Immunotherapeutics*, 10(6):1725–1728, 2014.

[30] J Lucidarme, D M C Hill, H B Bratcher, S J Gray, M du Plessis, R S W Tsang, J A Vazquez, M-K Taha, M Ceyhan, A M Efron, M C Gorla, J Findlow, K A Jolley, M C J Maiden, and R Borrow. Genomic resolution of an aggressive, widespread, diverse and expanding meningococcal serogroup B, C and W lineage. *Journal of Infection*, 71(5):544–552, 2015.

[31] A J Chalmers and W R O'Farrell. Preliminary remarks upon epidemic cerebrospinal meningitis as seen in the Anglo-Egyptian Sudan. *Journal of Tropical Medicine and Hygiene*, 19(9):101–116, 1916.

[32] L Lapeyssonnie. Cerebro-spinal meningitis in Africa [La méningite cérébrospinale en Afrique]. *Bulletin of the World Health Organization*, 28:3–114, 1963.

[33] B Greenwood. Editorial: 100 Years of epidemic meningitis in West Africa - Has anything changed? *Tropical Medicine and International Health*, 11(6):773–780, 2006.

[34] S Yezli, A M Assiri, R F Alhakeem, A M Turkistani, and B Alotaibi. Meningococcal disease during the Hajj and Umrah mass gatherings. *International Journal of Infectious Diseases*, 47:60–64, 2016.

[35] C L Trotter, C Lingani, K Fernandez, L V Cooper, A Bita, C Tevi-Benissan, O Ronveaux, M P Préziosi, and J M Stuart. Impact of MenAfriVac in nine countries of the African meningitis belt, 2010–15: an analysis of surveillance data. *The Lancet Infectious Diseases*, 17(8):867–872, 2017.

[36] S P Yazdankhah and D A Caugant. *Neisseria meningitidis*: An overview of the carriage state. *Journal of Medical Microbiology*, 53(9):821–832, 2004.

[37] I C Glitza, I Ehrhard, B Müller-Pebody, R Reintjes, T Breuer, A Ammon, and H G Sonntag. Longitudinal study of meningococcal carrier rates in teenagers. *International Journal of Hygiene and Environmental Health*, 211(3-4):263–272, 2008.

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

[39] C V Broome. The carrier state: *Neisseria meningitidis*. *Journal of Antimicrobial Chemotherapy*, 18:25–34, 1986.

[40] M C J Maiden. Dynamics of bacterial carriage and disease: lessons from the meningococcus. *Advances in Experimental Medicine and Biology*, 549:23–29, 2004.

[41] H Christensen, M May, L Bowen, M Hickman, and C L Trotter. Meningococcal carriage by age: A systematic review and meta-analysis. *The Lancet Infectious Diseases*, 10(12):853–861, 2010.

[42] V Vetter, R Baxter, G Denizer, M A P Sáfadi, S A Silfverdal, A Vyse, and R Borrow. Routinely vaccinating adolescents against meningococcus: targeting transmission & disease. *Expert Review of Vaccines*, 15(5):641–658, 2016.

[43] C L Trotter and M C J Maiden. Meningococcal vaccines and herd immunity: lessons learnt from serogroup C conjugate vaccination programmes. *Expert Review of Vaccines*, 8(7):851–861, 2009.

[44] J A Glover. Observations on the meningococcus carrier-rate in relation to density of population in sleeping quarters. *The Journal of Hygiene*, 17(4):367–79, 1918.

[45] M Baker, A McNicholas, N Garrett, N Jones, J Stewart, V Koberstein, and D Lennon. Household crowding a major risk factor for epidemic meningococcal disease in Auckland children. *Pediatric Infectious Disease Journal*, 19(10):983–990, 2000.

[46] D A Caugant and M C J Maiden. Meningococcal carriage and disease–population biology and evolution. *Vaccine*, 27(4):B64–70, 2009.

[47] J MacLennan, G Kafatos, K Neal, N J Andrews, J C Cameron, R Roberts, M R Evans, K Cann, D N Baxter, M C J Maiden, and J M Stuart. Social behavior and meningococcal carriage in British teenagers. *Emerging Infectious Diseases*, 12(6):950–957, 2006.

[48] J M Stuart, P M Robinson, K A V Cartwright, and N D Noah. Effect of smoking on meningococcal carriage. *Lancet*, 2(8665):723–725, 1989.

[49] C A Jeppesen, M D Snape, H Robinson, N Gossger, T M John, M Voysey, S Ladhani, I O Okike, C Oeser, A Kent, J Oliver, P Taylor, B Morales-Aza, S C Clarke, M Casey, F Martins, N R E Kitchin, A S Anderson, H Jones, K U Jansen, J Eiden, L Pedneault, P T Heath, A Finn, S N Faust, and A J Pollard. Meningococcal carriage in adolescents in the United Kingdom to inform timing of an adolescent vaccination strategy. *Journal of Infection*, 71(1):43–52, 2015.

[50] M C Maiden, J A Bygraves, E Feil, G Morelli, J E Russell, R Urwin, Q Zhang, J Zhou, K Zurth, D A Caugant, I M Feavers, M Achtman, and B G Spratt. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*, 95(6):3140–5, 1998.

[51] E C Holmes, R Urwin, and M C Maiden. The influence of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. *Molecular Biology and Evolution*, 16(6):741–9, 1999.

[52] H Claus, M C J Maiden, R Maag, M Frosch, and U Vogel. Many carried meningococci lack the genes required for capsule synthesis and transport. *Microbiology*, 148(6):1813–1819, 2002.

[53] C Brehony, K A Jolley, and M C J Maiden. Multilocus sequence typing for global surveillance of meningococcal disease. *FEMS Microbiology Reviews*, 31(1):15–26, 2007.

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

[55] W Hao, J H Ma, K Warren, R S W Tsang, D E Low, F B Jamieson, and D C Alexander. Extensive genomic variation within clonal complexes of *Neisseria meningitidis*. *Genome Biology and Evolution*, 3(1):1406–1418, 2011.

[56] J S Bennett, S D Bentley, G S Vernikos, M A Quail, I Cherevach, B White, J Parkhill, and M C J Maiden. 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*, 11:652, 2010.

[57] Y H Sun, R Exley, Y Li, D Goulding, and C Tang. Identification and characterization of genes required for competence in *Neisseria meningitidis*. *Journal of Bacteriology*, 187(9):3273–3276, 2005.

[58] Y Boucher and E Bapteste. Revisiting the concept of lineage in prokaryotes: a phylogenetic perspective. *BioEssays*, 31(5):526–536, 2009.

[59] B G Spratt and M C Maiden. Bacterial population genetics, evolution and epidemiology. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 354(1384):701–10, 1999.

[60] M C J Maiden, B Malorny, and M Achtman. A global gene pool in the neisseriae. *Molecular Microbiology*, 21(6):1297–1298, 1996.

[61] D H Huson and D Bryant. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2):254–267, 2006.

[62] K A Jolley and M C J Maiden. BIGSdb: Scalable analysis of bacterial genome variation at the population level. *BMC Bioinformatics*, 11:595, 2010.

[63] H W Smith. Microbial surfaces in relation to pathogenicity. *Bacteriological Reviews*, 41(2):475–500, 1977.

[64] P R Marri, M Paniscus, N J Weyand, M A Rendón, C M Calton, D R Hernández, D L Higashi, E Sodergren, G M Weinstock, S D Rounsley, and M So. Genome sequencing reveals widespread virulence gene exchange among human *Neisseria* species. *PLoS ONE*, 5(7):e11835, 2010.

[65] L A S Snyder and N J Saunders. The majority of genes in the pathogenic *Neisseria* species are present in non-pathogenic *Neisseria lactamica*, including those designated as 'virulence genes'. *BMC Genomics*, 7:128, 2006.

[66] O B Harrison, N J Evans, J M Blair, H S Grimes, R Colin, X Nassif, P Kriz, R Ure, S J Gray, J P Derrick, M C J Maiden, and I M Feavers. Epidemiological evidence for the role of the haemoglobin receptor, HmbR, in meningococcal virulence. *Journal of Infectious Diseases*, 200(1):94–98, 2009.

[67] E Hong, M Etienne, A E Deghmane, A Terrade, G Berthelot, F Caron, and M-K Taha. Genetic diversity and levels of expression of factor H binding protein among carriage isolates of *Neisseria meningitidis*. *PLoS ONE*, 9(9):e107240, 2014.

[68] X Sun, H Zhou, L Xu, H Yang, Y Gao, B Zhu, and Z Shao. Prevalence and genetic diversity of two adhesion-related genes, *pilE* and *nadA*, in *Neisseria meningitidis* in China. *Epidemiology and Infection*, 141(10):2163–2172, 2013.

[69] A Cehovin, M Winterbotham, J Lucidarme, R Borrow, C M Tang, R M Exley, and V Pelicic. Sequence conservation of pilus subunits in *Neisseria meningitidis*. *Vaccine*, 28(30):4817–4826, 2010.

[70] E Bille, J-R Zahar, A Perrin, S Morelle, P Kriz, K A Jolley, M C J Maiden, C Dervin, X Nassif, and C R Tinsley. A chromosomally integrated bacteriophage in invasive meningococci. *The Journal of Experimental Medicine*, 201(12):1905–1913, 2005.

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

[72] C Schoen, J Blom, H Claus, A Schramm-Gluck, P Brandt, T Muller, A Goesmann, B Joseph, S Konietzny, O Kurzai, C Schmitt, T Friedrich, B Linke, U Vogel, and M Frosch. Whole-genome comparison of disease and carriage strains provides insights into virulence evolution in *Neisseria*

meningitidis. *Proceedings of the National Academy of Sciences of the United States of America*, 105(9):3473–3478, 2008.

[73] I S Roberts. The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annual Review of Microbiology*, 50(1):285–315, 1996.

[74] Z Wen and J R Zhang. Bacterial Capsules. In Yi-Wei Tang, Max Sussman, Dongyou Liu, Ian Poxton, and Joseph Schwartzman, editors, *Molecular Medical Microbiology*, chapter 3, pages 33–53. Elsevier, London, 2nd edition, 2015.

[75] T Ophir and D L Gutnick. A role for exopolysaccharides in the protection of microorganisms from desiccation. *Applied and Environmental Microbiology*, 60(2):740–745, 1994.

[76] L M Willis and C Whitfield. Structure, biosynthesis, and function of bacterial capsular polysaccharides synthesized by ABC transporter-dependent pathways. *Carbohydrate Research*, 378:35–44, 2013.

[77] C Whitfield and I S Roberts. Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Molecular Microbiology*, 31:1307–1319, 1999.

[78] O B Harrison, H Claus, Y Jiang, J S Bennett, H B Bratcher, K A Jolley, C Corton, R Care, J T Poolman, W D Zollinger, C E Frasch, D S Stephens, I Feavers, M Frosch, J Parkhill, U Vogel, M A Quail, S D Bentley, and M C J Maiden. Description and nomenclature of *Neisseria meningitidis* capsule locus. *Emerging Infectious Diseases*, 19(4):566–573, 2013.

[79] L M Willis, J Stupak, M R Richards, T L Lowary, J Li, and C Whitfield. Conserved glycolipid termini in capsular polysaccharides synthesized by ATP-binding cassette transporter-dependent pathways in Gram-negative

pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, 110(19):7868–7873, 2013.

[80] Y-L Tzeng, J Thomas, and D S Stephens. Regulation of capsule in *Neisseria meningitidis*. *Critical Reviews in Microbiology*, 42(5):759–772, 2016.

[81] T Y Liu, E C Gotschlich, F T Dunne, and E K Jonssen. Studies on the meningococcal polysaccharides II. Composition and chemical properties of the group B and group C polysaccharide. *Journal of Biological Chemistry*, 246(15):4703–4712, 1971.

[82] A K Bhattacharjee, H J Jennings, C P Kenny, A Martin, and I C Smith. Structural determination of the polysaccharide antigens of *Neisseria meningitidis* serogroups Y, W-135, and BO. *Canadian Journal of Biochemistry*, 54(1):1–8, 1976.

[83] H Claus, U Vogel, M Mühlenhoff, R Gerardy-Schahn, and M Frosch. Molecular divergence of the *sia* locus in different serogroups of *Neisseria meningitidis* expressing polysialic acid capsules. *Molecular and General Genetics*, 257(1):28–34, 1997.

[84] H Claus, K Stummeyer, J Batzilla, M Mühlenhoff, and U Vogel. Amino acid 310 determines the donor substrate specificity of serogroup W-135 and Y capsule polymerases of *Neisseria meningitidis*. *Molecular Microbiology*, 71(4):960–971, 2009.

[85] Y-L Tzeng, C Noble, and D S Stephens. Genetic basis for biosynthesis of the (alpha) 1–4–linked N–acetyl–D–glucosamine 1–phosphate capsule of *Neisseria meningitidis* serogroup X. *Infection and Immunity*, 71(12):6712–20, 2003.

[86] J S Swartley, L J Liu, Y K Miller, L E Martin, S Edupuganti, and D S Stephens. Characterization of the gene cassette required for biosynthesis of the

(alpha1-6)-linked N-acetyl-D-mannosamine-1-phosphate capsule of serogroup A *Neisseria meningitidis*. *Journal of Bacteriology*, 180(6):1533–1539, 1998.

[87] T Fiebig, F Freiberger, V Pinto, M R Romano, A Black, C Litschko, A Bethe, D Yashunsky, R Adamo, A Nikolaev, F Berti, and R Gerardy-Schahn. Molecular cloning and functional characterization of components of the capsule biosynthesis complex of *Neisseria meningitidis* serogroup A: Toward in vitro vaccine production. *Journal of Biological Chemistry*, 289(28):19395–19407, 2014.

[88] T Fiebig, F Berti, F Freiberger, V Pinto, H Claus, M R Romano, D Proietti, B Brogioni, K Stummeyer, M Berger, U Vogel, P Costantino, and R Gerardy-Schahn. Functional expression of the capsule polymerase of *Neisseria meningitidis* serogroup X: A new perspective for vaccine development. *Glycobiology*, 24(2):150–158, 2014.

[89] C Litschko, M R Romano, V Pinto, H Claus, U Vogel, F Berti, R Gerardy-Schahn, and T Fiebig. The capsule polymerase CslB of *Neisseria meningitidis* serogroup L catalyzes the synthesis of a complex trimeric repeating unit comprising glycosidic and phosphodiester linkages. *Journal of Biological Chemistry*, 290(40):24355–24366, 2015.

[90] F Michon, R Roy, and H J Jennings. Structural elucidation of the capsular polysaccharide of *Neisseria meningitidis* group H. *Canadian Journal of Chemistry*, 62(8):1519–1524, 1984.

[91] HJ Jennings, Karl-Gunnar Rosell, and CP Kenny. Structural elucidation of the capsular polysaccharide antigen of *Neisseria meningitidis* serogroup Z using 13C nuclear magnetic resonance. *Canadian Journal of Chemistry*, 57(22):2902–2907, 1979.

[92] O B Harrison, J S Bennett, J P Derrick, M C J Maiden, and C D Bayliss. Distribution and diversity of the haemoglobin-haptoglobin iron-acquisition systems in pathogenic and non-pathogenic *Neisseria*. *Microbiology*, 159:1920–1930, 2013.

[93] G J Mcleod and B L Brandt. Immunological relationship between the capsular polysaccharides of *Neisseria meningitidis* serogroups Z and 29E. *Journal of General Microbiology*, 129(2):447–452, 1983.

[94] A Van der Kaaden, G J Gerwig, J P Kamerling, J F G Vliegenthart, and R H Tiesjema. Structure of the capsular antigen of *Neisseria meningitidis* serogroup K. *European Journal of Biochemistry*, 152(3):663–668, 1985.

[95] F Michon, J R Brisson, R Roy, H J Jennings, and F E Ashton. Structural determination of the capsular polysaccharide of *Neisseria meningitidis* group I: a two-dimensional NMR analysis. *Biochemistry*, 24(20):5592–5598, 1985.

[96] D E Barroso, T M P P Castineiras, M C Rebelo, M M Tulenko, J W Marsh, M G Krauland, and L H Harrison. Invasive *Neisseria meningitidis* strain expressing capsular polysaccharides W and Y in Brazil. *Journal of Clinical Microbiology*, 51(3):1059–1060, 2013.

[97] X Ji, P-P Yao, Le-Y Zhang, Y Li, F Xu, L-L Mei, S-R Zhu, Y-J Zhang, H-P Zhu, and Stijn van der Veen. Capsule switching of *Neisseria meningitidis* sequence type 7 serogroup A to serogroup X. *Journal of Infection*, 75:521–531, 2017.

[98] J S Swartley, A A Marfin, S Edupuganti, L J Liu, P Cieslak, B Perkins, J D Wenger, and D S Stephens. Capsule switching of *Neisseria meningitidis*. *Proceedings of the National Academy of Sciences of the United States of America*, 94(1):271–6, 1997.

[99] P Stefanelli, C Fazio, A Neri, T Sofia, and P Mastrantonio. First report of capsule replacement among Electrophoretic Type 37 *Neisseria meningitidis* strains in Italy. *Journal of Clinical Microbiology*, 41(12):5783–5786, 2003.

[100] P Stefanelli, C Fazio, P Vacca, A Palmieri, L Ambrosio, A Neri, A Piana, P Castiglia, F Argiolas, S Santus, L Masala, G Steri, F Riccardo, S Iannazzo, F P Maraglino, C D'Amario, and G Rezza. An outbreak of severe invasive meningococcal disease due to a capsular switched *Neisseria meningitidis* hypervirulent strain B:cc11. *Clinical Microbiology and Infection*, 25(1):111.e1–111.e4, 2019.

[101] A Lamelas, J Hauser, J-P Dangy, A-W M Hamid, K Röltgen, M R Abdul Sater, A Hodgson, A Sie, T Junghanss, S R Harris, J Parkhill, S D Bentley, and G Pluschke. Emergence and genomic diversification of a virulent serogroup W:ST-2881(CC175) *Neisseria meningitidis* clone in the African meningitis belt. *Microbial Genomics*, 3(8):e000120, 2017.

[102] R S W Tsang, D K S Law, S D Tyler, G S Stephens, M Bigham, and W D Zollinger. Potential capsule switching from serogroup Y to B: The characterization of three such *Neisseria meningitidis* isolates causing invasive meningococcal disease in Canada. *Canadian Journal of Infectious Diseases and Medical Microbiology*, 16(3):171–174, 2005.

[103] M M Mustapha, J W Marsh, M G Krauland, J O Fernandez, A P S de Lemos, J CC Dunning Hotopp, X Wang, L W Mayer, J G Lawrence, N L Hiller, and L H Harrison. Genomic investigation reveals highly conserved, mosaic, recombination events associated with capsular switching among invasive *Neisseria meningitidis* serogroup W sequence type (ST)-11 strains. *Genome Biology and Evolution*, 8(6):2065–2075, 2016.

[104] M M Mustapha, J W Marsh, M G Krauland, J O Fernandez, A P S de Lemos, J C Dunning Hotopp, X Wang, L W Mayer, J G Lawrence, N L Hiller, and L H Harrison. Genomic epidemiology of hypervirulent serogroup W, ST-11 *Neisseria meningitidis*. *EBioMedicine*, 2(10):1447–1455, 2015.

[105] M Frosch and U Vogel. Structure and genetics of the meningococcal capsule. In M C J Maiden and M Frosch, editors, *Handbook of Meningococcal Disease: Infection Biology, Vaccination, Clinical Management*, chapter 8, pages 145–162. Wiley-VCH GmbH & Co. KGaA, Weinheim, 2006.

[106] L M Willis and C Whitfield. KpsC and KpsS are retaining 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) transferases involved in synthesis of bacterial capsules. *Proceedings of the National Academy of Sciences of the United States of America*, 110(51):20753–8, 2013.

[107] Y-L Tzeng, A K Datta, C A Strole, M A Lobritz, R W Carlson, and D S Stephens. Translocation and surface expression of lipidated serogroup B capsular polysaccharide in *Neisseria meningitidis*. *Infection and Immunity*, 73(3):1491–1505, 2005.

[108] S Hammerschmidt, C Birkholz, U Zähringer, B D Robertson, J Van Putten, O Ebeling, and M Frosch. Contribution of genes from the capsule gene complex (*cps*) to lipooligosaccharide biosynthesis and serum resistance in *Neisseria meningitidis*. *Molecular Microbiology*, 11(5):885–896, 1994.

[109] A K Bhattacharjee, H J Jennings, and C P Kenny. Structural elucidation of the 3-deoxy-d-manno-octulosonic acid containing meninggococcal 29-e capsular polysaccharide antigen using carbon-13 nuclear magnetic resonance. *Biochemistry*, 17(4):645–651, 1978.

[110] H Petering, S Hammerschmidt, M Frosch, J P M Van Putten, C A Ison, and B D Robertson. Genes associated with the meningococcal capsule complex are also found in *Neisseria gonorrhoeae*. *Journal of Bacteriology*, 178(11):3342–3345, 1996.

[111] S N Bartley, S Mowlaboccus, C A Mullally, K A Stubbs, A Vrielink, M C J Maiden, O B Harrison, T T Perkins, and C M Kahler. Acquisition of the capsule locus by horizontal gene transfer in *Neisseria meningitidis* is often accompanied by the loss of UDP-GalNAc synthesis. *Scientific Reports*, 7:44442, 2017.

[112] D Ivanova, T Taylor, S L Smith, J U Dimude, A L Upton, M M Mehrjouy, O Skovgaard, D J Sherratt, R Retkute, and C J Rudolph. Shaping the landscape of the *Escherichia coli* chromosome: Replication-transcription encounters in cells with an ectopic replication origin. *Nucleic Acids Research*, 43(16):7865–7877, 2015.

[113] A J Pollard and C Frasch. Development of natural immunity to *Neisseria meningitidis*. *Vaccine*, 19(11-12):1327–1346, 2001.

[114] P De Wals, G De Serres, and T Niyonsenga. Effectiveness of a mass immunization campaign against serogroup C meningococcal disease in Quebec. *Journal of the American Medical Association*, 285(2):177–181, 2001.

[115] A Khatami and A J Pollard. The epidemiology of meningococcal disease and the impact of vaccines. *Expert Review of Vaccines*, 9(3):285–298, 2010.

[116] R Rappuoli, E De Gregorio, and P Costantino. On the mechanisms of conjugate vaccines. *Proceedings of the National Academy of Sciences of the United States of America*, 116(1):14–16, 2019.

[117] E. Miller, D Salisbury, and M Ramsay. Planning, registration, and implementation of an immunisation campaign against meningococcal serogroup C disease in the UK: a success story. *Vaccine*, 20:58–67, 2001.

[118] P Keshavan, M Pellegrini, K Vadivelu-Pechai, and M Nissen. An update of clinical experience with the quadrivalent meningococcal ACWY-CRM conjugate vaccine. *Expert Review of Vaccines*, 17(10):865–880, 2018.

[119] W H Chen, K M Neuzil, C R Boyce, M F Pasetti, M K Reymann, L Martellet, N Hosken, F M LaForce, R M Dhere, S S Pisal, A Chaudhari, P S Kulkarni, R Borrow, H Findlow, V Brown, M L McDonough, L Dally, and M R Alderson. Safety and immunogenicity of a pentavalent meningococcal conjugate vaccine containing serogroups A, C, Y, W, and X in healthy adults: a phase 1, single-centre, double-blind, randomised, controlled study. *The Lancet Infectious Diseases*, 18(10):1088–1096, 2018.

[120] B F Cress, J A Englaender, W He, D Kasper, R J Linhardt, and M A G Koffas. Masquerading microbial pathogens: capsular polysaccharides mimic host-tissue molecules. *FEMS Microbiology Reviews*, 38(4):660–97, 2014.

[121] D Serruto, M J Bottomley, S Ram, M M Giuliani, and R Rappuoli. The new multicomponent vaccine against meningococcal serogroup B, 4CMenB: immunological, functional and structural characterization of the antigens. *Vaccine*, 30:B87–B97, 2012.

[122] S L Harris, R G K Donald, J C Hawkins, C Tan, R O'Neill, L K McNeil, J L Perez, A S Anderson, K U Jansen, and T R Jones. *Neisseria meningitidis* serogroup B vaccine, bivalent rLP2086, induces broad serum bactericidal activity against diverse invasive disease strains including outbreak strains. *Pediatric Infectious Disease Journal*, 36(2):216–223, 2017.

[123] D Medini, M Stella, and J Wassil. MATS: Global coverage estimates for 4CMenB, a novel multicomponent meningococcal B vaccine. *Vaccine*, 33(23):2629–2636, 2015.

[124] C Brehony, C M C Rodrigues, R Borrow, A Smith, R Cunney, E R Moxon, and M C J Maiden. Distribution of Bexsero® Antigen Sequence Types (BASTs) in invasive meningococcal disease isolates: Implications for immunisation. *Vaccine*, 34(39):4690–4697, 2016.

[125] U Vogel, M-K Taha, J A Vazquez, J Findlow, H Claus, P Stefanelli, D A Caugant, P Kriz, R Abad, S Bambini, A Carannante, A-E Deghmane, C Fazio, M Frosch, G Frosi, S Gilchrist, M M Giuliani, E Hong, M Ledroit, P G Lovaglio, J Lucidarme, M Musilek, A Muzzi, J Oksnes, F Rigat, L Orlandi, M Stella, D Thompson, M Pizza, R Rappuoli, D Serruto, M Comanducci, G Boccadifuoco, J J Donnelly, D Medini, and R Borrow. Predicted strain coverage of a meningococcal multicomponent vaccine (4CMenB) in Europe: A qualitative and quantitative assessment. *The Lancet Infectious Diseases*, 13(5):416–425, 2013.

[126] C M C Rodrigues, J Lucidarme, R Borrow, A Smith, J C Cameron, E R Moxon, and M C J Maiden. Genomic surveillance of 4CMenB vaccine antigenic variants among disease-causing *Neisseria meningitidis* isolates, United Kingdom, 2010-2016. *Emerging Infectious Diseases*, 24(4):673–682, 2018.

[127] H Findlow, U Vogel, J E Mueller, A Curry, B-M Njanpop-Lafourcade, H Claus, S J Gray, S Yaro, Y Traoré, L Sangaré, P Nicolas, B D Gessner, and R Borrow. Three cases of invasive meningococcal disease caused by a capsule null locus strain circulating among healthy carriers in Burkina Faso. *Journal of Infectious Diseases*, 195(7):1071–7, 2007.

[128] K Ganesh, M Allam, N Wolter, H B Bratcher, O B Harrison, J Lucidarme, R Borrow, L de Gouveia, S Meiring, M Birkhead, M C J Maiden, A von Gottberg, and M du Plessis. Molecular characterization of invasive capsule null *Neisseria meningitidis* in South Africa. *BMC Microbiology*, 17:40, 2017.

[129] K O Johswich, J Zhou, D K S Law, F St Michael, S E McCaw, F B Jamieson, A D Cox, R S W Tsang, and S D Gray-Owen. Invasive potential of nonencapsulated disease isolates of *Neisseria meningitidis*. *Infection and Immunity*, 80(7):2346–53, 2012.

[130] D Kerr, W L Hooper, and D M Jones. A rare case of meningitis due to *Neisseria meningitidis* group 29E. *Journal of Infection*, 13:147–149, 1986.

[131] N J Ryan and G R Hogan. Severe meningococcal disease caused by serogroups X and Z. *American Journal of Diseases of Children*, 134(12):1173, 1980.

[132] Public Health England. Invasive meningococcal disease in England: annual laboratory confirmed reports for epidemiological year 2016 to 2017. Technical Report 38, Public Health England, 2018.

[133] U Vogel, H Claus, L Von Müller, D Bunjes, J Elias, and M Frosch. Bacteremia in an immunocompromised patient caused by a commensal *Neisseria meningitidis* strain harboring the capsule null locus (*cnl*). *Journal of Clinical Microbiology*, 42(7):2898–2901, 2004.

[134] D E Craven, K T Shen, and C E Frasch. Natural bactericidal activity of human serum against *Neisseria meningitidis* isolates of different serogroups and serotypes. *Infection and Immunity*, 37(1):132–137, 1982.

[135] M C Geoffroy, S Floquet, A Métais, X Nassif, and V Pelicic. Large-scale analysis of the meningococcus genome by gene distruption: Resistance to complement-mediated lysis. *Genome Research*, 13(3):391–398, 2003.

[136] P Densen. Interaction of complement with *Neisseria meningitidis* and *Neisseria gonorrhoeae*. *Clinical Microbiology Reviews*, 2:11–17, 1989.

[137] G A Jarvis and N A Vedros. Sialic acid of group B *Neisseria meningitidis* regulates alternative complement pathway activation. *Infection and Immunity*, 55(1):174–180, 1987.

[138] D S Stephens. Biology and pathogenesis of the evolutionarily successful, obligate human bacterium *Neisseria meningitidis*. *Vaccine*, 27:B71–7, 2009.

[139] M R Spinosa, C Progida, A Talà, L Cogli, P Alifano, and C Bucci. The *Neisseria meningitidis* capsule is important for intracellular survival in human cells. *Infection and Immunity*, 75(7):3594–603, 2007.

[140] A Jones, M Geörg, L Maudsdotter, and A B Jonsson. Endotoxin, capsule, and bacterial attachment contribute to *Neisseria meningitidis* resistance to the human antimicrobial peptide LL-37. *Journal of Bacteriology*, 191(12):3861–3868, 2009.

[141] M V R Weber, H Claus, M C J Maiden, M Frosch, and U Vogel. Genetic mechanisms for loss of encapsulation in polysialyltransferase-gene-positive meningococci isolated from healthy carriers. *International Journal of Medical Microbiology*, 296(7):475–484, 2006.

[142] B R Levin and J J Bull. Short-sighted evolution and the virulence of pathogenic microorganisms. *Trends in Microbiology*, 2(3):76–81, 1994.

[143] U Vogel and H Claus. The evolution of human pathogens: examples and clinical implications. *International Journal of Medical Microbiology*, 290(6):511–518, 2000.

[144] J P Cowen. Morphological study of marine bacterial capsules: implications for marine aggregates. *Marine Biology*, 114(1):85–95, 1992.

[145] C H Liu, S M Lee, J M VanLare, D L Kasper, and S K Mazmanian. Regulation of surface architecture by symbiotic bacteria mediates host colonization. *Proceedings of the National Academy of Sciences of the United States of America*, 105(10):3951–3956, 2008.

[146] N T Porter, P Canales, D A Peterson, and E C Martens. A subset of polysaccharide capsules in the human symbiont *Bacteroides thetaiotaomicron* promote increased competitive fitness in the mouse gut. *Cell Host and Microbe*, 22(4):494–506, 2017.

[147] U B Skov Sørensen, K Yao, Y Yang, H Tettelin, and M Kilian. Capsular polysaccharide expression in commensal *Streptococcus* species: Genetic and antigenic similarities to *Streptococcus pneumoniae*. *mBio*, 7(6):e01844–16, 2016.

[148] M Ulanova and R S W Tsang. *Haemophilus influenzae* serotype a as a cause of serious invasive infections. *The Lancet Infectious Diseases*, 14(1):70–82, 2014.

[149] R Podschun and U Ullmann. *Klebsiella* spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clinical Microbiology Reviews*, 11(4):589–603, 1998.

[150] M Lappann, J A J Haagensen, H Claus, U Vogel, and S Molin. Meningococcal biofilm formation: structure, development and phenotypes in a standardized continuous flow system. *Molecular Microbiology*, 62(5):1292–1309, 2006.

[151] MenAfriCar-Consortium. The diversity of meningococcal carriage across the African meningitis belt and the impact of vaccination with a group a meningococcal conjugate vaccine. *Journal of Infectious Diseases*, 212(8):1298–1307, 2015.

[152] M Virji. Meningococcal disease: epidemiology and pathogenesis. *Trends in Microbiology*, 4(12):466–469, 1996.

[153] D L Gibson, A P White, S D Snyder, S Martin, C Heiss, P Azadi, M Surette, and W W Kay. *Salmonella* produces an O-antigen capsule regulated by AgfD and important for environmental persistence. *Journal of Bacteriology*, 188(22):7722–30, 2006.

[154] J K Singh, F G Adams, and M H Brown. Diversity and function of capsular polysaccharide in *Acinetobacter baumannii*. *Frontiers in Microbiology*, 9:3301, 2019.

[155] E B Roberson and M K Firestone. Relationship between desiccation and exopolysaccharide production in a soil *Pseudomonas* sp. *Applied and Environmental Microbiology*, 58(4):1284–91, 1992.

[156] D H Limoli, C J Jones, and D J Wozniak. Bacterial extracellular polysaccharides in biofilm formation and function. *Microbiology Spectrum*, 3(3):MB–0011–2014, 2015.

[157] R L Walsh and A Camilli. *Streptococcus pneumoniae* is desiccation tolerant and infectious upon rehydration. *mBio*, 2(3):e00092–11, 2011.

[158] J Díaz, M Cárcamo, M Seoane, P Pidal, G Cavada, R Puentes, S Terrazas, P Araya, A B Ibarz-pavn, M Manríquez, J C Hormazábal, S Ayala, and M T Valenzuela. Prevalence of meningococcal carriage in children and adolescents aged 10-19 years in Chile in 2013. *Journal of Infection and Public Health*, 9(4):506–515, 2016.

[159] P Miller and D Schad. The resistance of meningococci to drying. *Journal of Bacteriology*, 47(1):71–77, 1944.

[160] C L Swain and D R Martin. Survival of meningococci outside of the host: implications for acquisition. *Epidemiology and Infection*, 135(2):315–320, 2007.

[161] Y-L Tzeng, L E Martin, and D S Stephens. Environmental survival of *Neisseria meningitidis*. *Epidemiology and Infection*, 142(1):187–90, 2014.

[162] C. L. Swain, D R Martin, D. Sim, T. W. Jordan, and J. K. MacKichan. Survival of *Neisseria meningitidis* outside of the host: environmental effects and differences among strains. *Epidemiology and Infection*, 145(16):3525–3534, 2017.

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

[164] R Grifantini, E Bartolini, A Muzzi, M Draghi, E Frigimelica, J Berger, F Randazzo, and G Grandi. Gene expression profile in *Neisseria meningitidis* and *Neisseria lactamica* upon host-cell contact. *Annals of the New York Academy of Sciences*, 975(1):202–216, 2002.

[165] S N Bartley, Y-L Tzeng, K Heel, C W Lee, S Mowlaboccus, T Seemann, W Lu, Y-H Lin, C S Ryan, C Peacock, D S Stephens, J K Davies, and C M Kahler. Attachment and invasion of *Neisseria meningitidis* to host cells is related to surface hydrophobicity, bacterial cell size and capsule. *PLoS ONE*, 8(2):e55798, 2013.

[166] K Yi, A W Rasmussen, S K Gudlavalleti, D S Stephens, and I Stojiljkovic. Biofilm formation by *Neisseria meningitidis*. *Infection and Immunity*, 72(10):6132–6138, 2004.

[167] R Brock Neil, J Q Shao, and M A Apicella. Biofilm formation on human airway epithelia by encapsulated *Neisseria meningitidis* serogroup B. *Microbes and Infection*, 11(2):281–287, 2009.

[168] M C J Maiden, A B Ibarz-pavn, R Urwin, S J Gray, N J Andrews, S C Clarke, A M Walker, M R Evans, J S Kroll, K R Neal, D A A Ala'Aldeen, D W Crook, K Cann, S Harrison, R Cunningham, D Baxter, E Kaczmarski, J Maclennan, J Claire, and J M Stuart. Impact of meningococcal serogroup C conjugate vaccines on carriage and herd immunity. *The Journal of Infectious Diseases*, 197(5):737–743, 2008.

[169] M Sadarangani and A J Pollard. Serogroup B meningococcal vaccines-an unfinished story. *The Lancet Infectious Diseases*, 10(2):112–124, 2010.

[170] K Yi, D S Stephens, and I Stojiljkovic. Development and evaluation of an improved mouse model of meningococcal colonization. *Infection and Immunity*, 71(4):1849–1855, 2003.

[171] F G Mackinnon, R Borrow, A R Gorringe, A J Fox, D M Jones, and A Robinson. Demonstration of lipooligosaccharide immunotype and capsule as virulence factors for *Neisseria meningitidis* using an infant mouse intranasal infection model. *Microbial Pathogenesis*, 15(5):359–366, 1993.

[172] K O Johswich, S E McCaw, E Islam, A Sintsova, A Gu, J E Shively, and S D Gray-Owen. *In vivo* adaptation and persistence of *Neisseria meningitidis* within the nasopharyngeal mucosa. *PLoS Pathogens*, 9(7):e1003509, 2013.

[173] N J Weyand. *Neisseria* models of infection and persistence in the upper respiratory tract. *Pathogens and Disease*, 75(3):ftx031, 2017.

[174] E Loh, E Kugelberg, A Tracy, Q Zhang, B Gollan, H Ewles, R Chalmers, V Pelicic, and C M Tang. Temperature triggers immune evasion by *Neisseria meningitidis*. *Nature*, 502:237–240, 2013.

[175] J Kortmann and F Narberhaus. Bacterial RNA thermometers: molecular zippers and switches. *Nature Reviews Microbiology*, 10(4):255–65, 2012.

[176] R P Barnwal, E Loh, K S Godin, J Yip, H Lavender, C M Tang, and G Varani. Structure and mechanism of a molecular rheostat, an RNA thermometer that modulates immune evasion by *Neisseria meningitidis*. *Nucleic Acids Research*, 44(19):9426–9437, 2016.

[177] O Lam, J Wheeler, and C M Tang. Thermal control of virulence factors in bacteria: a hot topic. *Virulence*, 5(8):852–62, 2014.

[178] A Schmidt, A Belaaouaj, R Bissinger, G Koller, L Malleret, C D'Orazio, M Facchinelli, B Schulte-Hubbert, A Molinaro, O Holst, J Hammermann, M Schniederjans, K C Meyer, S Damkiaer, G Piacentini, B Assael, K Bruce, S Hausler, J J LiPuma, J Seelig, D Worlitzsch, and G Doring. Neutrophil elastase-mediated increase in airway temperature during inflammation. *Journal of Cystic Fibrosis*, 13(6):623–631, 2014.

[179] L H Harrison, C W Armstrong, S R Jenkins, M W Harmon, G W Ajello, G B Miller, and C V Broome. A cluster of meningococcal disease on a school bus following epidemic influenza. *Archives of Internal Medicine*, 151(5):1005–1009, 1991.

[180] P Makras, S Alexiou-Daniel, A Antoniadis, and D Hatzigeorgiou. Outbreak of meningococcal disease after an influenza B epidemic at a Hellenic air force recruit training center. *Clinical Infectious Diseases*, 33(6):e48–e50, 2002.

[181] K A V Cartwright, D M Jones, A J Smith, J M Stuart, E B Kaczmarski, and S R Palmer. Influenza A and meningococcal disease. *The Lancet*, 338(8766):554–557, 1989.

[182] J H Jacobs, C Viboud, E T Tchetgen, J Schwartz, C Steiner, L Simonsen, and M Lipsitch. The association of meningococcal disease with influenza in the united states, 1989-2009. *PLoS ONE*, 9(9):e107486, 2014.

[183] B Hubert, L Watier, P Garnerin, and S Richardson. Meningococcal disease and influenza-like syndrome: a new approach to an old question. *Journal of Infectious Diseases*, 166(3):542–545, 1992.

[184] A R Tuite, L M Kinlin, S P Kuster, F Jamieson, J C Kwong, A McGeer, and D N Fisman. Respiratory virus infection and risk of invasive meningococcal disease in central Ontario, Canada. *PLoS ONE*, 5(11):e15493, 2010.

[185] R Allard, M Couillard, P Pilon, M Kafka, and L Bédard. Invasive bacterial infections following influenza: A time-series analysis in Montréal, Canada, 1996-2008. *Influenza and other Respiratory Viruses*, 6(4):268–275, 2012.

[186] K A Dunlop, P V Coyle, P Jackson, C C Patterson, and M D Shields. Respiratory viruses do not trigger meningococcal disease in children. *Journal of Infection*, 54(5):454–458, 2007.

[187] J M Stuart, K Cartwright, and N J Andrews. Respiratory syncytial virus infection and meningococcal disease. *Epidemiology and Infection*, 117(1):107–111, 1996.

[188] J E Mueller, M Woringer, S Porgho, Y Madec, H Tall, N Martiny, and B W Bicaba. The association between respiratory tract infection incidence and localised meningitis epidemics: an analysis of high-resolution surveillance data from Burkina Faso. *Scientific Reports*, 7:11570, 2017.

[189] M A Rameix-Welti, M L Zarantonelli, D Giorgini, C Ruckly, M Marasescu, S Van Der Werf, J M Alonso, N Naffakh, and M-K Taha. Influenza A virus neuraminidase enhances meningococcal adhesion to epithelial cells through interaction with sialic acid-containing meningococcal capsules. *Infection and Immunity*, 77(9):3588–3595, 2009.

[190] S A Shah, S M Ganesan, S Varadharaj, S M Dabdoub, J D Walters, and P S Kumar. The making of a miscreant: tobacco smoke and the creation of pathogen-rich biofilms. *NPJ Biofilms and Microbiomes*, 3:26, 2017.

[191] K Reimann. Attempts to demonstrate a polysaccharide capsule in *Neisseria gonorrhoeae*. *APMIS*, 96(8):735–740, 1988.

[192] K F Starr, E A Porsch, C Heiss, I Black, P Azadi, and J W St. Geme. Characterization of the *Kingella kingae* polysaccharide capsule and exopolysaccharide. *PLoS ONE*, 8(9):e75409, 2013.

[193] C Schoen, H Tettelin, J Parkhill, and M Frosch. Genome flexibility in *Neisseria meningitidis*. *Vaccine*, 27:B102–B111, 2009.

[194] T Ikemura. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli*. *Journal of Molecular Biology*, 151(3):389–409, 1981.

[195] J Hacker, G Blum-Oehler, I Mühldorfer, and H Tschäpe. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Molecular Microbiology*, 23(6):1089–1097, 1997.

[196] B R Clarke, R Pearce, and I S Roberts. Genetic organization of the *Escherichia coli* K10 capsule gene cluster: Identification and characterization of two conserved regions in group III capsule gene clusters encoding

polysaccharide transport functions. *Journal of Bacteriology*, 181(7):2279–2285, 1999.

[197] S M Nielsen, C de Gier, C Dimopoulou, V Gupta, L H Hansen, and N Nørskov-Lauritsen. The capsule biosynthesis locus of *Haemophilus influenzae* shows conspicuous similarity to the corresponding locus in *Haemophilus sputorum* and may have been recruited from this species by horizontal gene transfer. *Microbiology*, 161(6):1182–1188, 2015.

[198] H Ito. The genetic organization of the capsular polysaccharide biosynthesis region of *Actinobacillus pleuropneumoniae* serotype 14. *Journal of Veterinary Medical Science*, 77(5):583–586, 2015.

[199] Z Xu, P Du, B Zhu, L Xu, H Wang, Y Gao, H Zhou, W Zhang, C Chen, and Z Shao. Phylogenetic study of clonal complex (CC)198 capsule null locus (*cnl*) genomes: A distinctive group within the species *Neisseria meningitidis*. *Infection, Genetics and Evolution*, 34:372–377, 2015.

[200] J A Vázquez, L de la Fuente, S Berron, M O'Rourke, N H Smith, J Zhou, and B G Spratt. Ecological separation and genetic isolation of *Neisseria gonorrhoeae* and *Neisseria meningitidis*. *Current Biology*, 3(9):567–572, 1993.

[201] J S Kroll, K E Wilks, J L Farrant, and P R Langford. Natural genetic exchange between *Haemophilus* and *Neisseria*: Intergeneric transfer of chromosomal genes between major human pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, 95(21):12381–12385, 1998.

[202] T-T Lam, H Claus, M Frosch, and U Vogel. Sequence analysis of serotype-specific synthesis regions II of *Haemophilus influenzae* serotypes c and d: evidence for common ancestry of capsule synthesis in *Pasteurellaceae* and *Neisseria meningitidis*. *Research in Microbiology*, 162(5):483–487, 2011.

[203] O Rendueles, M Garcia-Garcerà, B Néron, M Touchon, and E P C Rocha. Abundance and co-occurrence of extracellular capsules increases environmental breadth: implications for the emergence of pathogens. *PLoS Pathogens*, 13(7):e1006525, 2017.

[204] M E A Clemence, M C J Maiden, and O B Harrison. Characterization of capsule genes in non-pathogenic *Neisseria* species. *Microbial Genomics*, 4(9):e000208, 2018.

[205] A Unkmeir, U Kämmerer, A Stade, C Hübner, S Haller, A Kolb-Mäurer, M Frosch, and G Dietrich. Lipooligosaccharide and polysaccharide capsule: virulence factors of *Neisseria meningitidis* that determine meningococcal interaction with human dendritic cells. *Infection and Immunity*, 70(5):2454–62, 2002.

[206] L Cuthbertson, V Kos, and C Whitfield. ABC transporters involved in export of cell surface glycoconjugates. *Microbiology and Molecular Biology Reviews*, 74(3):341–362, 2010.

[207] K Rutherford, J Parkhill, J Crook, T Horsnell, P Rice, M A Rajandream, and B Barrell. Artemis: sequence visualization and annotation. *Bioinformatics*, 16(10):944–945, 2000.

[208] S Kumar, G Stecher, and K Tamura. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7):1870–1874, 2016.

[209] NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research*, 45:D12–D17, 2017.

[210] R D Finn, P Coggill, R Y Eberhardt, S R Eddy, J Mistry, A L Mitchell, S C Potter, M Punta, M Qureshi, A Sangrador-Vegas, G A Salazar, J Tate, and

A Bateman. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Research*, 44(D1):D279–D285, 2016.

[211] P Rice, L Longden, and A Bleasby. EMBOSS: the European molecular biology open software suite. *Trends in Genetics*, 16(6):276–277, 2000.

[212] J C Abbott, D M Aanensen, and S D Bentley. WebACT: An Online Genome Comparison Suite. In N H Bergman and J M Walker, editors, *Methods in Molecular Biology*, chapter 1.4, pages 57–74. Humana Press Inc., Totowa, 2007.

[213] K Katoh, K Misawa, K Kuma, and T Miyata. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, 30(14):3059–3066, 2002.

[214] S Guindon, J F Dufayard, V Lefort, M Anisimova, W Hordijk, and O Gascuel. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3):307–321, 2010.

[215] D Darriba, G L Taboada, R Doallo, and D Posada. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9(8):772, 2012.

[216] X Didelot and D J Wilson. ClonalFrameML: efficient inference of recombination in whole bacterial genomes. *PLoS Computational Biology*, 11(2):e1004041, 2015.

[217] J Huerta-Cepas, F Serra, and P Bork. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution*, 33(6):1635–1638, 2016.

[218] X Wang, Y Liu, J Chen, X Jiao, P Wang, Y Zhang, H Lin, and M Liu. Purification of group C meningococcal capsular polysaccharide by chromatography as well as safety and immunogenicity of prepared polysaccharide-protein conjugate. *Chinese Journal of Biologicals*, 26(12):1705–1714, 2013.

[219] A González-Díaz, F Tubau, M Pinto, Y Sierra, M Cubero, J Càmara, J Ayats, P Bajanca-Lavado, C Ardanuy, and S Marti. Identification of polysaccharide capsules among extensively drug-resistant genitourinary *Haemophilus parainfluenzae* isolates. *Scientific Reports*, 9(1):4481, 2019.

[220] J T Bosse, Y Li, R Fernandez Crespo, S Lacouture, M Gottschalk, R Sárközi, L Fodor, M Casas Amoribieta, Ø Angen, K Nedbalcova, M T G Holden, D J Maskell, A W Tucker, B W Wren, A N Rycroft, and P R Langford. Comparative sequence analysis of the capsular polysaccharide loci of *Actinobacillus pleuropneumoniae* serovars 1–18, and development of two multiplex PCRs for comprehensive capsule typing. *Veterinary Microbiology*, 220:83–89, 2018.

[221] M Ma, D A Powell, N J Weyand, K A Rhodes, M A Rendón, J A Frelinger, and M So. A natural mouse model for *Neisseria* colonization. *Infection and Immunity*, 86(5):e00839–17, 2018.

[222] D R Bundle, H J Jennings, and C P Kenny. An improved procedure for the isolation of meningococcal, polysaccharide antigens, and the structural determination of the antigen from serogroup X. *Carbohydrate Research*, 26(1):268–270, 1973.

[223] D R Bundle, I C P Smith, and HJ Jennings. Determination of the structure and conformation of bacterial polysaccharides by carbon 13 nuclear magnetic

resonance. Studies on the group-specific antigens of *Neisseria meningitidis* serogropus A and X. *Journal of Biological Chemistry*, 249(7):2275–2281, 1974.

[224] C Hill. Virulence or niche factors: What's in a name? *Journal of Bacteriology*, 194(21):5725–5727, 2012.

[225] E Lerat, V Daubin, H Ochman, and N A Moran. Evolutionary origins of genomic repertoires in bacteria. *PLoS Biology*, 3(5):e130, 2005.

[226] M W J Van Passel, P R Marri, and H Ochman. The emergence and fate of horizontally acquired genes in *Escherichia coli*. *PLoS Computational Biology*, 4(4):e1000059, 2008.

[227] M E A Clemence, O B Harrison, and M C J Maiden. *Neisseria meningitidis* has acquired sequences within the capsule locus by horizontal genetic transfer [version 2; peer review: 2 approved]. *Wellcome Open Research*, 4:99, 2019.

[228] B G Spratt, L D Bowler, Q Y Zhang, J Zhou, and J M Smith. Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal *Neisseria* species. *Journal of Molecular Evolution*, 34(2):115–125, 1992.

[229] P Zhu, M J Klutch, M C Bash, R S W Tsang, L K Ng, and C M Tsai. Genetic diversity of three *lgt* loci for biosynthesis of lipooligosaccharide (LOS) in *Neisseria* species. *Microbiology*, 148(6):1833–1844, 2002.

[230] Y Qvarnstrom and G Swedberg. Variations in gene organization and DNA uptake signal sequence in the *folP* region between commensal and pathogenic *Neisseria* species. *BMC Microbiology*, 6:11, 2006.

[231] B Linz, M Schenker, P Zhu, and M Achtman. Frequent interspecific genetic exchange between commensal neisseriae and *Neisseria meningitidis*. *Molecular Microbiology*, 36(5):1049–1058, 2000.

[232] M E Wörmann, C L Horien, J S Bennett, K A Jolley, M C J Maiden, C M Tang, E L Aho, and R M Exley. Sequence, distribution and chromosomal context of class I and class II pilin genes of *Neisseria meningitidis* identified in whole genome sequences. *BMC Genomics*, 15:253, 2014.

[233] H B Bratcher, J C Cameron, C M C Rodrigues, A Finn, M Wootton, A Smith, P Heath, S Ladhani, R Cunningham, M D Snape, A J Pollard, R Borrow, C Trotter, J M Maclennan, S J Gray, and M C J Maiden. UKMenCar4 : A cross-sectional survey of asymptomatic meningococcal carriage amongst UK adolescents at a period of low invasive meningococcal disease incidence [version 1; peer review: awaiting peer review]. *Wellcome Open Research*, 4:118, 2019.

[234] S Budroni, E Siena, J C Dunning, K L Seib, D Serruto, C Nofroni, M Comanducci, D R Riley, S C Daugherty, S V Angiuoli, A Covacci, M Pizza, R Rappuoli, E R Moxon, H Tettelin, and D Medini. *Neisseria meningitidis* is structured in clades associated with restriction modification systems that modulate homologous recombination. *Proceedings of the National Academy of Sciences of the United States of America*, 108(11):4494–4499, 2011.

[235] C Schoen, J Weber-Lehmann, J Blom, B Joseph, A Goesmann, A Strittmatter, and M Frosch. Whole-genome sequence of the transformable *Neisseria meningitidis* serogroup A strain WUE2594. *Journal of Bacteriology*, 193(8):2064–2065, 2011.

[236] D A Benson, M Cavanaugh, K Clark, I Karsch-Mizrachi, D J Lipman, J Ostell, and E W Sayers. GenBank. *Nucleic Acids Research*, 41:D36–D42, 2017.

[237] R C Edgar. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797, 2004.

[238] D Bryant and V Moulton. Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution*, 21(2):255–265, 2004.

[239] D P Martin, B Murrell, M Golden, A Khoosal, and B Muhire. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, 1(1):vev003, 2015.

[240] M O Salimen, J K Carr, D S Burke, and F E McCutchan. Identification of Breakpoints in Intergenotypic Recombinants of HIV Type 1 by Bootscanning. *AIDS Research and Human Retroviruses*, 11(11):1423–1425, 1995.

[241] T H Jukes and C R Cantor. Evolution of Protein Molecules. In H N Munro, editor, *Mammalian Protein Metabolism III*, chapter 24, pages 21–132. Academic Press, New York, 1969.

[242] R Furuya, Y Onoye, A Kanayama, T Saika, T Iyoda, M Tatewaki, K Matsuzaki, I Kobayashi, and M Tanaka. Antimicrobial resistance in clinical isolates of *Neisseria subflava* from the oral cavities of a Japanese population. *Journal of Infection and Chemotherapy*, 13(5):302–304, 2007.

[243] A J Beddek, M-S Li, J S Kroll, T W Jordan, and D R Martin. Evidence for capsule switching between carried and disease-causing *Neisseria meningitidis* strains. *Infection and Immunity*, 77(7):2989–2994, 2009.

[244] M Frosch, U Edwards, K Bousset, B Krausse, and C Weisgerber. Evidence for a common molecular origin of the capsule gene loci in Gram-negative bacteria expressing group II capsular polysaccharides. *Molecular Microbiology*, 5(5):1251–63, 1991.

[245] U Vogel, H Claus, and M Frosch. Genetic lineages and their traits in *Neisseria meningitidis*. *International Journal of Medical Microbiology*, 294:75–82, 2004.

[246] E V Koonin, K S Makarova, and L Aravind. Horizontal gene transfer in prokaryotes: quantification and classification. *Annual Review of Microbiology*, 55:709–742, 2001.

[247] S P Yazdankhah, P Kriz, G Tzanakaki, J Kremastinou, J Kalmusova, M Musilek, T Alvestad, K A Jolley, D J Wilson, N D McCarthy, D A Caugant, and M C J Maiden. Distribution of serogroups and genotypes among disease-associated and carried isolates of *Neisseria meningitidis* from the Czech Republic, Greece, and Norway. *Journal of Clinical Microbiology*, 42(11):5146–5153, 2004.

[248] L Ambrosio, A Neri, C Fazio, G M Rossolini, P Vacca, E Riccobono, F Voller, A Miglietta, and P S Id. Genomic analysis of *Neisseria meningitidis* carriage isolates during an outbreak of serogroup C clonal complex 11, Tuscany, Italy. *PLoS ONE*, 14(5):e0217500, 2019.

[249] P S Rubilar, G N Barra, J M Gabastou, P Alarcón, P Araya, J C Hormazábal, and J Fernandez. Increase of *Neisseria meningitidis* W:cc11 invasive disease in Chile has no correlation with carriage in adolescents. *PLoS ONE*, 13(3):e0193572, 2018.

[250] N Rius Gordillo, C Fernández San José, S González Peris, and FA Moraga-Llop. Invasive meningitis due to serogroup 29E meningococci [Enfermedad invasora por meningococo del serogrupo 29E]. *Anales de Pediatria*, 78(1):68–69, 2013.

[251] E Wachter, A E Brown, T E Kiehn, B J Lee, and D Armstrong. *Neisseria meningitidis* serogroup 29E (Z') septicemia in a patient with far advanced multiple myeloma (plasma cell leukemia). *Journal of Clinical Microbiology*, 21(3):464–466, 1985.

[252] R Whittaker, J G Dias, M Ramliden, C Ködmön, A Economopoulou, N Beer, and L Pastore Celentano. The epidemiology of invasive meningococcal disease in EU/EEA countries, 2004–2014. *Vaccine*, 35(16):2034–2041, 2017.

[253] E Severi, D W Hood, and G H Thomas. Sialic acid utilization by bacterial pathogens. *Microbiology*, 153(9):2817–2822, 2007.

[254] U Vogel, A Weinberger, R Frank, A Müller, J Köhl, J P Atkinson, and M Frosch. Complement factor C3 deposition and serum resistance in isogenic capsule and lipooligosaccharide sialic acid mutants of serogroup B *Neisseria meningitidis*. *Infection and Immunity*, 65(10):4022–4029, 1997.

[255] M M Giuliani, J Adu-Bobie, M Comanducci, B Aricò, S Savino, L Santini, B Brunelli, S Bambini, A Biolchi, B Capecchi, E Cartocci, L Ciucchi, F Di Marcello, F Ferlicca, B Galli, E Luzzi, V Masignani, D Serruto, D Veggi, M Contorni, M Morandi, A Bartalesi, V Cinotti, D Mannucci, F Titta, E Ovidi, J A Welsch, D Granoff, R Rappuoli, and M Pizza. A universal vaccine for serogroup B meningococcus. *Proceedings of the National Academy of Sciences of the United States of America*, 103(29):10834–9, 2006.

[256] G W Zlotnick, T R Jones, P Liberator, L Hao, S Harris, L K McNeil, D Zhu, J Perez, J Eiden, K U Jansen, and A S Anderson. The discovery and development of a novel naccine to protect against *Neisseria meningitidis* serogroup B disease. *Human Vaccines & Immunotherapeutics*, 11(1):5–13, 2014.

[257] E Lujan, E Partridge, S Giuntini, S Ram, and D M Granoff. Breadth and duration of meningococcal serum bactericidal activity in health care workers and microbiologists immunized with the MenB-fHbp vaccine. *Clinical and Vaccine Immunology*, 24:e00121–17, 2017.

[258] L Ostergaard, T Vesikari, J Absalon, J Beeslaar, B J Ward, S Senders, J J Eiden, K U Jansen, A S Anderson, L J York, T R Jones, S L Harris, R O'Neill, D Radley, R Maansson, J-L Prégaldien, J Ginis, N B Staerke, and J L Perez. A bivalent meningococcal B vaccine in adolescents and young adults. *New England Journal of Medicine*, 377(24):2349–2362, 2017.

[259] M-K Taha, J C Hawkins, P Liberator, A-E Deghmane, L Andrew, Li Hao, T R Jones, L K McNeil, R E O'Neill, J L Perez, K U Jansen, and A S Anderson. Bactericidal activity of sera from adolescents vaccinated with bivalent rLP2086 against meningococcal serogroup B outbreak strains from France. *Vaccine*, 35(11):1530–1537, 2017.

[260] D A Caugant. Genetics and evolution of *Neisseria meningitidis*: Importance for the epidemiology of meningococcal disease. *Infection, Genetics and Evolution*, 8(5):558–565, 2008.

[261] I Stojijkovic, V Hwa, L de Saint Martin, P O'Gaora, X Nassif, F Heffron, and M So. The *Neisseria meningitidis* haemoglobin receptor: its role in iron utilization and virulence. *Molecular Microbiology*, 15(3):531–541, 1995.

[262] M C Schneider, R M Exley, H. Chan, I Feavers, Y-H Kang, R B Sim, and C M Tang. Functional significance of factor H binding to *Neisseria meningitidis*. *Journal of Immunology*, 176(12):7566–7575, 2006.

[263] S Kumar, G Stecher, M Li, C Knyaz, and K Tamura. MEGA X : Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, 35(6):1547–1549, 2018.

[264] Salvador Capella-Gutiérrez, José M. Silla-Martínez, and Toni Gabaldón. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15):1972–1973, 2009.

[265] K P Schliep. phangorn: Phylogenetic analysis in R. *Bioinformatics*, 27(4):592–593, 2011.

[266] L J Revell. phytools: An R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution*, 3(2):217–223, 2012.

[267] L Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[268] D Firth. Bias Reduction of Maximum Likelihood Estimates. *Biometrika*, 80(1):27–38, 1993.

[269] R Rappuoli, M Pizza, V Masignani, and K Vadivelu. Meningococcal B vaccine (4CMenB): the journey from research to real world experience. *Expert Review of Vaccines*, 17(12):1111–1121, 2018.

[270] M C J Maiden and M Frosch. Can we, should we, eradicate the meningococcus? *Vaccine*, 30:52–56, 2012.

[271] H Marshall, M McMillan, A Koehler, A Lawrence, J MacLennan, M C J Maiden, M Ramsay, S Ladhani, C Trotter, R Borrow, A Finn, T Sullivan, P Richmond, C M Kahler, J Whelan, and K Vadivelu. Impact of MenB vaccine on pharyngeal carriage of *Neisseria meningitidis* in adolescents and implementation of a MenB vaccine program for infants and adolescents in South Australia. In *The 15th Congress of the European Meningococcal and Haemophilus Disease Society*, Lisbon, 2019.

[272] U Jounio, A Saukkoriipi, H B Bratcher, A Bloigu, R Juvonen, S Silvennoinen-Kassinen, A Peitso, T Harju, O Vainio, M Kuusi, M C J Maiden, M Leinonen, H Käyhty, and M Toropainen. Genotypic and phenotypic characterization of carriage and invasive disease isolates of *Neisseria meningitidis* in Finland. *Journal of Clinical Microbiology*, 50(2):264–273, 2012.

[273] A B Brueggemann, D T Griffiths, E Meats, T Peto, D W Crook, and B G Spratt. Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. *Journal of Infectious Diseases*, 187(9):1424–1432, 2003.

[274] R P Wenzel, J A Davies, J R Mitzel, and W E Beam. Non-usefulness of meningococcal carriage-rates. *The Lancet*, 302(7822):195, 1973.

[275] E A Edwards, L F Devine, C H Sengbusch, and H W Ward. Immunological investigations of meningococcal disease III: brevity of group C acquisition prior to disease occurrence. *Scandinavian Journal of Infectious Diseases*, 9(2):105–110, 1977.

[276] K Schipper and A Van der Ende. The zebrafish embryo as an infection model for *Neisseria meningitidis*. In *The 15th Congress of the European Meningococcal and Haemophilus Disease Society*, Lisbon, 2019.

[277] M Ravenhall, N Škunca, F Lassalle, and C Dessimoz. Inferring Horizontal Gene Transfer. *PLoS Computational Biology*, 11(5):e1004095, 2015.

[278] M Mustapha, A P de Lemos, M Pacey, JW Marsh, C Siebra, L Timm, L H Harrison, and C T Sacchi. Two cases of a newly characterized *Neisseria* species [abstract OP176]. In *The 21st International Pathogenic Neisseria Conference*, Asilomar, CA, 2018.

[279] C O Buckee, S Gupta, P Kriz, M C J Maiden, and K A Jolley. Long-term evolution of antigen repertoires among carried meningococci. *Proceedings of the Royal Society B: Biological Sciences*, 277(1688):1635–1641, 2010.

[280] J W Eerkens, R V Nichols, G G R Murray, K Perez, E Murga, P Kaijankoski, J S Rosenthal, L Engbring, and B Shapiro. A probable prehistoric case of meningococcal disease from San Francisco Bay: Next generation sequencing of

Neisseria meningitidis from dental calculus and osteological evidence. *International Journal of Paleopathology*, 22:173–180, 2018.

Appendices

A

Isolates used in this thesis

Table A.1: Isolates used in Chapter 2. For a full list of isolate details, please refer to the dataset deposited on FigShare at `https://doi.org/10.6084/m9.figshare.6016112.v1` under the tab "pubMLST isolates".

Isolate	Species	PubMed ID	PubMLST ID
CCUG 808	*N. animalis*	24097834	19940
DSM 21642	*N. animaloris*		30325
CCUG 50858 T	*N. bacilliformis*	22422752	19077
CCUG 30380	*N. bacilliformis*	24097834	21038
CCUG 38158	*N. bacilliformis*	24097834	21039
CCUG 50611	*N. bacilliformis*	24097834	21040
15883	*'N. bergeri'*	22422752	49339
CCUG 56775 T	*N. canis*	22422752	49340
ATCC 14685	*N. cinerea*	20676376	14731
CCUG 53043	*N. cinerea*	24097834	21041
CCUG 346 T	*N. cinerea*	22422752	49341
CCUG 5746	*N. cinerea*	22422752	49342
CCUG 53898	*N. dentiae*	22422752	49345
CCUG 4554	*N. elongata*	24097834	21042
ATCC 29315	*N. elongata* subsp. *glycolytica*	20676376	14740
CCUG 2043T	*N. elongata* subsp. *elongata*	24097834	20516
CCUG 30802T	*N. elongata* subsp. *nitroreducens*	24097834	20515
FA1090	*N. gonorrhoea*		2855
NCCP11945	*N. gonorrhoea*	18586945	13685
FA19	*N. gonorrhoea*	26358608	46275
FA6140	*N. gonorrhoea*	26358608	46276
ATCC 23970	*N. lactamica*		5544
004-12	*N. lactamica*	24097834	8778
017-02	*N. lactamica*	24097834	8837
020-06	*N. lactamica*	21092259	8851

14	*N. meningitidis*	18305155	30
MC58	*N. meningitidis*	10710307	240
Z2491	*N. meningitidis*	10761919	613
FAM18	*N. meningitidis*	17305430	698
053442	*N. meningitidis*	18031983	12672
ATCC 29256	*N. mucosa*	22422752	2863
ATCC 25996	*N. mucosa*	20676376	3565
ATCC 9913	*N. mucosa*	24097834	5197
ATCC 19696	*N. mucosa*	24097834	5354
CCUG 26474	*N. mucosa*	24097834	21043
CCUG 41451 T	*N. mucosa*	22422752	49351
CCUG 12106	*N. mucosa*	22422752	49352
CCUG 431	*N. mucosa*	22422752	49360
CCUG 24847	*N. mucosa*	22422752	49361
AP2031	*N. musculi*	27298306	29520
CCUG 26878 T	*N. oralis*	22422752	19091
F0314	*N. oralis*		21044
CCUG 10421	*N. oralis*	24097834	21045
CCUG 804	*N. oralis*	24097834	21046
ATCC 43768	*N. polysaccharea*	20676376	14730
CCUG 4790	*N. polysaccharea*	24097834	21047
CCUG 24846	*N. polysaccharea*	22422752	49358
CCUG 27182	*N. polysaccharea*	22422752	49359
871	*N. shayeganii*		21048
NRL30031	*N. subflava*	20676376	14732
NJ9703	*N. subflava*	20676376	14733
C102	*N. subflava*		21060
CCUG 24841	*N. subflava*	23813677	21063
CCUG 24844	*N. subflava*	23813677	21064
12007_2012	*N. subflava*		26870
C6A	*N. subflava*		46753
CCUG 4788	*N. subflava*	22422752	49346
CCUG 29761	*N. subflava*	22422752	49347
CCUG 17913 T	*N. subflava*	22422752	49348
CCUG 806	*N. subflava*	22422752	49349
CCUG 7826	*N. subflava*	22422752	49353
CCUG 24918	*N. subflava*	22422752	49362
CCUG 23930 T	*N. subflava*	22422752	49363
CCUG 800	*N. subflava*	22422752	49364
CCUG 801	*N. subflava*	22422752	49365
CCUG 24960	*N. subflava*	22422752	49366
CCUG 25198	*N. subflava*	22422752	49367
SK114	*N. subflava*		49373
9715	*N. wadsworthii*		21049
CCUG 4007 T	*N. weaveri*	22422752	49368
NCTC 12230 T	*N. zoodegmatis*		36317

Table A.2: Isolates used in Chapter 3. For a full list of isolate details, please refer to the dataset deposited on FigShare at `https://doi.org/10.6084/m9.figshare.8256572.v1`. cc indicates clonal complex.

Isolate	Species	Disease	Capsule	cc	PubMed ID	PubMLST or GenBank ID
CCUG 808	*N. animalis*		NPN		24097834	19940
DSM 21642	*N. animaloris*		NPN			30325
CCUG 50858 T	*N. bacilliformis*	other	NPN		22422752	19077
CCUG 30380	*N. bacilliformis*		NPN		24097834	21038
CCUG 38158	*N. bacilliformis*		NPN		24097834	21039
CCUG 50611	*N. bacilliformis*		NPN		24097834	21040
15883	*'N. bergeri'*	carrier	cnl		22422752	49339
CCUG 56775 T	*N. canis*		NPN		22422752	49340
ATCC 14685	*N. cinerea*		cnl			14731
CCUG 53043	*N. cinerea*		cnl		24097834	21041
OX40632	*N. cinerea*	carrier	cnl			44493
CCUG 346 T	*N. cinerea*		cnl		22422752	49341
CCUG 5746	*N. cinerea*		cnl		22422752	49342
CCUG 53898	*N. dentiae*		NPN		22422752	49345
CCUG 4554	*N. elongata*		cnl		24097834	21042
ATCC 29315	*N. elongata* subsp. *glycolytica*		cnl		20676376	14740
CCUG 2043T	*N. elongata* subsp. *elongata*	other	NPN		24097834	20516
CCUG 30802T	*N. elongata* subsp. *nitroreducens*	other	NPN		24097834	20515
FA1090	*N. gonorrhoeae*		cnl			2855
NCCP11945	*N. gonorrhoeae*		cnl		18586945	13685
DGI2	*N. gonorrhoeae*		cnl			15698
FA19	*N. gonorrhoeae*		cnl		26358608	46275
004-12	*N. lactamica*	carrier	cnl	cc595	24097834	8778
020-06	*N. lactamica*	carrier	cnl	cc640	21092259	8851
ST42026	*N. lactamica*	carrier	cnl	cc613		44268
CA44314	*N. lactamica*	carrier	cnl	cc624		49630
PL41991	*N. lactamica*	carrier	cnl	cc1494		50939
H44/76	*N. meningitidis*	invasive	B	cc32	25523208	237
500	*N. meningitidis*		C	cc11	25523208	343
Z2491	*N. meningitidis*	meningitis	A	cc4	25523208	613
WUE 2594	*N. meningitidis*	meningitis	A	cc5	21296965	19260
M01-240355	*N. meningitidis*		B	cc213	21368196	19265
M10 240675	*N. meningitidis*	invasive	Y	cc167		20059
M11 240073	*N. meningitidis*	invasive	Y	cc174		20232
M11 240437	*N. meningitidis*	invasive	Y	cc92		20456
M11 240441	*N. meningitidis*	invasive	W	cc174		20458
M11 240488	*N. meningitidis*	invasive	B	cc865		21125
M11 240593	*N. meningitidis*	invasive	B	cc269		21143
M12 240332	*N. meningitidis*	invasive	Y	cc103		21505
M13 240398	*N. meningitidis*	invasive	C	cc103		28101
10.2397.V	*N. meningitidis*	invasive	B	cc35		35286
M15 240053	*N. meningitidis*	invasive	B	cc103		37832
M15 240201	*N. meningitidis*	invasive	B	cc41/44		37945
M15 240355	*N. meningitidis*	invasive	C	cc32		38055
M15 240850	*N. meningitidis*	invasive	Y	cc23		39440
M15 240953	*N. meningitidis*	invasive	NG	cc175		41526
ST41111	*N. meningitidis*	carrier	cnl	cc1117		44160
M16 240244	*N. meningitidis*	invasive	B	cc1157		44723
GL40098	*N. meningitidis*	carrier	cnl	cc1136		46548
GL42620	*N. meningitidis*	carrier	B	cc865		46694
B6116/77	*N. meningitidis*	invasive	B	cc8	25523208	47014
CA40118	*N. meningitidis*	carrier	B	cc41/44		49395
CA41628	*N. meningitidis*	carrier	cnl	cc53		49517
CA41900	*N. meningitidis*	carrier	Y	cc103		49544
CA41967	*N. meningitidis*	carrier	Y	cc167		49552
LO40728	*N. meningitidis*	carrier	Y	cc174		49696
PL41818	*N. meningitidis*	carrier	NG	cc175		49960
WG40024	*N. meningitidis*	carrier	B	cc1157		50115
WG40823	*N. meningitidis*	carrier	Y	cc23		50138

ST41713	*N. meningitidis*	carrier	Y	cc32		50361
ST41884	*N. meningitidis*	carrier	Y	cc1157		50391
ST41910	*N. meningitidis*	carrier	Z	cc1157		50394
ST42119	*N. meningitidis*	carrier	cnl	cc192		50414
CA40160	*N. meningitidis*	carrier	cnl	cc198		52562
ST41648	*N. meningitidis*	carrier	B	cc35		52695
α707	*N. meningitidis*	carrier	E	cc254	23628376	HF562982.1
WUE 3608	*N. meningitidis*	carrier	L		23628376	HF562986.1
α275	*N. meningitidis*	invasive	W	cc22	23628376	HF562987.1
α388	*N. meningitidis*	carrier	X	cc254	23628376	HF562988.1
WUE 173	*N. meningitidis*		Z		23628376	HF562991.1
WUE 171	*N. meningitidis*		W	cc11	23628376	HF562992.1
ATCC 9913	*N. mucosa*	carrier	cnl		24097834	5197
ATCC 19696	*N. mucosa*	carrier	cnl		24097834	5354
CCUG 26474	*N. mucosa*		cnl		24097834	21043
CCUG 41451 T	*N. mucosa*	carrier	cnl		22422752	49351
CCUG 12106	*N. mucosa*		cnl		22422752	49352
CCUG 431	*N. mucosa*		NPN		22422752	49360
CCUG 24847	*N. mucosa*		cnl		22422752	49361
AP2031	*N. musculi*		cnl		27298306	29520
CCUG 26878 T	*N. oralis*	carrier	NPN			19091
F0314	*N. oralis*		NPN			21044
CCUG 10421	*N. oralis*		NPN		24097834	21045
CCUG 804	*N. oralis*		NPN		24097834	21046
ATCC 43768	*N. polysaccharea*		cnl			14730
CCUG 4790	*N. polysaccharea*		cnl		24097834	21047
ST41786	*N. polysaccharea*	carrier	cnl			44084
BR40761B	*N. polysaccharea*	carrier	cnl			44102
OX42150	*N. polysaccharea*	carrier	cnl			44630
PL40288	*N. polysaccharea*	carrier	cnl			46727
CCUG 24846	*N. polysaccharea*	carrier	cnl		22422752	49358
CCUG 27182	*N. polysaccharea*		cnl		22422752	49359
871	*N. shayeganii*		cnl			21048
NJ9703	*N. subflava*		NPN		22422752	14733
BR40103	*N. subflava*	carrier	NPN			43570
BR41637	*N. subflava*	carrier	NPN			43658
BR40599	*N. subflava*	carrier	partial			43740
BR40608	*N. subflava*	carrier	NPN			43741
BR40664	*N. subflava*	carrier	NPN			43793
BR41470	*N. subflava*	carrier	NPN			43794
ST41782	*N. subflava*	carrier	NPN			44065
OX41971	*N. subflava*	carrier	NPN			44613
OX42005	*N. subflava*	carrier	NPN			44615
OX42006	*N. subflava*	carrier	partial			44616
CCUG 7826	*N. subflava*		NPN		22422752	49353
CCUG 24918	*N. subflava*	carrier	NPN		22422752	49362
CA42438	*N. subflava*	carrier	NPN			49571
PL40548	*N. subflava*	carrier	partial			49863
MT40035	*N. subflava*	carrier	partial			50889
PL42224	*N. subflava*	carrier	NPN			50940
9715	*N. wadsworthii*		NPN			21049
CCUG 4007 T	*N. weaveri*		NPN		22422752	49368
10022	*N. weixii*		NPN			56407
NCTC 12230 T	*N. zoodegmatis*		NPN			36317
NBRC 15645	*V. vulnificus*					CP012881.1
AP76	*A. pleuropneumoniae*					CP001091.1
9953L55	*A. porcitonsillarum*				30363939	CP029206.1
130Z	*A. succinogenes*					CP000746.1
NCTC 12996	*A. suis*					LT906456.1
D11S-1	*A. actinomycetem-comitans*				26420795	CP001733.2
USDA-ARS-USMARC-189	*B. trehalosi*					CP006955.1
15-184	*Glaesserella sp.*				29572210	CP023057.1
18010	*H. influenzae*					FQ312006.1
KWG1	*K. kingae*					LN869922.1

Table A.3: Isolates used in Chapter 4. Disease isolates sourced from the MRF-MGL. Carriage isolates sourced from UKMenCar4. cc indicates clonal complex.

Isolate	Epi year	Disease	Capsule	cc	PubMLST ID
11.1489.Z	07/2010-06/2011	invasive	B	cc60	35304
11.5445.E	07/2011-06/2012	invasive	B	cc1157	35322
12.1222.T	07/2011-06/2012	invasive	B	cc60	35331
13.6117.K	07/2013-06/2014	invasive	B	cc1157	35389
13.8092.Q	07/2013-06/2014	invasive	B	cc1157	35396
14-8704599	07/2014-06/2015	invasive	B	cc1157	42619
15-8708102	07/2015-06/2016	invasive	B	cc1157	44680
15-8708164	07/2015-06/2016	invasive	B	cc1157	44683
M10 240618	07/2010-06/2011	invasive	E	cc60	20023
M10 240636	07/2010-06/2011	invasive	B	cc1157	20035
M10 240687	07/2010-06/2011	invasive	B	cc60	20068
M10 240693	07/2010-06/2011	invasive	B	cc60	20071
M10 240718	07/2010-06/2011	invasive	B	cc60	20088
M10 240811	07/2010-06/2011	invasive	B	cc1157	20152
M10 240825	07/2010-06/2011	invasive	B	cc60	20162
M11 240028	07/2010-06/2011	invasive	B	cc60	20190
M11 240117	07/2010-06/2011	invasive	B	cc60	20257
M11 240167	07/2010-06/2011	invasive	B	cc60	20287
M11 240246	07/2010-06/2011	invasive	E	cc60	20328
M11 240338	07/2010-06/2011	invasive	B	cc1157	20390
M11 240339	07/2010-06/2011	invasive	B	cc60	20391
M11 240394	07/2010-06/2011	invasive	B	cc60	20428
M11 240412	07/2010-06/2011	invasive	B	cc60	20441
M11 240440	07/2010-06/2011	invasive	B	cc60	20457
M11 240471	07/2011-06/2012	invasive	B	cc60	21112
M11 241026	07/2011-06/2012	invasive	B	cc1157	21243
M11 241037	07/2011-06/2012	invasive	B	cc1157	21252
M11 241059	07/2011-06/2012	invasive	B	cc1157	21267
M11 241075	07/2011-06/2012	invasive	B	cc60	21280
M12 240093	07/2011-06/2012	invasive	E	cc60	21352
M12 240135	07/2011-06/2012	invasive	B	cc1157	21383
M12 240158	07/2011-06/2012	invasive	B	cc60	21396
M12 240222	07/2011-06/2012	invasive	B	cc60	21432
M12 240641	07/2012-06/2013	invasive	E	cc1157	28107
M13 240153	07/2012-06/2013	invasive	B	cc60	27973
M13 240264	07/2012-06/2013	invasive	B	cc1157	28027
M13 240410	07/2012-06/2013	invasive	E	cc1157	28113
M13 240417	07/2012-06/2013	invasive	B	cc60	28048
M13 240710	07/2013-06/2014	invasive	B	cc60	35511
M14 240123	07/2013-06/2014	invasive	B	cc60	35605
M14 240208	07/2013-06/2014	invasive	E	cc1157	35641
M14 240295	07/2013-06/2014	invasive	B	cc60	35698
M14 240314	07/2013-06/2014	invasive	B	cc60	35710
M14 240437	07/2014-06/2015	invasive	B	cc1157	35776
M14 240554	07/2014-06/2015	invasive	B	cc60	37715

M14 240563	07/2014-06/2015	invasive	B	cc60	37723
M14 240641	07/2014-06/2015	invasive	B	cc1157	37777
M15 240004	07/2014-06/2015	invasive	B	cc1157	37792
M15 240133	07/2014-06/2015	invasive	B	cc60	37896
M15 240716	07/2015-06/2016	invasive	E	cc1157	39354
M15 240789	07/2015-06/2016	invasive	B	cc1157	39405
M15 240984	07/2015-06/2016	invasive	B	cc60	41548
M15 240991	07/2015-06/2016	invasive	B	cc60	41555
M16 240244	07/2015-06/2016	invasive	B	cc1157	44723
M16 240410	07/2015-06/2016	invasive	B	cc60	44791
MT40051	07/2014-06/2015	carrier	B	cc461	50893
MT40144	07/2014-06/2015	carrier	B	cc41/44	50903
MT40165	07/2014-06/2015	carrier	E	cc1157	60961
MT40353	07/2014-06/2015	carrier	B	cc865	50915
MT40393	07/2014-06/2015	carrier	B	cc865	50919
MT40397	07/2014-06/2015	carrier	E	cc1157	60966
BR40013	07/2014-06/2015	carrier	B	Undefined	50242
BR40069	07/2014-06/2015	carrier	E	cc178	43567
BR40076	07/2014-06/2015	carrier	B	cc213	50245
BR40099	07/2014-06/2015	carrier	B	cc213	43569
BR40158	07/2014-06/2015	carrier	B	cc213	50247
BR40196	07/2014-06/2015	carrier	E	cc1157	43575
BR40263	07/2014-06/2015	carrier	B	cc213	43580
BR40286	07/2014-06/2015	carrier	B	cc41/44	50253
BR40329	07/2014-06/2015	carrier	B	cc1157	52541
BR40346A	07/2014-06/2015	carrier	E	cc1157	43588
BR40359	07/2014-06/2015	carrier	E	cc1157	43590
BR40454	07/2014-06/2015	carrier	B	Undefined	43596
BR40458A	07/2014-06/2015	carrier	B	cc41/44	50256
BR40459	07/2014-06/2015	carrier	B	cc41/44	43599
BR40460	07/2014-06/2015	carrier	B	cc41/44	43600
BR40482B	07/2014-06/2015	carrier	E	cc1157	50857
BR40485A	07/2014-06/2015	carrier	E	cc1157	43603
BR40491	07/2014-06/2015	carrier	E	cc1157	50259
BR40497	07/2014-06/2015	carrier	E	cc1157	43732
BR40498	07/2014-06/2015	carrier	B	cc461	43733
BR40503	07/2014-06/2015	carrier	B	Undefined	50262
BR40533	07/2014-06/2015	carrier	E	cc1157	43738
BR40622B	07/2014-06/2015	carrier	E	cc1157	43743
BR40632	07/2014-06/2015	carrier	B	cc213	44112
BR40670	07/2014-06/2015	carrier	B	cc213	43745
BR40675	07/2014-06/2015	carrier	B	cc213	44086
BR40743	07/2014-06/2015	carrier	E	cc1157	43747
BR40748	07/2014-06/2015	carrier	E	Undefined	43748
BR40753	07/2014-06/2015	carrier	B	cc35	43749
BR40767	07/2014-06/2015	carrier	E	cc1157	44116
BR40843	07/2014-06/2015	carrier	B	cc269	43776
BR40912	07/2014-06/2015	carrier	E	cc1157	43777

BR40934	07/2014-06/2015	carrier	B	cc213	52543
BR40999	07/2014-06/2015	carrier	E	cc1157	50268
BR41090	07/2014-06/2015	carrier	E	cc1157	43789
BR41099	07/2014-06/2015	carrier	E	cc1157	50269
BR41123	07/2014-06/2015	carrier	B	cc461	52544
BR41131	07/2014-06/2015	carrier	B	cc32	43631
BR41160	07/2014-06/2015	carrier	B	cc213	52545
BR41173	07/2014-06/2015	carrier	B	Undefined	43551
BR41176	07/2014-06/2015	carrier	E	cc60	43552
BR41208	07/2014-06/2015	carrier	B	cc213	52546
BR41225	07/2014-06/2015	carrier	E	cc1157	43558
BR41252	07/2014-06/2015	carrier	B	cc41/44	50271
BR41257	07/2014-06/2015	carrier	E	cc1157	50272
BR41285	07/2014-06/2015	carrier	B	cc35	43564
BR41292	07/2014-06/2015	carrier	B	cc32	50273
BR41294	07/2014-06/2015	carrier	B	cc35	52548
BR41298	07/2014-06/2015	carrier	B	cc32	43610
BR41315	07/2014-06/2015	carrier	B	cc35	43614
BR41338	07/2014-06/2015	carrier	B	cc35	43615
BR41347	07/2014-06/2015	carrier	B	cc35	43616
BR41378	07/2014-06/2015	carrier	B	cc41/44	52550
BR41407	07/2014-06/2015	carrier	B	cc41/44	43624
BR41423	07/2014-06/2015	carrier	B	cc213	43626
BR41426	07/2014-06/2015	carrier	B	cc35	43627
BR41438	07/2014-06/2015	carrier	B	cc41/44	43634
BR41483	07/2014-06/2015	carrier	E	cc178	43641
BR41486	07/2014-06/2015	carrier	E	cc1157	43642
BR41492	07/2014-06/2015	carrier	B	Undefined	52556
BR41517	07/2014-06/2015	carrier	B	cc32	43649
BR41525	07/2014-06/2015	carrier	E	cc178	43651
BR41532	07/2014-06/2015	carrier	E	cc1157	43652
BR41664	07/2014-06/2015	carrier	B	cc213	50283
BR41694	07/2014-06/2015	carrier	B	cc41/44	43663
BR41778	07/2014-06/2015	carrier	E	cc1157	43671
BR41780	07/2014-06/2015	carrier	B	cc269	43672

B

Manuscript and conference abstracts arising from this DPhil

The 14th Congress of the European Meningococcal and Haemophilus Disease Society, Prague, 2017. Abstract O5.043 (talk)

Capsule genes are common among *Neisseria* commensals.

Marianne Clemence, Martin CJ Maiden and Odile B Harrison.

Introduction

The *Neisseria* are a diverse group of commensal bacteria, which contribute to the richness of the human mucosa microbiota. The pathogenic potential of *Neisseria meningitidis* (*Nme*) is rare, and usually restricted to strains possessing one of the six capsular serogroups that enable invasion of the bloodstream, leading to meningococcal disease. The capsule locus (*cps*) appears to have been acquired by horizontal transfer, conferring pathogenic potential upon *Nme* uniquely within the *Neisseria* genus. Nevertheless, host death ends the transmission chain, so there is presumably no benefit to causing disease. This, the six non-pathogenic seorgroups, and the recent discovery of *cps* loci in other commensal *Neisseria*, imply the capsule may have some other role in the biology of *Nme*, and *Neisseria* in general.

Aims

To further characterise *cps* and its distribution in commensal *Neisseria* species.

Materials and methods

Whole genome sequences from 19 *Neisseria* commensals in pubMLST were surveyed for capsule transport Regions B and C using BLAST. The distribution of cps throughout the *Neisseria* phylogeny was assessed, and new NEIS loci defined for candidate Region A capsule synthesis loci using Artemis. Homology and structure of complete *cps* loci were compared between species using the Artemis Comparison Tool.

Results

The *cps* was common among Neisseria species, including several human-associated commensals, although no cps was identified in isolates from species most closely related to *Nme*, including *Neisseria gonorrhoeae*, *Neisseria polysaccharea*, *Neisseria lactamica* and *Neisseria cinerea*. Homologous capsule synthesis loci were shared between commensals, as well as with *Nme*, though not necessarily between closely related species. *cps* synteny was not conserved throughout the genus, with no commensal *cps* arranged with the same synteny as that of *Nme*; in *Neisseria bacilliformis*, Regions A, B and C were located separately across the genome. *Neisseria subflava* contained isolates both with a complete *cps*, and isolates from which capsule appeared to have been lost, without the *cnl* locus typical of *Nme* capsule-null isolates.

There was also variation in candidate Region A loci within *N. subflava*, and *Neisseria oralis*, with differences between variants comparable to the differences between *Nme* serogroups.

Conclusion

Findings suggest that the capsule was lost in the common ancestor of a subset of *Neisseria*, and later re-acquired in *Nme*. Elsewhere in the genus, *cps* is common and diverse. Therefore, the capsule should not be viewed solely as a virulence factor of the meningococcus, but as a potentially important survival factor across the *Neisseria* genus.

The American Society of Microbiology (ASM) Microbe Conference, Atlanta, GA, 2018. Poster 1074

Genomic analysis of the effects of capsule switching on the epidemiology of *Neisseria meningitidis*.

Marianne Clemence and Martin Maiden.

Neisseria meningitidis (*Nme*) is a commensal of the human nasopharynx, which very occasionally becomes invasive, leading to meningitis and/or septicaemia. Meningococci are classified into clonal complexes (cc), defined by multi-locus sequence typing of seven house-keeping loci. Some ccs are described as hyper-invasive due to their high frequency among disease isolates compared to carriage isolates. Possession of a capsule from one of Serogroups A, B, C, W, X or Y is also usually necessary for disease. However, disease is an evolutionary dead-end for the bacterium, and transmission principally occurs in carriage. This study aimed to investigate the effects of capsule switching within the ST-60 complex (cc60) of *Nme* on incidence in both disease and carriage populations.

The study compared meningococci belonging to the ST-60 complex from a carriage study in British teenagers, and contemporaneous disease isolates within the UK-based MRF Genome Library. Whilst all cc60 isolates from carriage (n=33) possessed a serogroup E capsule, 90% of cc60 isolates associated with disease (n=31) possessed a serogroup B capsule. Results from sequence analysis were consistent with the occurrence of repeated, independent capsule switching from serogroup E to serogroup B within this clonal complex. Using the Random Forest algorithm with alleles as predictors, no other locus from the *Nme* core genome (1605 loci) was identified as being as important as capsule loci, in the switching from Serogroup B to Serogroup E. This single horizontal genetic transfer event appears to confer a higher invasive potential upon these isolates, whilst reducing their ability to colonise individuals, consistent with reduced transmission ability.

cc60 is not usually considered to be hyper-invasive, as it is relatively rare in disease. However, these results suggest it may actually have a high invasive potential if it acquires a disease-associated capsule through horizontal transfer. This acquisition may also inhibit its ability to persist in asymptomatic carriage. This emphasises the importance of combinatorial genetic factors on determining the overall transmission and invasive potential of a bacterial pathogen. These results may also have important implications for the impact of a protein-based "MenB" vaccine on meningococcal carriage and disease.

The 15th Congress of the European Meningococcal and Haemophilus Disease Society, Lisbon, 2019. Abstract PO-044 (poster)

The menignococcal capsule locus contains sequences acquired by horizontal genetic transfer from *Neisseria subflava*, and other *Neisseria* species.

Marianne Clemence, Odile Harrison, Jenny MacLennan and Martin Maiden

Introduction & Aims

Possession of a meningococcal capsule from one of serogroups A, B, C, W, X or Y is virtually essential for *Neisseria meningitidis* to cause invasive meningococcal disease. It has been proposed that the capsule locus may have been acquired in a formerly capsule null isolate by horizontal genetic transfer (HGT) from another species, conferring pathogenic potential upon the bacterium. The aim of this study was to further characterise the evolutionary history of the meningococcal capsule, and determine whether phylogenetic analyses of sequences within the capsule locus are consistent with an HGT acquisition model.

Materials and Methods

Whole genome sequencing data were obtained from pubmlst.org/neisseria, hosted on BIGSdb, and GenBank, making use of isolates from the 107 global dataset, UKMenCar4, the MRF genome library and others. Isolates included meningococci from a broad range of serogroups and clonal complexes, as well as other *Neisseria*, and species from other genera. Isolates were investigated using phylogenetic analyses including the Neighbor-Net algorithm, BOOTSCANning within Recombination Detection Programme 4, and maximum likelihood.

Results

In splits graphs generated using the Neighbor-Net algorithm, meningococcal Region B sequences grouped with the most weight with putative homologues in *N. subflava*, whilst Region C grouped with isolate 10022 (CP023429.1) belonging to the novel species "*N. weixii*". However, splits graphs were highly reticulate, and BOOTSCANning of the sequences of interest was consistent with Region B from some meningococci, and Region C from all meningococci, being mosaic.

Sequences for which BOOTSCANning best supported *N. subflava* as the nearest neighbour grouping were further analysed using maximum likelihood phylogenies; meningococcal sequences were found to be nested within homologous *N. subflava* sequences, consistent with horizontal genetic transfer of sequences from *N. subflava* to *N. meningitidis*.

Region D, and its duplicate Region D' in the meningococcal capsule locus, were also investigated using the Neighbor-Net algorithm and BOOTSCANning. In many cases, analyses were consistent with acquisition of *rfbBAC'+galE1* by HGT from other *Neisseria* species, as opposed to duplication of *rfbBAC+galE* native to capsule null meningococci, although analyses were impeded by re-orientation events that take place within the capsule locus.

Conclusion

The evolutionary history of the meningococcal capsule is complex, with evidence of HGT events involving *N. subflava*, as well as other *Neisseria* species. This has resulted in highly mosaic sequences within the capsule locus. Analyses were consistent with, but not proof of, *en bloc* acquisition of capsule in meningococci, for which *N. subflava* could be a plausible donor.

Microbial Genomics 4(9):e000208, 2018 (manuscript)

MICROBIAL GENOMICS

RESEARCH ARTICLE
Clemence *et al.*, *Microbial Genomics* 2018;4
DOI 10.1099/mgen.0.000208

Characterization of capsule genes in non-pathogenic *Neisseria* species

Marianne Elizabeth Alexandra Clemence, Martin Christopher James Maiden* and Odile Barbara Harrison

Abstract

The genus *Neisseria* comprises a diverse group of commensal bacteria, which typically colonize the mucosal surfaces of humans and other animals. *Neisseria meningitidis*, the meningococcus, is notable for its potential to cause invasive meningococcal disease (IMD) in humans; however, IMD is comparatively rare, and meningococci normally colonize the nasopharynx asymptomatically. Possession of a polysaccharide capsule has been shown to be a prerequisite for disease in almost all IMD cases, and was previously considered unique to *N. meningitidis*, and potentially acquired by horizontal genetic transfer (HGT). Nevertheless, the capsule must also have some role in asymptomatic colonization and/or transmission, consistent with the existence of six non-disease-associated meningococcal capsule serogroups. In this study, full complements of putative capsule genes were identified in non-pathogenic *Neisseria* species, including *Neisseria subflava* and *Neisseria elongata*. These species contained genes for capsule transport and translocation homologous to those of *N. meningitidis*, as well as novel putative capsule synthesis genes. Phylogenetic analyses were consistent with the proposal that these genes were acquired by the meningococcus through HGT. In contrast with previous evolutionary models, however, the most parsimonious explanation of these data was that capsule transport genes had been lost in the common ancestor of the meningococcus, gonococcus, and their close relatives, and then reacquired by some meningococci. The most likely donor of the meningococcal transport genes was another *Neisseria* species.

DATA SUMMARY

This work made use of sequencing data obtained from the pubMLST database (https://pubmlst.org/neisseria), the NCTC 3000 project (https://www.sanger.ac.uk/resources/downloads/bacteria/nctc/) and GenBank (https://www.ncbi.nlm.nih.gov/genbank/index.html). A comprehensive list of strain IDs, metadata and accession numbers can be found on FigShare at https://doi.org/10.6084/m9.figshare.6016112.v1. Annotated pubMLST isolates can be searched in the pubMLST isolate database, and allele sequences retrieved using the NEIS numbers from the pubMLST sequence database.

INTRODUCTION

The genus *Neisseria* is a diverse group of Gram-negative bacteria, many of which are asymptomatic colonizers of the mucosal surfaces of animals and man [1]. In humans, they have been isolated from the mouth, nose, throat and urogenital tract, but whilst many *Neisseria* species belong to the human oral microbiota, research has focused on those associated with disease: *Neisseria gonorrhoeae* and *Neisseria meningitidis*. In common with many other *Neisseria* species, *N. meningitidis* usually colonizes the nasopharynx asymptomatically; however, it occasionally invades the bloodstream, leading to life-threatening invasive meningococcal disease (IMD), comprising meningitis and/or septicaemia [2]. In contrast, there are very few case reports of other *Neisseria* species causing invasive disease. Compared to asymptomatic colonization, IMD is an extremely rare transmission-terminating event, associated with particular meningococcal genotypes that normally express a polysaccharide capsule [3].

Capsules are associated with virulence in several human pathogens, including *Escherichia coli*, *Haemophilus influenzae* and *Klebsiella pneumoniae* [4–6]. A number of successful vaccines have been developed that target capsular antigens, for example the polysaccharides forming the capsules of the meningococcal serogroups A, C, W and Y [7]. Capsules can aid evasion of immune responses, including the complement system and phagocytosis by macrophages, facilitating

Received 22 March 2018; Accepted 15 July 2018
Author affiliation: Department of Zoology, University of Oxford, Oxford, UK.
***Correspondence:** Martin Christopher James Maiden, martin.maiden@zoo.ox.ac.uk
Keywords: meningitis; colonization; virulence determinants; transmission; invasive.
Abbreviations: HGT, horizontal genetic transfer; IMD, invasive meningococcal disease; ML, maximum likelihood; NCBI, National Center for Biotechnology Information ; NPN, non-pathogenic *Neisseria*; rMLST, ribosomal multi-locus sequence typing; WGS, whole-genome sequence.
One supplementary table is available with the online version of this article.

Wellcome Open Research 4:99, 2019 (manuscript)

Wellcome Open Research Wellcome Open Research 2019, 4:99 Last updated: 02 AUG 2019

Check for updates

RESEARCH ARTICLE

REVISED *Neisseria meningitidis* has acquired sequences within the capsule locus by horizontal genetic transfer [version 2; peer review: 2 approved]

Marianne E. A. Clemence, Odile B. Harrison, Martin C. J. Maiden

Department of Zoology, University of Oxford, Oxford, OX1 3SY, UK

v2 **First published:** 21 Jun 2019, **4**:99 (https://doi.org/10.12688/wellcomeopenres.15333.1)
Latest published: 02 Aug 2019, **4**:99 (https://doi.org/10.12688/wellcomeopenres.15333.2)

Abstract

Background: Expression of a capsule from one of serogroups A, B, C, W, X or Y is usually required for *Neisseria meningitidis* (*Nme*) to cause invasive meningococcal disease. The capsule is encoded by the capsule locus, *cps*, which is proposed to have been acquired by a formerly capsule null organism by horizontal genetic transfer (HGT) from another species. Following identification of putative capsule genes in non-pathogenic *Neisseria* species, this hypothesis is re-examined.
Methods: Whole genome sequence data from *Neisseria* species, including *Nme* genomes from a diverse range of clonal complexes and capsule genogroups, and non-*Neisseria* species, were obtained from PubMLST and GenBank. Sequence alignments of genes from the meningococcal *cps*, and predicted orthologues in other species, were analysed using Neighbor-nets, BOOTSCANing and maximum likelihood phylogenies.
Results: The meningococcal *cps* was highly mosaic within regions B, C and D. A subset of sequences within regions B and C were phylogenetically nested within homologous sequences belonging to *N. subflava*, consistent with HGT event in which *N. subflava* was the donor. In the *cps* of 23/39 isolates, the two copies of region D were highly divergent, with *rfbABC'* sequences being more closely related to predicted orthologues in the proposed species *N. weixii* (GenBank accession number CP023429.1) than the same genes in *Nme* isolates lacking a capsule. There was also evidence of mosaicism in the *rfbABC'* sequences of the remaining 16 isolates, as well as *rfbABC* from many isolates.
Conclusions: Data are consistent with the *en bloc* acquisition of *cps* in meningococci from *N. subflava*, followed by further recombination events with other *Neisseria* species. Nevertheless, the data cannot refute an alternative model, in which native meningococcal capsule existed prior to undergoing HGT with *N. subflava* and other species. Within-genus recombination events may have given rise to the diversity of meningococcal capsule serogroups.

Keywords

Neisseria, meningitis, capsule, recombination, horizontal genetic transfer, subflava

Open Peer Review

Reviewer Status ✓ ✓

	Invited Reviewers	
	1	**2**
REVISED **version 2** published 02 Aug 2019		✓ report
version 1 published 21 Jun 2019	✓ report	✓ report

1 **David S. Stephens**, Emory University, Atlanta, USA
Yih-Ling Tzeng, Emory Univeristy, Atlanta, USA

2 **H. Steven Seifert**, Northwestern University Feinberg School of Medicine, Chicago, USA

Any reports and responses or comments on the article can be found at the end of the article.