The temporal and geographical distribution and diversity of disease-associated *Neisseria meningitidis* genetic types in Europe

Carina Brehony

Hertford College
University of Oxford
Department of Zoology

Michaelmas term 2009

Thesis submitted in partial fulfilment for the Degree of Doctor of Philosophy

ABSTRACT

Meningococcal disease, caused by the bacterium *Neisseria meningitidis,* is an important cause of morbidity and mortality in young children and adolescents worldwide. There are 12 serogroups with most disease due to meningococci expressing one of five capsular polysaccharide antigens corresponding to serogroups A, B, C, Y and W135. In Europe, the majority of disease-causing strains are of serogroups B and C. No comprehensive vaccine is available against the bacterium due to the difficulty in producing serogroup B vaccines. A number of countries, *e.g.* UK and the Republic of Ireland have implemented routine meningococcal conjugate C (MCC) vaccine strategies. Due to the high proportion of disease accounted for by serogroup B in Europe and other developed countries, much research is currently being carried out to unearth vaccine candidates that would be protective and give as wide coverage as possible. Such candidates include the antigens PorA, FetA and factor H-binding protein. Potential drawbacks with antigens such as these which are under immune selection are high degrees of variability and lack of cross-immunity. Determination of the distribution, both geographically and temporally, of antigens and their association with clonal complex can aid in the formulation of novel vaccines and assess their potential coverage across Europe.

Serological typing schemes involving characterisation of the polysaccharide capsule (serogroup) and outer membrane proteins such as PorA (serosubtype) and PorB (serotype) have been used for a number of years with some success. However, drawbacks associated with these methods include insufficient discrimination, limitations in panels of monoclonal antibodies used in the typing procedures and difficulty in comparison of results among labs. Consequently, in recent years genotypic methods such as multi-locus enzyme electrophoresis (MLEE) and subsequently multi-locus sequence typing (MLST) have been developed. These methods measure the variation in slowly evolving housekeeping genes whereas serological methods measure variation in antigens which are under immune pressure and are therefore more diverse. Combination of phenotypic and genotypic typing methods can offer high levels of discrimination. Molecular studies into meningococcal diversity have offered many important insights into its population biology, which have implications for prevention and control of meningococcal disease. These have included the identification of

hyperinvasive lineages and the correlation of genetic type with antigenic type and disease epidemiology.

The EU-MenNet programme was established as a pan-European infrastructure for the research and surveillance of European meningococcal disease. Its aim was to coordinate and disseminate the latest molecular isolate characterisation techniques (MLST) and electronic data transfer *via* the Internet to exploit epidemiological and population genetic studies. Within the EU-MenNet, the European Meningococcal MLST Centre (EMMC) was set up to carry out molecular typing — MLST, PorA and FetA — of European disease isolates from 18 countries over three years 2000, 2001 and 2002. The output of this project will be the largest representative molecular epidemiological study of meningococcal disease in Europe. Assessment of the data produced will give insights into the geographic and temporal distribution and structuring of disease-associated clonal complexes and antigens and their associations. This will give an indication of the meningococcal disease population in Europe and will be invaluable for the current, and ongoing, development and introduction of new meningococcal vaccines.

DECLARATION

The work included in this thesis is my own except for a number of isolates which were typed at the beginning of the EU-MenNet project by Dr. Keith Jolley and some samples which were sequence typed at the respective labs indicated in Table 2.1. Epidemiological information collected separately for EU-MenNet isolates was carried out by the European Meningococcal Epidemiology Centre (EMEC/EU-IBIS). The 107 collection nucleotide sequence data analysed, but not generated by me were freely available via the Internet at http://pubmlst.org/. Some of the work included in this thesis has been published (Brehony *et al.* , 2007; Brehony *et al.* , 2009) and copies of these papers can be found in Appendix 2. No part of this work has been submitted for any other degree.

Carina Brehony

December 2009

ACKNOWLEDGEMENTS

I would like to begin by thanking my supervisor Prof. Martin Maiden for first of all taking me on as part of the group. It has been a great opportunity for me in many, many ways and I have learned a great deal since coming to Oxford from the Emerald Isle! His guidance and advice has been invaluable especially in the difficult last months of writing up. I would also like to thank all the members of the Maiden group for being a very pleasant bunch of people to work with, get advice from and just as importantly go the Turf with ☺. Thanks to Keith for all the help with bioinformatics and the multitude general computing queries. And not forgetting our absent members, my fellow travelling companion Ana and the Tartan Army's finest Roisin, both of whom have become great friends. See you in Argentina! Thanks to all of the European labs who provided the samples, without which this project quite literally would not have been possible. Also, thanks to the sequencing facility girls Lynne, Becky and Amanda. Gratitude to all my friends who have offered support and fun times that have made the difficult times easier to manage. I can't name everyone but would like to especially mention Sinéad, Róisín, Niamh and Sheila. Finally and most importantly, I would like to thank my family, Mam, Aidan and Damien who have been there for me since the year dot. I would like to dedicate this work to my parents, Christine and Colum Brehony. Mam you have done such an amazing job with the three of us and put us first always. I hope you are proud and should take enormous credit. Dad, I know you are looking down on us and hope you are proud of how we have all turned out.

Where the wandering water gushes
From the hills above Glen-Car,
In pools among the rushes
That scarce could bathe a star,
We seek for slumbering trout
And whispering in their ears
Give them unquiet dreams;
Leaning softly out
From ferns that drop their tears
Over the young streams.
Come away, O human child!
To the waters and the wild
With a faery, hand in hand,
For the world's more full of weeping than you can understand.

Extract from *The Stolen Child* from *The Wanderings of Oisín and Other Poems* by W.B. Yeats.

In memory of Colum Brehony

LIST OF FIGURES

Figure 1.1: Scanning electron-micrograph of *Neisseria meningitidis* diplococci.
Figure 1.2: Age-specific carriage of *N. meningitidis* and *N. lactamica*.
Figure 1.3: Transmission cycle of *Neisseria meningitidis.*
Figure 1.4: Average laboratory confirmed cases of all invasive meningococcal disease by age and calendar year, England & Wales 1999–2008.
Figure 1.5: 'The Tumbler test' demonstrating non-blanching purpuric rash due to septicaemia.
Figure 1.6: Age-related disease prevalence and serum bactericidal activity (SBA).
Figure 1.7: Cross-section showing meningococcal capsule, cell membrane and surface structures.
Figure 1.8: Chemical composition of meningococcal capsular polysaccharides.
Figure 1.9: Molecular surface of a porin eyelet from the PorA model.
Figure 1.10 (a) and (b): Proposed topology models for FetA.
Figure 1.11: (a) Ribbon diagram of fHbp. (b) Cartoon of the fHbp-fH complex with.
Figure 1.12: Schematic figure depicting meningococcal serological classification.
Figure 1.13: MLST work flow.
Figure 1.14: The reverse vaccinology approach to vaccine candidate discovery.
Figure 1.15: Models of bacterial population structure based on descent.
Figure 1.16: The formation of mosaic genes.
Figure 1.17: Evidence of recombination – Incongruence.
Figure 1.18: The emergence of clones from the background diverse population of *N. meningitidis*.
Figure 1.19: (a) Emergence of lineages. (b) Schematic representation of the effect of immune selection on diverse pathogen populations
Figure 1.20 (a) and (b): Worldwide distribution of invasive meningococcal serogroups.
Figure 1.21: ST-5 complex pandemic spread.
Figure 1.22: (a) Countries in the African meningitis belt. (b) Fitted model of observed meningitis epidemics based on environmental variables. (c) Trends in epidemic meningococcal meningitis in the Africa 1948–2002.
Figure 1.23: Incidence of invasive meningococcal disease in Europe 2005.

Figure 1.24: Incidence of serogroup C invasive meningococcal disease in Europe by routine vaccine MCC and epidemiological year.
Figure 1.25: Percentage distribution of serogroups causing laboratory-diagnosed confirmed and probable meningococcal disease in Europe 2006.
Figure 1.26: Clonal complexes by serogroup from European disease isolates 1999–2004.
Figure 2.1: Circular representation of the *N. meningitidis* Z2471 genome showing the locations of the seven housekeeping loci used for multilocus sequence typing.
Figure 2.2: Locations of oligonucleotide primers used for the PCR amplification and sequencing of the *porA* gene.
Figure 2.3: Locations of oligonucleotide primers used for the PCR amplification and sequencing of the *fHbp* gene.
Figure 3.1: Snapshot of the European disease isolate dataset using eBURST
Figure 3.2: Rank-abundance plot illustrating ST diversity in the European disease population.
Figure 3.3: Diversity of clonal complexes in terms of ST using Simpson's index of diversity arranged in order of decreasing *D* value.
Figure 3.4: Distribution of clonal complexes in Europe 2000–2002.
Figure 3.5 (a) and (b): Genetic diversity and evenness of each country in terms of clonal complex.
Figure 3.6: Percentage of unique STs only found in the respective country.
Figure 3.7: Clonal complex distribution seen in participating countries over the three year period 2000–2002.
Figure 3.8: Change in index of diversity for each country over the time period 2000–2002 in terms of clonal complex.
Figure 3.9: UPGMA tree constructed in SPLITSTREE using a distance matrix of F_{ST} values of pairwise comparisons of STs in 18 European countries.
Figure 4.1: Rank-abundance plot illustrating PorA diversity in the European disease population.
Figure 4.2 (a) and (b): Genetic diversity and evenness of each country in terms of PorA type.
Figure 4.3: Rank-abundance plot illustrating FetA diversity in the European disease population.

Figure 4.4 (a) and (b): Genetic diversity and evenness of each country in terms of FetA type.
Figure 4.5: Distribution of PorA types in Europe 2000–2002.
Figure 4.6: Distribution of FetA types in Europe 2000–2002.
Figure 4.7: Change in diversity index in terms of PorA type over three years in each European country.
Figure 4.8: Change in diversity index in terms of FetA type over three years in each European country.
Figure 4.9 (a) and (b): Observed and expected PorA VR combinations.
Figure 4.10: PorA/FetA combinations.
Figure 5.1: Breakdown of dataset by serogroup.
Figure 5.2: Prevalence of strain types.
Figure 5.3: Distribution of clonal complexes by age group.
Figure 5.4: Breakdown of serogroup by age group.
Figure 5.5: PorA distribution by age group.
Figure 5.6: FetA distribution by age group.
Figure 5.7: Potential coverage of PorA/FetA recipe in EU-MenNet dataset with number of antigens covered (hits).
Figure 5.8 (a) – (j): Breakdown of each vaccine antigen-associated variant family.
Figure 5.9 (a), (b) and (c): Coverage of VR1 VR2 and FetA variant families in vaccine recipe.
Figure 5.10: Distribution of clonal complexes in Europe restricted to serogroup B.
Figure 5.11: PorA/FetA recipe coverage by age group.
Figure 5.12: Potential coverage of PorA/FetA recipe compared to currently available vaccines NonaMen, ACYW, MCC, New Zealand, Norwegian and Cuban.
Figure 5.13: Potential coverage of PorA/FetA recipe compared to NonaMen recipe in countries which have implemented MCC *versus* those that have not.
Figure 6.1: Aligned unique fHbp peptide sequence variable sites.
Figure 6.2: 75% majority-rule consensus CLONALFRAME radial tree of 107 aligned nucleotide sequences with colour coding according to clonal complex and confidence values in nodes.
Figure 6.3: (a) Neighbour-joining tree constructed from aligned peptide sequences indicating association with clonal complex. (b) Maximum likelihood tree for the 107 *fHbp* gene sequences.

Figure 6.4: (a) SPLITSTREE analysis of all nucleotide sequences corresponding to 27 peptide fHbp types. (b) SPLITSTREE analysis of subfamily A/variant 2/3 (17 types). (c) Subfamily B/variant 1 (10 types).
Figure 6.5: Neighbour-joining tree constructed from aligned peptide sequences including a *Neisseria gonorrhoeae* sequence from GenBank.
Figure 6.6: Distribution of clonal complexes within fHbp peptide types.
Figure 6.7: Subfamily-serogroup association.
Figure 6.8: Association of disease state with subfamily/variant.
Figure 6.9: Representation of fHbp gene recombination events.
Figure 6.10: SNAP analysis of 107 aligned fHbp sequences with average number of non-synonymous and synonymous substitutions along the sequence indicated.
Figure 6.11: OMEGAMAP program output using subfamily A/variant 2 and subfamily B/variant 1 nucleotide sequences.
Figure 6.12: Structure of the fH-fHbp complex with temperature colouring using per-site point estimate of ω for (a) subfamily B/variant 1 sequences, (b) subfamily A/variant 2 and (c) subfamily A/variant 3 sequences.

LIST OF TABLES

Table 1.1: MCC vaccination programmes in the European countries as at September 2005.
Table 2.1: Summary of isolates received from 18 participant countries of EU-MenNet.
Table 2.2: Oligonucleotide primers used for the PCR amplification and sequencing of MLST loci.
Table 2.3: Oligonucleotide primers used for the sequencing of the VRs of the *porA* gene.
Table 2.4: Oligonucleotide primers used for the sequencing of the VR of the *fetA* gene.
Table 2.5: Oligonucleotide primers used for the sequencing of the *fHbp* gene.
Table 3.1: Genetic variation in *N. meningitidis* gene fragments.
Table 3.2: Association of clonal complex and country.
Table 3.3: Change in clonal complex over time (2000–2002).
Table 3.4: Significant clonal complex change from 2000–2002 by country by frequency.
Table 3.5: 'Standardised' Index of association (I^S_A) for each country per year.
Table 3.6: AMOVA for isolate data grouped geographically, performed on nucleotide sequences, allele designations, allelic profiles and STs.
Table 3.7: F_{ST} values of pairwise comparisons of concatenated MLST nucleotide sequences in 18 European countries.
Table 3.8: AMOVA for isolate data grouped temporally (years 2000, 2001, 2002), performed on nucleotide sequences (allele and concatenated alleles), allele designations, allelic profiles and STs.
Table 3.9: Temporal structuring in Europe. F_{ST} values of pairwise comparisons of each year using concatenated MLST locus nucleotide sequences, allele sequences, allele designations, allelic profiles and STs.
Table 3.10: F_{ST} values of pairwise comparisons of each year using concatenated MLST nucleotide sequences in 16 European countries.
Table 4.1: Breakdown of PorA and VRs in EU-MenNet dataset.
Table 4.2: Breakdown of PorA types by frequency and association with country.
Table 4.3: Breakdown of FetA types by frequency and association with country.
Table 4.4: F_{ST} values of pairwise comparisons of PorA types in 18 European countries.

Table 4.5: F_{ST} values of pairwise comparisons of FetA types in 16 European countries.
Table 4.6: Change in PorA type over time (2000–2002).
Table 4.7: Change in FetA type over time (2000–2002).
Table 4.8: Temporal structuring in Europe. F_{ST} values of pairwise comparisons of each year using PorA, VR1 and VR2 types in 18 European countries.
Table 4.9: Temporal structuring in Europe. F_{ST} values of pairwise comparisons of each year using FetA in 16 European countries.
Table 4.10: Temporal structuring in Europe. F_{ST} values of pairwise comparisons of each year using PorA types in 16 European countries.
Table 4.11: Temporal structuring in Europe. F_{ST} values of pairwise comparisons of each year using FetA types in 14 European countries.
Table 4.12: PorA/FetA antigen combinations.
Table 4.13: Breakdown of PorA types by frequency and association with clonal complex.
Table 4.14: Breakdown of FetA types by frequency and association with clonal complex.
Table 5.1: Breakdown of clonal complex by frequency and association with age group.
Table 5.2: Breakdown of clonal complex and association with age group by odds ratio with and without controlling for serogroup.
Table 5.3: Potential coverage of PorA/FetA recipe in EU-MenNet dataset with number of antigens covered.
Table 5.4: Breakdown of associated PorA/FetA vaccine recipe antigens with serogroup and clonal complex.
Table 5.5: Breakdown of coverage of PorA/FetA vaccine by year including number of antigens covered.
Table 5.6: Breakdown of coverage of PorA/FetA vaccine by country EU database including number of antigens covered.
Table 5.7: Breakdown of coverage of PorA/FetA vaccine by serogroup in EU database including number of antigens covered.
Table 5.8: Breakdown of serogroup B by clonal complex.
Table 5.9: Breakdown of coverage of PorA/FetA vaccine by country restricted to serogroup B in EU database including number of antigens covered.
Table 6.1: Study isolates and association with clonal complex, sequence type (ST), year, country, disease, serogroup, peptide, nucleotide allele and subfamily/variant.

Table 6.2: Association of disease/carriage and subfamily/variant.
Table 6.3: Maximum χ^2 test for recombination. Putative recombination sites in *fHbp* locus.

LIST OF ABBREVIATIONS

A	adenine
AMOVA	analysis of molecular variance
bp	base pairs
BSA	bovine serum albumin
BURST	Based Upon Related Sequence Types Analysis
˚C	degrees Celsius
χ^2	chi-squared
C	cytosine
cc	clonal complex
CDC	Centers for Disease Control and prevention
CI	confidence interval
CO_2	carbon dioxide
CRM197	cross reacting material 197
CSF	cerebrospinal fluid
D	Simpson's index of diversity
DAMBE	Data Analysis in Molecular Biology and Evolution
DLV	double locus variant
DNA	deoxyribonucleic acid
dNTPs	deoxy-nucleotide triphosphates
ddNTPs	dideoxy-nucleotide triphosphates
d_N/d_S	non-synonymous to synonymous substitutions
E	evenness index
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme-linked immunosorbent assay
EMEC	European Meningococcal Epidemiology Centre
EMMC	European Meningococcal MLST Centre
ET	electrophoretic type
EU-IBIS	European Union Invasive Bacterial Infections Surveillance
EU-MenNet	European Union Meningococcal Network
FACS	fluorescence-activated cell sorting analysis
FepA	ferric enterobactin protein A
FetA	ferric enterobactin transporter A
FhuA	ferric hydroxamate uptake receptor A
fH	factor H
fHbp	factor H-binding protein
FrbP	iron-regulated protein B
F_{ST}	Wright's *F* statistic (fixation index)
g	grams
g	force of gravity
G	guanine
GIS	geographical information system
GNA	genome-derived Neisserial antigen
HCl	hydrochloride
HPA	Health Protection Agency
H_2O	water
I_A	index of association
Ig	immunoglobulin

IS	insertion element
kb	kilobase pairs
kDa	kilo Dalton
LB	Luria-Bertani
LOS	lipooligosaccharide
LPS	lipopolysaccharide
LP2086	lipoprotein 2086
M	molar
MAb	monoclonal antibody
MCC	meningococcal C conjugate
MEGA	Molecular Evolutionary Genetics Analysis
mg	milligram
$MgCl_2$	magnesium chloride
µg	microgram
µl	microlitre
mg	milligrams
min	minute
ml	millilitre
mM	millimolar
µM	micromolar
ML	maximum likelihood
MLEE	multi-locus enzyme electrophoresis
MLST	multi-locus sequence typing
MLVA	multiple-locus variable-number tandem repeat analysis
MVP	Meningitis Vaccine Project
n/a	not applicable
NaCl	sodium chloride
NadA	Neisserial adhesin A
NaOAc	sodium acetate
NCAM	neural cell adhesion molecule
Nhba	Neisserial heparin-binding antigen
ng	nanogram
NIBSC	National Institute for Biological Standards and Control
nM	nanomolar
nm	nanometre
NJ	neighbour-joining
NST	non-subtypeable
NspA	Neisserial surface protein A
NT	non-typeable
NVI	Nederlands Vaccin Instituut
ω	omega
OMP	outer membrane protein
OMV	outer membrane vesicle
OpaA	meningococcal opacity protein
OpC	meningococcal opacity like protein
OR	odds ratio
ORF	open reading frame
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PDB	Protein Data Bank

PEG	polyethylene glycol
PFGE	pulsed-field gel electrophoresis
pmoles	picomoles
PorA	meningococcal class 1 outer membrane porin protein A
PorB	meningococcal class 2 or 3 outer membrane porin protein B
RAPD	random amplified fragment length polymorphism
RIVM	National Institute of Public Health and the Environment, the Netherlands
Rmp	reduction-modifiable protein
rpm	revolutions per minute
SBA	serum bactericidal antibody
sec	second
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
Σ	sigma (sum of)
SLV	single locus variant
SNAP	Syonymous Non-synonymous Analysis Program
ST	sequence type
STARS	*Sequence Typing Analysis and Retrieval System*
START 2	Sequence Type Analysis and Recombinational Tests 2
sVR	semi-variable region
T	thymine
Taq	*Thermus aquaticus*
TBE	tris-borate-EDTA
Tbp	transferrin-binding protein
Tris	tris-hydroxymethyl-aminomethane
UPGMA	unweighted pair group method with arithmetic mean
USELAT	ultrasound-enhanced latex agglutination test
UV	ultraviolet
V	volts
VR	variable region
v/v	volume by volume
WHO	World Health Organisation
w/v	weight by volume
x	times

TABLE OF CONTENTS

ABSTRACT……………………………………………………………………… i
DECLARATION………………………………………………………………… iii
ACKNOWLEDGMENTS………………………………………………………….. iv
LIST OF FIGURES…………………………………………………………… vi
LIST OF TABLES…………………………………………………………...... x
LIST OF ABBREVIATIONS…………………………………………………… xiii
TABLE OF CONTENTS……………………………………………………...... xvi

CHAPTER 1: INTRODUCTION

1.1 The meningococcus……………………………………………….. 1
1.1.1 Meningococcal carriage…………………………………….. 2
1.2 Meningococcal disease…………………………………………… 4
1.2.1 Clinical disease……………………………………………… 5
1.2.2 Diagnosis……………………………………………………. 7
1.2.3 Treatment…………………………………………………… 7
1.2.4 Immunity……………………………………………………. 8
1.3 Surface structures of the meningococcus………………………….. 10
1.3.1 The capsule………………………………………………….. 11
1.3.2 Pili…………………………………………………………… 13
1.3.3 Lipooligosaccharide………………………………………… 14
1.3.4 The outer membrane proteins………………………………. 14
1.3.4.1 PorA……………………………………………….. 14
1.3.4.2 PorB……………………………………………….... 16
1.3.4.3 Reduction modifiable protein……………………… 16
1.3.4.4 The opacity proteins……………………………….. 16
1.3.4.5 FetA……………………………………………….... 17
1.3.4.6 Factor H-binding protein…………………………... 19
1.4 Characterisation of *Neisseria meningitidis*………………………… 22
1.4.1 Serological typing…………………………………………… 22
1.4.2 Genotyping………………………………………………….. 24

1.4.2.1 Multilocus Sequence Typing (MLST)…………….. 24
1.5 Vaccines……………………………………………………… 27
1.5.1 Capsular polysaccharide vaccines…………………………….. 27
1.5.2 Glycoconjugate vaccines……………………………………… 28
1.5.3 Serogroup B vaccines…………………………………………. 29
1.5.3.1 Outer membrane protein vaccines………………… 30
1.5.3.1.1 Reverse vaccinology………………….. 32
1.5.3.2 Other vaccines…………………………………….. 35
1.6 Population genetics……………………………………………… 36
1.6.1 Models of bacterial population structure……………………… 36
1.6.1.1 *N. meningitidis* – 'an epidemic clone'?................... 40
1.7 Epidemiology of meningococcal disease…………………………. 44
1.7.1 Worldwide epidemiology……………………………………... 44
1.7.1.1 Major disease-associated lineages………………… 46
1.7.2 Africa and the 'Meningitis Belt'……………………………… 49
1.7.3 European epidemiology……………………………………….. 52
1.8 Aims and outline of study………………………………………… 58

CHAPTER 2: MATERIALS & METHODS

4.2 Materials………………………………………………………… 60
4.3 Meningococcal isolates…………………………………………… 60
4.3.1 European meningococcal disease isolates…………………… 60
4.3.2 107 meningococcal isolate collection………………………… 62
4.3.3 GenBank fHbp peptide sequence…………………………...62
4.4 Microbiological methods………………………………………... 62
4.5 Molecular biology techniques……………………………………… 63
4.5.1 DNA preparation…………………………………………… 63
4.5.2 PCR amplification…………………………………………. 63
4.5.3 Agarose gel electrophoresis……………………………….... 69
4.5.4 Purification of PCR products……………………………… 70
4.5.5 Automated nucleotide sequence determination…………… 70
4.5.6 Purification of sequencing reactions……………………… 71
4.6 Data analysis……………………………………………………….. 71
4.6.1 Assembly and analysis of sequence data…………………... 71

4.6.2 fHbp database…………………………………………… 72
4.6.3 Genetic analyses and software programs………………….. 73
4.6.3.1 Analysis of selection………………………………. 73
4.6.3.2 Analysis of recombination……………………….. 74
4.6.3.3 Analysis of population structuring………………… 75
4.6.3.4 Software programs………………………………… 76
4.6.3.4.1 START..……………………………… 76
4.6.3.4.2 MEGA...……………………………… 77
4.6.3.4.3 PHYML……………………………….. 77
4.6.3.4.4 SPLITSTREE……………………………. 78
4.6.3.4.5 ARLEQUIN…………………………….. 78
4.6.3.4.6 CLONALFRAME……………………... 79
4.6.3.4.7 OMEGAMAP………………………..... 79
4.6.3.4.8 SNAP..………………………………….. 80
4.6.3.4.9 R program…………………………... 80
4.6.3.4.10 STATA..……………………………... 80
4.6.3.4.11 eBURST..……………………………. 80
4.6.4 Statistical analyses………………………………………… 81
4.6.4.1 Pearson's Chi-squared and Fisher's exact tests…… 81
4.6.4.2 Odds ratios……………………………………….. 82
4.6.4.3 Logistic regression………………………………… 82
4.6.4.4 Simpson's index of diversity……………………… 83

CHAPTER 3: DISTRIBUTION OF INVASIVE MENINGOCOCCAL GENETIC TYPES IN EUROPE 2000–2002

3.2 Introduction…………………………………………………... 85
3.3 Results………………………………………………………... 87
3.3.3 Analysis of MLST housekeeping genes…………………… 87
3.3.4 Genetic diversity of European disease isolates……………. 87
3.3.5 Distribution of genetic types in Europe…………………… 91
3.3.5.1 Genetic diversity by country………………………. 95
3.3.6 Changes in genetic types over time……………………... 98
3.3.6.1 Changes in genetic diversity over time……………. 104
3.3.7 Spatio-temporal structuring in Europe…………………….. 108

3.3.7.1 Geographic structuring…………………………….. 108
3.3.7.2 Temporal structuring………………………………. 112
3.4 Discussion………………………………………………………. 116

CHAPTER 4: DISTRIBUTION AND STRUCTURING OF POTENTIAL VACCINE ANTIGENS PorA AND FetA IN EUROPE
4.1 Introduction…………………………………………………... 121
4.2 Results………………………………………………………... 123
4.2.1 Genetic diversity of antigens in Europe…………………… 123
4.2.1.1 PorA……………………………………………….. 123
4.2.1.2 FetA………………………………………………... 127
4.2.2 Antigenic structuring in Europe…………………………… 130
4.2.2.1 Geographic structuring……………………………. 130
4.2.2.2 Temporal structuring………………………………. 140
4.2.2.3 Non-overlapping antigenic strain structure…………149
4.2.3 Antigen/clonal complex association………………………. 153
4.3 Discussion……………………………………………………… 156

CHAPTER 5: USING POPULATION BIOLOGY TO INFORM VACCINE DESIGN
5.1 Introduction…………………………………………………….. 160
5.2 Merging EU-MenNet and EMEC data……………………………. 162
5.2.1 Breakdown of combined dataset………………………….. 162
5.2.1.1 Age group clonal complex association…………… 165
5.2.1.2 Age group antigen association……………………. 168
5.3 Rationale for PorA/FetA vaccine…………………………………. 170
5.3.1 Stability of vaccine-antigen associations………………….. 173
5.3.2 Vaccine antigen variant families………………………….. 175
5.3.3 PorA/FetA recipe coverage……………………………….. 180
5.3.3.1 By year…………………………………………….. 180
5.3.3.2 By country…………………………………………. 181
5.3.3.3 By serogroup……………………………………..... 182
5.3.3.4 Restricted to serogroup B……………………….... 182
5.3.3.5 By age group…………………………………….... 185

5.3.4 Comparison with other vaccines.............................. 186
5.4 Discussion... 188

CHAPTER 6: VARIATION OF THE FACTOR H-BINDING PROTEIN
4.2 Introduction.. 195
4.3 Results.. 197
4.3.2 Diversity of fHbp gene and protein.......................... 197
4.3.3 Phylogenetic analysis of fHbp................................. 203
4.3.4 Clonal complex/serogroup association....................... 208
4.3.5 Evidence for recombination.................................... 211
4.3.6 Evidence for selection... 214
4.4 Discussion... 219

CHAPTER 7: CONCLUDING REMARKS AND FUTURE WORK
7.1 Conclusions.. 224
7.2 Future work.. 228

REFERENCES

APPENDICES

CHAPTER 1: INTRODUCTION

1.1 The meningococcus

The bacterium later to become known as *Neisseria meningitidis* was first isolated by Anton Weichselbaum in Vienna in 1887 from the cerebrospinal fluid of six fatal cases of 'spotted fever' (Arkwright, 1907). At the time he named the organism *Diplococcus intracellularis meningitidis*. It later became known as *Neisseria meningitidis* after the German scientist and clinician Albert Neisser who discovered the related organism, *Neisseria gonorrhoeae* (the gonococcus). *N. meningitidis* — the meningococcus — is a Gram-negative, aerobic, diplococcus (paired spherical bacterium) β-proteobacterium and a pathogenic member of the family *Neisseriaceae* and genus *Neisseria* (Figure 1.1). Other members of this genus include many harmless animal and human commensals which inhabit mucosal surfaces, and are classified microbiologically as: *Neisseria lactamica*, *Neisseria cinerea* , *Neisseria sicca* , *Neisseria flava* , *Neisseria subflava, Neisseria perflava, Neisseria flavescens*, *Neisseria polysaccharea, Neisseria mucosa* . *N. meningitidis* and the closely-related pathogen *N. gonorrhoeae* are found exclusively in humans and are important causes of disease, namely, meningitis and septicaemia and gonorrhoea respectively. *N. meningitidis* resides in the epithelial mucosal cells of the nasopharynx while *N. gonorrhoeae* primarily infects the mucosal surfaces of the genito-urinary tract.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.1: Scanning electron-micrograph of *Neisseria meningitidis* diplococci. From www.textbookofbacteriology.net (Kenneth Todar, University of Wisconsin-Madison, Department of Bacteriology).

1.1.1 Meningococcal carriage

Despite a reputation as a grave pathogen, the meningococcus is frequently carried harmlessly in the human nasopharynx. It is found at a prevalence of approximately 10% of the human population in non-epidemic conditions, and can be considered to be part of the normal commensal flora (Broome, 1986; Cartwright *et al.*, 1987; Caugant *et al.*, 1994; Claus *et al.*, 2005). Carriage may persist in the nasopharynx for periods of weeks up to several months. Carriage rates in the population vary with age, being the lowest in infants and young children and rising to a peak in adolescents and young adults (Caugant *et al.*, 1994). This contrasts with the carriage rate of the commensal *N. lactamica* which peaks in young children and infants and declines with age (Figure 1.2). Thus it is speculated that *N. lactamica* provides protective immunity against *N. meningitidis* colonisation (Bennett *et al.*, 2005; Gold *et al.*, 1978) and is therefore being evaluated as a vaccine candidate (Gorringe *et al.*, 2009).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.2: Age-specific carriage of *N. meningitidis* and *N. lactamica*. Adapted from (Cartwright *et al.*, 1987).

The meningococcus has many adaptations that help it survive in the habitat of the human nasopharynx: adhesion molecules known as pili help the bacterium adhere to epithelial cell surfaces; receptors that bind to human transferrin and lactoferrin, which are sources of iron, an essential growth component; production of an IgA protease enzyme which destroys the immunoglobulin which would normally prevent bacterial adherence and colonisation. Risk factors for carriage include cigarette smoking (active and passive) and social behaviour such as attendance of pubs/clubs and intimate kissing (Caugant *et al.*, 1994; Kremastinou *et al.*, 2003; MacLennan *et al.*, 2006; Neal, 2000). Spread of *N. meningitidis* is by prolonged close contact or *via* airborne droplets expelled from the nose and throat of colonised persons. It is even thought that sharing of drinking cups and utensils may be a possible means of transmission since the bacterium has been shown to survive for periods of hours on glass or plastic (Swain, 2006). Transmission rates in the population are generally low, but increase in small enclosed communities such as military bases and in university halls of residence (Andersen *et al.*, 1998; Neal, 2000; Riordan *et al.*, 1998).

In temperate climates, disease incidence shows seasonality being at its highest during the winter months with carriage rates not varying with season (Caugant *et al.*, 1994). In sub-Saharan Africa, disease increases rapidly with the dry and windy (Harmattan) season, but declines during the rainy season. However, carriage rates are relatively constant throughout the year (Blakebrough *et a l.*, 1982; Greenwood *et al.*, 1984; Trotter & Greenwood, 2007).

In most cases, carriage does not lead to disease but is an immunising process producing a systemic protective antibody response (Kremastinou *et al.*, 1999). Less than 1% of those colonised go on to develop systemic infection and in Europe, only one case of disease occurs per 100,000 inhabitants per year (Aycock & Mueller, 1950; EU-IBIS *et al.*, 2007; Vogel, 2001). Meningococcal carriage, and not disease, is now recognised to be the major force in transmission of the bacteria, since most disease cases have not been in contact with other cases and the organism is an obligate human pathogen. Disease is a relatively rare event and not useful to the meningococcus since it is a dead-end situation in that it does not promote transmission (Figure 1.3). Therefore, *N. meningitidis* is considered an 'accidental pathogen' (Maiden & Urwin, 2006).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.3: Transmission cycle of *Neisseria meningitidis.* From (Frosch & Maiden, 2006).

1.2 Meningococcal disease

The Swiss physician Vieusseux is credited with the first description of an outbreak of meningococcal disease which occurred in Geneva and its environs in 1805 (Cartwright, 1995). He tracked the epidemic which resulted in 33 deaths and noted the classical clinical symptoms which included vomiting, skin rashes, stiffness of the spine and severe headaches. Almost contemporaneously was the description of an outbreak in Massachusetts in the United States by Danielson and Mann:

> "*The history of a singular and very mortal disease, which lately made its appearance in Medfield*"

In the majority of instances, meningococcal colonisation of the nasopharynx leads to asymptomatic carriage, which may last for a period of weeks or months and then be cleared. However, in a small number of cases it can lead to invasive disease when the bacterium penetrates the epithelial surface by means of process called 'parasite-directed endocytosis' and enters the bloodstream. The bacteria can proliferate rapidly in the blood and release endotoxin (septicaemia/meningococcaemia) and in some cases cross the blood-brain barrier (Nassif *et al.*, 2002) into the cerebrospinal fluid

(CSF) causing inflammation of the meninges (meningitis). Invasive disease occurs usually within 10 days of acquisition of the bacterium. The exact reasons behind bacterial invasion and survival in the blood are still poorly understood but are likely to be due to a combination of host, environmental and meningococcal factors. Those at highest risk of developing disease include young infants and young adults (Figure 1.4). Close contacts and family of an index case are also at increased risk of disease. There is a slight predominance overall of disease among males (Jones, 1995).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.4: Average laboratory confirmed cases of all invasive meningococcal disease by age and calendar year, England & Wales 1999–2008. Adapted from www.hpa.org.uk.

1.2.1 Clinical disease

Meningococcal infection can result in a number of different clinical manifestations which may occur separately or in combination: fulminant septicaemia, meningitis, fulminant septicaemia *with* meningitis and mild systemic meningococcal infection. Other more rare manifestations of disease include pneumonia, arthritis, epiglottitis, pericarditis and otitis media. More than 50% of people infected with meningococci get

both meningitis and septicaemia, over 30% get septicaemia alone and fewer than 15% get meningitis alone (2007). The clinical symptoms of meningococcal disease in the early stages can be difficult to distinguish from those of more common and milder infections such as influenza. The symptoms of meningitis, the most common presentation, include sudden onset of headache, fever, neck stiffness, nausea, vomiting, photophobia and signs of altered mental function. Septicaemia, which can occur either on its own or with meningitis, results in higher mortality. Its symptoms include rapid onset of fever and a distinctive non-blanching purpuric rash in more than half of cases (Figure 1.5).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.5: 'The Tumbler test' demonstrating non-blanching purpuric rash due to septicaemia. From www.meningitis.org.

Before the advent of the antibiotic era, meningococcal disease was associated with very high mortality rates, typically 70% or more. Now with treatment, the case-fatality rate is approximately 10% (Goldacre *et al.* , 2003; Rosenstein *et al.* , 2001; van Deuren *et al.* , 2000). Serious sequelae such as hearing loss, neurologic and brain damage, sensorineural deafness and limb loss can occur in a small number of disease survivors (Edwards & Baker, 1981; Heckenberg *et al.* , 2008; Kirsch *et al.* , 1996; Schildkamp *et al.* , 1996). Predictors of poor prognosis and severe disease include the following clinical presentations: absence of meningitis, septic shock, low peripheral white blood cell count, deranged coagulation indices and thrombocytopaenia (Stiehm & Damrosch, 1966; van Brakel *et al.*, 2000).

1.2.2 Diagnosis

Early recognition and diagnosis of meningococcal disease are essential as it can progress very rapidly to death if left untreated, particularly in the case of fulminant septicaemia. The gold standard for diagnosis of meningococcal disease is bacterial culture from blood or CSF samples (normally sterile sites), Gram staining and visualisation of the bacterium *via* microscopy. Isolation of the organism can be difficult and false negatives occur especially when there has been prior treatment with antibiotics. Rates of isolation can drop from 50% to less than 5% if parenteral antibiotics have been administered (Cartwright *et al.*, 1992; Ni *et al.*, 1992). Culture of the organism can be performed in most laboratories and advantages of this are that further testing such as antibiotic-susceptibility and characterisation, *e.g.* serogroup, serotype, MLST for epidemiological purposes can be carried out.

More recently, rapid non-culture techniques such as PCR, latex agglutination and serology have been combined with culture to improve diagnostics, especially since use of antibiotics can make isolation and culture of organisms difficult or impossible. PCR is a rapid, highly specific and sensitive technique for the diagnosis of meningococcal DNA in clinical samples (Kristiansen *et a l.*, 1991; Newcombe *et al.* , 1996; Ni *et al.* , 1992). It has been used in the UK by Health Protection Agency reference labs since 1996 and has greatly increased the numbers of confirmed cases (Kaczmarski *et al.*, 1998). A drawback is that it is relatively expensive to perform and is thus not readily available to all labs especially those in developing countries. The ultrasound-enhanced latex agglutination test (USELAT) is a rapid and relatively inexpensive method that detects the presence of bacterial antigens in CSF and blood serum (Gray *et a l.*, 1999; Sobanski *et a l.*, 2001). The enzyme-linked immunosorbent assay (ELISA) is a reproducible and specific test that has been used for a number of years for the detection of serum antibody levels in meningococcal disease cases and can be used in tandem with other diagnostic methods (Gray *et al.*, 2001).

1.2.3 Treatment

Sulphonamides, first developed in 1937, were the first antibacterial agent specifically designed to tackle meningococcal infection and found widespread use during the period

of World War II and after (Schwentker *et al.*, 1937). However, their use has now been largely abandoned as a result of widespread resistance that began to appear in the 1950s and 1960s (Millar *et al.*, 1963). At present, penicillin, which was first developed as a chemotherapeutic agent in the 1940s (Chain *et al.*, 1940), is the preferred drug for treatment. A 5–7 day course of intravenous antibiotic therapy is generally recommended for effective treatment of both meningitis and septicaemia. A more recently developed class of antimicrobials, the cephalosporins, *e.g.* ceftriaxone are also very effective against the meningococcus. Their advantages are that they are effective against other microbes (if *N. meningitidis* has not been confirmed) and, as yet, there is no resistance to them, although penicillin resistance is still a relatively rare occurrence (Vazquez, 2001). As close contacts of disease are at higher risk, antimicrobial chemoprophylaxis is recommended to prevent further cases. During epidemics in the developing world, it is not practical or economically viable to use repeated doses of penicillin or cephalosporins. It has been found that a single dose of a suspension of oily chloramphenicol (Pecoul *et al.*, 1991) is effective. The effectiveness of adjuvant therapies such as the use of steroids to reduce the effects neuronal damage in meningitis cases is still under debate (McIntyre, 2005).

1.2.4 Immunity

A large proportion (>50%) of new-borns have short-term passive immunity to the meningococcus as a result of acquisition of maternal antibodies through the placenta and in the colostrum through breast-feeding (Goldschneider *et al.*, 1969b; Griffiss, 1995). Serum bactericidal activity (SBA) is a measure of immune response and is considered a correlate of protection against meningococcal disease and is therefore used in vaccine evaluation (Borrow *et al.*, 2005). Anti-meningococcal SBA has been found to peak in neonates, when disease is uncommon, and then wane rapidly after 6–12 months when it is at its lowest level, when cases of meningococcal disease peak (Figure 1.6). SBA rises steadily again through childhood and it reaches a point where 65–85% of adults have SBA against the meningococcus (Pollard & Frasch, 2001). The rise in SBA with age after 12 months to adulthood reflects the increase in active immunity developed upon exposure to *N. meningitidis* through colonisation and by cross reaction

against antigens of other related harmless commensals, *e.g. N. lactamica* (Gold *et al.* , 1978).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.6: Age-related disease prevalence and serum bactericidal activity (SBA). From adaptation by (Pollard & Frasch, 2001) from (Goldschneider *et al.*, 1969b).

Factors which damage the mucosa such as smoking, exposure to dry wind, *e.g.* Harmattan in Africa, and respiratory infections such as influenza increase risk of invasive disease (Cartwright *et al.* , 1991; El Ahmer *et al.* , 1999). Individuals with immune deficiencies such as hypogammaglobulinaemia, which is a deficiency of plasma gamma globulins and impairment of antibody formation, are at greater risk of meningococcal infection and disease, underlining the importance of antibodies in the immune response (Salit, 1981). Other groups with increased risk of disease include individuals with deficiencies in the complement pathway which is one of the first lines of defence against bacterial infection. For example, a polymorphism that causes an increase in the levels of the negative complement regulator factor H is associated with lower levels of bactericidal activity and therefore increased susceptibility to meningococcal disease (Haralambous *et al.*, 2006).

1.3 Surface structures of the meningococcus

Like other Gram-negative organisms, the meningococcus has a cell envelope which arranges itself into a triple layer of: a cytoplasmic membrane, periplasmic space and peptidoglycan layer, and an outer membrane which contains a myriad of surface structures such as pili and proteins (Figure 1.7). It is also able to produce a polysaccharide layer, also known as a capsule, on top of the outer membrane. These outermost parts of the bacterial cell, the outer membrane and capsule, are very important in the interaction of the bacterium and host immune system.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.7: Cross-section showing meningococcal capsule, cell membrane and surface structures. From (Morley & Pollard, 2001).

1.3.1 The capsule

Unlike other members of the genus *Neisseria,* the meningococcus is able to produce a polysaccharide capsule. The capsule is crucial for invasive disease as it aids survival in the nasopharynx by offering protection against the human immune response, *e.g.* bacteriolysis and phagocytosis. Furthermore, as water is an important constituent of the capsule it prevents desiccation in the external environment thus aiding the transmission process. The meningococcal polysaccharide ("Type 1 specific substance") was first identified by Scherp and Rake as an important component of the bacterial cell (Scherp & Rake, 1935).

The polysaccharide capsule is chemically and antigenically diverse, and this diversity is exploited in serological characterisation to define serogroups. There are 13 different serogroups, with most invasive disease caused by meningococci producing the following five types: A, B, C, W-135, Y. More recently however, serogroup X has emerged and has been responsible for an increased number of disease cases in Africa (Boisier *et al*., 2007; Rosenstein *et al*., 2001; Stephens, 2007). All the major invasive serogroups are composed of sialic acid derivatives (5-*N*-acetyl-neuramninic acid), except for A and X which are composed of *N*-acetyl mannosamine-1-phosphate and *N*-acetyl-D-glucosamine-1-phosphate respectively (Frosch & Vogel, 2006) (Figure 1.8). The α(2-8) linked sialic acid serogroup B homopolymer is chemically and immunologically identical to a molecule found on the human Neural Cell Adhesion Molecule (NCAM) (Finne *et al*., 1983; Finne *et al.*, 1987). This is thought to account for the poor mammalian immune response against the B capsule.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.8: Chemical composition of meningococcal capsular polysaccharides. From (Morse, 1996).

The 24 kilobase chromosomal gene cluster *cps* contains all the genes involved in meningococcal capsular synthesis and consists of five regions: A, B, C, D and E (Frosch *et al.*, 1989; Frosch & Vogel, 2006). Region A contains the genes required to encode enzymes involved in polysaccharide biosynthesis: the *siaD* gene required for polysialyltransferase, which has variants corresponding to the different sialic acid containing serogroups (B, C, Y and W-135), *siaA*, *siaB* and *siaC* required for synthesis of activated sialic acid (Claus *et al.* , 1997) and the *myn* gene operon required for serogroup A capsule expression (Swartley *et al.*, 1998). Regions B and C encode genes involved in the translocation of the polysaccharide to the cell surface. Region D contains genes involved in lipid modification (Hammerschmidt *et al.*, 1994) and region E contains a gene encoding the putative transcription factor *tex*. These two regions are

also present in *N. gonorrhoeae* and *N. lactamica* and hence may not solely be related to capsule biosynthesis (Claus *et al.*, 2002; Petering *et al.*, 1996). Therefore, it is postulated that encapsulation is due horizontal DNA transfer of genes from unrelated species to the meningococcus (Petering *et al.*, 1996).

A feature of the meningococcus is its ability to vary expression and switch on/off surface molecules including the capsule. This is known as 'phase variation'. In particular situations it is advantageous for the bacteria to possess a capsule, *e.g.* during transmission for protection, while conversely, the acapsulate state allows for adhesion and then invasion into mucosal epithelial cells (Hammerschmidt *et al.*, 1996). Mechanisms by which capsules may be inactivated include slipped-strand mispairing and mobile element insertions into region A (Hammerschmidt *et al.*, 1996; Weber *et al.*, 2006). Meningococci which completely lack the regions A, B and C of the *cps* locus and therefore are unable to produce a capsule have been found in carriers (Claus *et al.*, 2002). These acapsulate isolates contained a region 133–144 bp in length known as the capsule null locus (*cnl*). Capsule switching, from one capsular polysaccharide type to another, is another feature in the arsenal of the meningococcus. Isolates have been shown to switch from one serogroup to another probably as the result of horizontal gene transfer of *siaD* genes (Alcala *et al.*, 2002; Stefanelli *et al.*, 2003; Swartley *et al.*, 1997).

1.3.1 Pili

In common with other Gram-negative bacteria, such as the gonococcus, *N. meningitidis* produces type IV pili which are filamentous polymer projections 6 nm in diameter extending to several micrometres in length (Merz & So, 2000). The pilus is made up of pilin subunits which are encoded for by the *pilE* gene (Morand & Rudel, 2006). There are two different classes of pilin, class I and class II, which are structurally and antigenically distinct (Heckels, 1989). Pili have many functions important in the pathophysiology of meningococcal infection including, adhesion to host cells, bacterial movement, bacterial aggregation, DNA transformation and conjugation. They are not considered feasible as vaccine candidates due to their very high antigenic variability (Frasch, 1995).

1.3.2 Lipooligosaccharide

Lipooligosaccharide (LOS) is a significant component (~50%) of the meningococcal cell membrane and is related to lipopolysaccharide (LPS) found in many Gram-negative bacteria. It has a molecular mass of approximately 4.8 kDa and is made up of two regions, a core hydrophilic oligosaccharide and hydrophobic lipid A region embedded in the outer membrane (Wright *et a l.*, 2006). The lipid A (meningococcal endotoxin) element of LOS which is a constituent of outer membrane vesicles (OMVs) or 'blebs' — characteristically shed by the meningococcus — plays a part in the induction of proinflammatory responses in meningococcal sepsis and meningitis. There are at least 12 distinct structural variations known as immunotypes, L1–L12 (Scholten *et al.*, 1994), which can be detected by various immunoassays such as ELISA (Kuipers *et al.*, 2001). L3, L7 and L9 are the immunotypes most commonly associated with disease.

1.3.3 The outer membrane proteins

There are five major structural classes of meningococcal outer membrane proteins (OMPs) as elucidated by SDS-PAGE (Hitchcock, 1989; Tsai *et al.*, 1981). These are: the Class 1 porin protein PorA; Class 2 and 3, mutually exclusive, PorB porins encoded for by the same *porB* gene; Class 4 Rmp; Class 5 Opa/Opc proteins (Hitchcock, 1989). Other important proteins associated with the outer membrane include iron-regulated proteins such as FetA and the more recently identified factor H-binding protein.

1.3.3.1 PorA

PorA, which is encoded by the *porA* gene, is one of the most abundant and immunogenic proteins found in the outer membrane. It is found in the majority of meningococcal isolates although there can be varying levels of expression depending on the presence of a polyguanidine stretch in the promoter region (van der Ende *et al.* , 1995). According to the presence of 11, 10 or 9 contiguous guanidine residues, expression levels will be high, medium or absent respectively. Isolates with deletions of the *porA* gene and with the gene rendered inactive by insertion elements have also been observed (Newcombe *et al.* , 1998; van der Ende *et al.* , 1999). The protein has a

molecular mass of 44–47 kDa (Derrick *et al.* , 2006) and forms protein trimers which function as cation selective pores allowing hydrophilic nutrients pass into the cell which would otherwise be impenetrable (Tommassen *et al.* , 1990). The structure of the protein is that of a β barrel with eight surface-exposed loops (Figure 1.9). Two of these loops (I and IV) contain variable regions (VR1 and VR2) (McGuinness *et a l.*, 1990) which contain immunologically important epitopes. There is also a minor semi-variable region (sVR, VR3). Indeed, the antigenic variation in these VRs is exploited by phenotypic (serosubtyping) and genotypic typing methods as one means of differentiating between isolates.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.9: Molecular surface of a porin eyelet from the PorA model. The external loops (red) are indicated by numbers I to VIII. The β sheets are indicated by the blue directional arrows. From (Derrick *et al.*, 1999).

Due to its high immunogenicity and despite high variability, PorA is attractive as a vaccine candidate, and numerous studies and trials have been carried out (Oster *et al.*, 2005; van den Dobbelsteen *et al.*, 2007). Tailor-made OMV vaccines containing the respective outbreak strain PorA proteins have been used with some success in tackling outbreak situations in Norway, Cuba and New Zealand (Bjune *et al.*, 1991; O'Hallahan *et al.*, 2005; Sierra *et al.*, 1991).

1.3.4.2 PorB

Like PorA, PorB is an abundant porin protein in the meningococcal outer membrane. It also assembles into trimeric structures that form an anionic selective pore allowing diffusion of nutrients into the cell. Meningococci express either one or other of its two protein forms encoded for by the *porB* gene, *i.e.* Class 2 (PorB2) or Class 3 (PorB3) which are 40–42 kDa and 37–39 kDa respectively (Derrick *et al.*, 2006). It forms a β barrel with eight surface-exposed loops with four loops showing high levels of amino acid variability (loops I, V, VI and VII) (Sacchi *et al.* , 1998). A major difference between the two porins is that loops I and IV are shorter in PorB than those in PorA and this has been suggested as a possible reason for a lack of anti-PorB3 bactericidal response in immunised individuals (Guttormsen *et al.*, 1993; Michaelsen *et al.*, 2001).

1.3.4.3 Reduction modifiable protein (Rmp)

The class 4 OMP reduction modifiable protein (Rmp, also known as RmpM) is so named because it undergoes modification of its electrophoretic mobility in the presence of a reducing agent. Its structure and function have yet to be fully elucidated but it is thought to be closely associated and form complexes with the porins (PorA, PorB) (Jansen *et a l.*, 2000) and the iron-regulated OMPs lactoferrin-binding protein (LbpA), transferring-binding protein A (TbpA) and ferric enterobactin transporter A (FetA) (Prinz & Tommassen, 2000).

1.3.4.4 The opacity proteins

The class 5 OMP opacity proteins (Opa, Opc) are so named because they are associated with opaque agar-grown colony phenotypes. There are high levels of sequence diversity in terms of *opa* genes and Opa proteins found in meningococcal isolates (Callaghan *et al.*, 2006) and they undergo high levels of antigenic and phase variation. Meningococci express three or four of these highly variable Opa protein types which are encoded for by a family of genes which are found throughout the genome: *opaA*, *opaB*, *opaD* and *opaJ*. Opc on the other hand is not as diverse antigenically and is encoded for by just one gene (*opc*) (Olyhoek *et al.* , 1991). Opacity proteins are involved in key processes in host cell-bacterium interaction including, assisting bacterial adhesion to the

host cell epithelium and invasion of the bacterium into epithelial cells (Dehio *et al.*, 1998; Virji *et al.*, 1993). Furthermore, high variability and phase and antigenic variation of these proteins assist the bacterium in evasion of the host immune system.

1.3.4.5 FetA

FetA, which was previously known as FrpB (iron-regulated protein B) (Ala'Aldeen *et al.*, 1994; Pettersson *et al.*, 1995), is a 70 kDa iron-regulated OMP coded for by the *fetA* gene. It is expressed in the majority of meningococci and deletions of the gene are rare but can occur in invasive isolates (Claus *et al.*, 2007; Marsh *et al.*, 2007). It is expressed under iron-limited conditions in the human host and shows sequence homology to TonB-dependent OMPs of other Gram-negative bacteria (Kortekaas *et al.*, 2007). Its function is in transportation of iron into the bacterial cell. Following the general structural criteria for OMPs of other Gram-negative bacteria, a topology model was proposed that predicts a β-sheet formation with 26 membrane-spanning domains and 13 surface-exposed loops (Figure 1.10 (a)) (Pettersson *et al.*, 1995).

However, a more recent study suggests that this topology may need to be revised following the elucidation of the crystal structures of five other TonB-dependent receptors, *e.g.* FepA and FhuA in *Escherichia coli* which showed a 22-strand β-barrel structure and 11 surface-exposed loops (Figure 1.10 (b)) (Kortekaas *et al.*, 2007). In the first topology model, loop 7 is the longest of the 13 loops, while in the second topology model it corresponds to loop 5 of the 11 loops in total. The extended loop (7/5) contains a variable region (VR) containing epitopes that are targets for many MAbs raised against the protein (van der Ley *et al.*, 1996). Genotypic typing, similar to PorA VR typing, targets this VR to differentiate between isolates and has been used to assess levels of antigenic diversity.

FetA has been proposed as a vaccine candidate for a number of reasons: antibodies against FetA have been found in the sera of patients recovering from meningococcal disease (Black *et al.*, 1986); murine antibodies raised against FetA were found to be bactericidal and strain-specific (Pettersson *et al.*, 1990); FetA can elicit an immune response in vaccinees (Wedege *et al.*, 1998). Studies have shown a highly diverse but structured distribution of FetA types in the meningococcal population (Russell *et al.*, 2008; Thompson *et al.*, 2003; Urwin *et al.*, 2004).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.10 (a): Proposed topology model for FetA. The protein consists of a β barrel with 13 surface-exposed loops. Boxed areas denote regions of variability and the shaded section corresponds to epitopes on loop 7 recognised by monoclonal antibodies. From (Pettersson *et al.*, 1995).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.10 (b): Proposed new topology model for FetA. The FetA β-barrel consists of 11 surface-exposed loops labelled L1–L11. Amino acids in the exposed loops and periplasmic turns are shown as circles while those in the β-barrel are diamonds (grey diamonds indicate residues exposed to the lipid bi-layer). Black circles indicate amino acid sequences against which anti-sera were raised in the experiment. The length of β strands was arbitrarily chosen to be identical to those in FepA (*E. coli* receptor). From (Kortekaas *et al.*, 2007).

1.3.4.6 Factor H-binding protein

Factor H-binding protein (fHbp), previously known as lipoprotein 2086 (LP2086) and Genome Derived Neisserial Antigen 1870 (GNA1870), was discovered by an iterative process of immunisation following differential detergent extraction and protein purification (Bernfield *et al.*, 2002; Fletcher *et al.*, 2004). It was also identified as GNA1870 (Masignani *et al.*, 2003) by the technique known as 'reverse vaccinology' (Rappuoli, 2000) – refer to section 1.5.3.1.1. Subsequently, it has been given the name factor H-binding protein because of its role in modulating the activity of the alternative complement pathway where it binds the regulatory protein factor H (fH) (Madico *et al.*, 2006).

fH has a critical role in maintaining homeostasis of the complement system and also, by attachment to host cells and tissue, preventing potential damage to them by inhibiting complement activation (Rodríguez de Córdoba *et al.*, 2004). Several organisms, including *N. meningitidis* mimic human tissue by recruiting fH and coating their surface, therefore avoiding complement-mediated lysis (Lambris *et al.*, 2008; Schneider *et al.*, 2006). In the case of the meningococcus, fHbp is the only receptor for fH on its surface (Schneider *et al.*, 2009). The protein is present in all meningococci so far tested although levels of expression may vary (Fletcher *et al.*, 2004; Masignani *et al.*, 2003). In comparison to other vaccine antigens, it is relatively sparse in its epitope surface-exposure in most meningococcal strains (Welsch *et al.*, 2004). Expression of fHbp has been found to be key for survival in *ex vivo* human blood and human serum particularly in high-expressing strains (Seib *et al.*, 2009; Welsch *et al.*, 2008). Structurally, fHbp is a surface-exposed 29 kDa globular lipoprotein composed of two β barrels connected by a short linker and is bound to the outer membrane by an N-terminal lipid anchor (Figure 1.11) (Cantini *et al.*, 2009; Mascioni *et al.*, 2008; Schneider *et al.*, 2009). Recent analysis of the fH-fHbp interaction indicate that the fH recognition site spans the whole surface of fHbp and that previously described bactericidal epitope sites do not lie in this region, but epitopes that bind to antibodies that affect fH binding are found around the edge of the site (Schneider *et al.*, 2009).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.11: (a) Ribbon diagram of fHbp. Secondary structure elements are shown. β-strands of the N-terminal domain are shown in *cyan* and helices are shown in *red*, whereas β-strands of fHbp C terminal domain are shown in *dark blue* (Cantini *et al.* , 2009). (b) Cartoon of the fHbp-fH complex with fHbp coloured *cyan, green* and *yellow* and fH in *dark blue*. Side chains from both proteins involved in forming salt bridges across the interaction surface are shown in *red* as ball-and-stick representations (zoomed and reoriented in inset box) (Schneider *et al.*, 2009).

fHbp is a principal component of two recombinant protein vaccines in clinical trials at the time of writing (Anderson *et al.*, 2009; Biolchi *et al.*, 2009). It is unique as a vaccine candidate in that it is able to elicit serum antibodies that activate classical complement pathway bacteriolysis and also prevent fH binding to the meningococcal cell surface thus making it more susceptible to bactericidal activity (Madico *et al.*, 2006; Welsch *et al.*, 2008). Like the related human-restricted organism *N. gonorrhoeae*, there is specificity of binding to human fH (Granoff *et al.*, 2009; Ngampasutadol *et al.*, 2008). This may help to explain the higher bactericidal titres obtained when using vaccine-induced antibodies with rabbit complement *versus* human complement and also the organisms' exclusively human-related pathogenicity.

1.4 Characterisation of *Neisseria meningitidis*

A multitude of typing techniques, both phenotypic and genotypic, exist for bacterial characterisation (Achtman, 1996; van Belkum *et al.*, 2007) and many of these methods have contributed greatly to the understanding of the epidemiology and population biology of a diverse array of organisms. The choice of typing method depends on the question being asked and the level of discrimination required.

1.4.1 Serological typing

After the microbiological identification of a bacterium such as *N. meningitidis*, for epidemiological purposes it is necessary to carry out classification of the isolate to obtain as much information as possible. One way of doing this is by means of serological (phenotypic) characterisation which identifies the types of surface antigens that are present. The serological classification scheme of *N. meningitidis* comprises: serogroup, based on capsular polysaccharide variants; serotype, based on variants of the class 2 or 3 OMP PorB; serosubtype, based on variants of the class 1 OMP PorA; immunotypes, based on variants of LOS (Figure 1.12). In this typing scheme an isolate is given a designation whereby each type is separated by a colon, *e.g.* A:4,21:P13,6:L3.

Serological differences between meningococci in terms of serum agglutination were observed as long ago as 1909 by Dopter. Originally a four-type classification system (I–IV) as defined by Gordon and Murray in 1915 was used by the majority of laboratories until a more clear and unified scheme was introduced which initiated the use of the serogroup names — A, B, C *etc.*— that are used today (Branham, 1953). Subsequent additional serogroups were added over the years (Eldridge *et al.*, 1978) so that today 13 meningococcal serogroups are known. Using serum or cerebrospinal fluid, serogrouping is performed by agglutination using polyclonal or monoclonal antibodies.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.12: Schematic figure depicting meningococcal serological classification. Serogroup is determined by the capsular polysaccharide type (A, B, C, *etc.*); serotype is determined by the PorB type (2a, 2b, *etc.*); serosubtype is determined by PorA type (P1.5,2, *etc.*); immnunotype is determined by the lipooligosaccharide type (L1–L12). Adapted from: (Maiden & Feavers, 1995).

The presence of various antigenic types within serogroups was identified in early studies using bactericidal assays (Frasch & Chapman, 1972; Gold & Wyle, 1970; Goldschneider *et al.*, 1969a). These 'serotype' antigens were later found to consist of OMPs and LOS (Mandrell & Zollinger, 1977; Poolman *et al.*, 1980). In 1985, a unified serotyping nomenclature system was introduced (Frasch *et al.*, 1985) and has been widely adopted since (Kuipers *et al.*, 2001) with extra refinement of serosubtyping over the years (Poolman & Abdillahi, 1988). Various immunoassays such as ELISA, dot blot, colony blot and immunoblot are employed in serological classification. There are several limitations associated with serological characterisation methods namely: current panels of MAbs may not give complete coverage of types present; absence of epitopes due to phase variation or mutation; genetic relationships between isolates missed; results may sometimes be ambiguous. In an effort to override these problems and in addition to serological typing, genotypic typing methods have been developed more recently with much success.

1.4.2 Genotyping

Various DNA-based typing techniques have been devised as alternatives to and complementary to typing methods employing serology. These include: pulsed-field gel electrophoresis (PFGE) (Ribot *et al.*, 2001), ribotyping (Woods *et al.*, 1992), restriction fragment length polymorphism (RFLP) (Campos *et al.*, 1992) and multiple-locus variable number tandem repeat analysis (MLVA) (Schouls *et al.*, 2006). While these methods have proved to be discriminatory and useful for short-term epidemiology they have some drawbacks such as poor portability and the inability to determine long-term relationships between isolates.

Since meningococci exhibit a non-clonal population structure due to high levels of recombination (Holmes *et al.*, 1999), it is necessary to employ a typing technique that samples variation from around the genome (multilocus) and from genes that are evolving relatively slowly due to stabilising selection for conservation of function, *e.g.* house-keeping genes. Multi-locus enzyme electrophoresis (MLEE) (Selander *et al.*, 1986) was the first technique used meeting these criteria. MLEE indirectly detects the variation of house-keeping genes by analysing the electrophoretic mobility of their products, *i.e.* metabolic enzymes on a starch gel. Each isolate is assigned an electrophoretic type (ET) based on the resulting electrophoretic pattern. MLEE was able to identify the major groupings (clonal complexes) of *N. meningitidis* and those, the hyperinvasive lineages, most associated with disease (Caugant *et al.*, 1986a; Olyhoek *et al.*, 1987). A key difficulty with MLEE however is with comparisons of results between laboratories. Furthermore, it is quite technically demanding requiring up to 20 loci for resolution.

1.4.2.1 Multilocus sequence typing (MLST)

MLST builds upon the principles of MLEE but, instead of movement of enzymes on starch gels, it uses the DNA sequences of internal fragments of house-keeping genes to index variation (Maiden *et al.*, 1998). It therefore detects synonymous as well as non-synonymous substitutions, which are detected in MLEE by changes in the protein product. It has numerous advantages over MLEE including: (i) portability; sequence data can be stored and shared electronically; (ii) high resolution; seven loci can be used

to give high discrimination among isolates; (iii) unambiguity of results; (iv) can be integrated into a high-throughput framework; (v) further analysis can be carried out using the sequence data (Figure 1.13).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.13: MLST work flow. Illustration of the processes involved in obtaining an allelic profile for an isolate. From (Maiden, 2006).

Approximately 450–500 bp internal fragments of seven housekeeping genes are sequenced. Each locus is assigned an allele number based on its nucleotide sequence and a string of seven integers (allelic profile) gives the sequence type (ST). Allele numbers are assigned in order of discovery and are based on whether sequences differ at one or many nucleotides, thus not weighting according to the number of nucleotide differences, which could have occurred during a single genetic transfer event.

Both MLEE and MLST have shown that the invasive meningococcal population structure comprises clusters of closely related organisms known as clonal complexes. A clonal complex is defined a group of STs in which four of the seven house-keeping loci are shared in common with a central genotype. This central type is the putative 'ancestral genotype' from which all the other genotypes descended. A clonal complex

is confirmed by looking at a combination of clustering techniques such as eBURST (Feil *et al.*, 2004), UPGMA and split decomposition (Bandelt & Dress, 1992; Huson, 1998). The data concerning the proposed clonal complex is reviewed and confirmed by an international management committee. MLST is now a "gold standard" typing technique which is widely used and becoming more so as time wears on. It has been used to study the epidemiology and population biology of a wide-range of pathogens (Bisharat *et al.*, 2005; McCarthy *et al.*, 2007; Pullinger *et al.*, 2007).

A nomenclature scheme has been proposed which is compatible with earlier serological and MLEE schemes but incorporates molecular characterisation including MLST and antigenic gene typing (Jolley *et a l.*, 2007). The recommendation for the nomenclature is that it be of the form: serogroup: PorA type: FetA type: sequence type (clonal complex), *e.g.*: B: P1.19,15: F5-1: ST-33 (cc32).

1.5 Vaccines

Virtually all meningococcal disease is due to meningococci expressing one of six capsular polysaccharide antigens corresponding to serogroups A, B, C, X, Y and W135. Currently, there is no comprehensive vaccine against all meningococcal disease types due to the difficulty in producing a successful vaccine against serogroup B disease. Attempts to develop a vaccine against meningococcal disease began in the early part of the twentieth century with the use of killed whole cell preparations and then meningococcal toxoids in immunisations (Feavers, 2001; Frasch, 1995). The development of vaccines came in response to the numerous epidemics that had swept across the globe at the time. With the advent of antibiotics, interest in the development of vaccines waned slightly. However, with the increase in antimicrobial resistance and further epidemics, particularly in the armed forces, vaccine research resumed in earnest. Vaccines, in tandem with the discovery of sulphonamides and penicillin, have had an immense impact on the levels of morbidity and mortality due to the meningococcus.

1.5.1 Capsular polysaccharide vaccines

Following the discovery of the polysaccharide component of the meningococcal cell and its immunogenic properties, research into the development of polysaccharides as vaccine components commenced (Scherp & Rake, 1935; Scherp & Rake, 1945). The first capsular polysaccharide vaccines were developed ~40 years ago when the importance of the human antibody response to the meningococcal capsule was first elucidated (Artenstein *et al.* , 1970; Goldschneider *et al.* , 1969a; Goldschneider *et a l.*, 1969b; Gotschlich *et al.*, 1969). These vaccines were specifically designed to target the serogroups A and C which had been responsible for various epidemics worldwide since WWII. The vaccines proved effective in prevention of serogroup A and C disease in US military recruits and in epidemic control in Africa (Greenwood & Wali, 1980; Hassan-King *et al.*, 1988; Makela *et al.*, 1975).

The current WHO guidelines for control of epidemic meningococcal disease in Africa include the use of bivalent A/C polysaccharide vaccines in mass vaccination campaigns once a certain threshold level of disease incidence has been reached (WHO, 2003; WHO, 2007). Due to the emergence of epidemic serogroup W-135 in the African

country Burkina Faso, a trivalent (ACW-135) vaccine was developed and it successfully interrupted an outbreak situation in the country (Ahmad, 2004). Bivalent and tetravalent (ACYW-135) polysaccharide vaccines have been licensed and successfully used in various countries to impede epidemics and also to protect those at higher risk such as undergraduate students, patients with immunodeficiencies, lab-workers and those travelling to areas of high disease incidence. Those travelling to the *Hajj* pilgrimage are required to have a quadrivalent vaccination (ACWY Vac), following recent outbreaks during the event. Polysaccharide vaccines have a number of drawbacks however. They are poorly immunogenic in young children and infants who are those at most risk (Reingold *et al.* , 1985). Also, as polysaccharides are T-cell independent, they do not induce immunological memory so the period of protection is short (3–5 years) and as they do not prevent mucosal colonisation, there is no protection provided through herd immunity.

1.5.2 Glycoconjugate vaccines

Covalent conjugation of the meningococcal polysaccharide to a protein carrier (non-toxic diphtheria mutant toxin Cross Reacting Material 197 (CRM197) or tetanus toxoid) allows for the induction of a T-cell dependent immunological response and therefore immunological memory (Costantino *et al.*, 1992; Jennings & Lugowski, 1981; Robbins & Schneerson, 1990). Improvement of the immunological response by conjugation of a bacterial polysaccharide to a protein was first demonstrated by Avery and Goebel in 1929 with the pneumococcus (Avery & Goebel, 1929). The first conjugate vaccine to find great success, as it has almost eradicated the disease, was the *Haemophilus influenzae* type b vaccine (Heath & McVernon, 2002). Meningococcal conjugate vaccines have been shown to induce protective immunity in persons of all ages and are considered to be safe (Bramley *et al.*, 2001; Choo *et al.*, 2000; Richmond *et al.*, 1999).

The meningococcal C conjugate (MCC) vaccine was introduced by the Department of Health in the UK in October 1999 in response to increasing levels of serogroup C disease in the country. It was the first country to include the meningococcal conjugate vaccine in a routine infant immunisation programme (Miller *et al.*, 2001). Infants were immunised at two, three and four months of age and all children up to 18 years of age were targeted in a catch-up campaign from November

1999 to December 2000. The impact of the campaign has been a dramatic reduction in serogroup C disease of over 80% within 18 months of onset of the programme (Miller *et al.*, 2001). There was also a reduction of 66% in carriage of serogroup C in adolescents a year after the programme began (Maiden *et al.*, 2002). The effects of herd immunity, *i.e.* protection of the unvaccinated population, have been shown with a reduction of disease by 67% in this group thanks to the high vaccine coverage rate and the reduction in carriage (Ramsay *et al.*, 2003). Evidence of a drop in effectiveness of primary infant immunisation (at two, three and four months) more than one year after immunisation (Trotter *et al.*, 2004) led to a subsequent change in the schedule in September 2006 to two doses at three and five months and a booster at 12 months. Following the UK, MCC vaccine campaigns have been undertaken in various countries such as the Republic of Ireland, Spain, Netherlands, and Belgium with similar success.

A quadrivalent vaccine (Menactra®, Sanofi-Pasteur) targeting serogroups A, C, W135 and Y was licensed in the USA in 2005 for adolescents aged 11 to 12, those who have not been vaccinated before high-school entry at 15 years, college freshmen and various individuals at risk (Bilukha & Rosenstein, 2005). Also underway is the Meningitis Vaccine Project (MVP), a collaboration with WHO, PATH (Program for Appropriate Technology in Health) and various worldwide parties which aims to produce a conjugate serogroup A vaccine for Africa at an affordable price and eventually eliminate epidemic serogroup A meningococcal disease in Africa (Jodar *et al.*, 2003; LaForce *et al.*, 2007). Prevention of epidemics by use of a conjugate vaccine providing a longer period of protection and herd immunity would be more practical than the current strategy of epidemic control using polysaccharide vaccines.

1.5.3 Serogroup B vaccines

Serogroup B is one of the leading causes of meningococcal disease worldwide, particularly in industrialised countries (Jones, 1995; Pollard *et al.*, 2001; Trotter *et al.*, 2006a). However, to date no successful polysaccharide or conjugate vaccines have been developed against serogroup B disease. This is a consequence of its similarity to host antigens, which also raises concerns about the safety of serogroup B polysaccharide as a vaccine component (Finne *et al.*, 1983). There is poor mammalian immune response to the (α2-8)-linked polysialic acid containing polysaccharide capsule. This molecule is a

surface component of many foetal and mammalian tissues and NCAM and is present in developing neural tissue of the embryo which when exposed to a developing foetus induces immunological tolerance (Finne *et al.*, 1983; Finne *et al.*, 1987). An evaluation of the immunogenicity of purified serogroup B polysaccharide in humans showed little serum antibody response (Wyle *et al.* , 1972). Therefore, a capsular serogroup B vaccine could be poorly immunogenic and could also possibly induce an auto-immune response, although this is disputed (Stein *et al*., 2006). Chemical modification of the molecule by replacement of the *N*-acetyl group with *N*-propionyl and conjugation with a protein carrier has been tested and proven immunogenic in lab animals and in adults with no autoantibody production, but *in vitro* tests have shown no functional antibody activity in humans (Bruge *et al.*, 2004; Jennings *et al.*, 1986). Further extensive work is needed to evaluate the extent of immunogenicity and safety of this method. According to Poolman and Berthet (Poolman & Berthet, 2001), ideally a serogroup B vaccine should: (i) be safe and immunogenic in the paediatric population, (ii) be able to elicit protection against a wide range of clinical strains and (iii) be cost-effective and easy to manufacture at a large scale.

1.5.3.1 Outer membrane protein vaccines

As a result of the lack of success with vaccines directed against the B capsular polysaccharide, and the finding that in serogroup B disease bactericidal antibodies are mainly directed at non-capsular surface antigens, most vaccine research has focussed on a variety of sub-capsular cellular components, particularly OMPs (Frasch, 1995; Jodar *et al.* , 2002). There are several drawbacks of these antigens as vaccine candidates however: (i) by their very nature they are highly variable making a comprehensive vaccine difficult to formulate; (ii) the protective efficacy in young children has yet to be shown; (iii) the duration of protection is relatively short-lived.

Vaccines have been developed from OMVs which are produced by cells during culture. They contain antigenic cell components such as OMPs (mainly PorA) and can be purified for use in vaccines (Frasch *et al.*, 2001). Numerous trials have been carried out using OMV vaccines and they have been used to tackle single clone epidemics with some success in Cuba, Norway and New Zealand (Bjune *et al.*, 1991; O'Hallahan *et al.*, 2005; Rodriguez *et al.*, 1999). The Cuban and Norwegian vaccines were developed in response to epidemics occurring in the respective countries during the 1980s (Bjune *et*

al., 1991; Sierra *et al.*, 1991). The Cuban vaccine was developed from a single strain (B:4:P1.19,15) and showed 83% efficacy in teenagers. The same vaccine was trialled in Brazil and gave varying degrees of efficacy depending on age (74% in 2–6 years age group) with little or no protection in infants (de Moraes *et al.*, 1992). The Norwegian vaccine based on the ET-5 (ST-11) clone B:15:P1.7,16 strain was found to have 57% efficacy in teenage school students (Bjune *et al.*, 1991). In New Zealand since 1991, there has been an epidemic caused by a single strain (B:4:P1.7b,4) (Baker *et al.*, 2001). This lead to the development of a tailor-made vaccine MeNZB (Oster *et al.*, 2005) which was implemented in a vaccine campaign begun in 2004 targeting those under the age of 20.

Overall, OMV vaccines have shown good efficacy in older children but not in infants and there are questions over duration of protection. Also, their immunogenicity is limited to the vaccine strain. PorA, the main protein found in the vaccines, has been shown to be the major immunogen (Milagres *et al.*, 1994). As it is highly variable and able to undergo variation, OMV vaccines containing just one PorA type may be very restricted in coverage in an endemic setting. Wider coverage against variable antigens can be attained by inclusion of multiple variants in vaccine formulations and multivalent PorA vaccines such as NonaMen have been developed (van den Dobbelsteen *et al.*, 2007). Preclinical trials of NonaMen, which was combined with a pneumococcal vaccine, showed it was immunogenic in mice and rabbits (van den Dobbelsteen *et al.*, 2007).

FetA, despite its variability (Thompson *et a l.*, 2003), is another immunogenic OMP with vaccine potential (Wedege *et al.*, 1998). The diversity of the antigen is structured and maintained over time and is related to the hyperinvasive lineages (Russell *et al.*, 2008; Urwin *et al.*, 2004). Widespread and persistent antigenic structuring is also found with PorA and other OMPs. Therefore, a potentially broad coverage vaccine could be created from a combination of a relatively small number of PorA and FetA antigen components (Russell *et al.*, 2008; Urwin *et al.*, 2004).

Alternatively, other, more conserved antigens such as NadA and NspA have been evaluated. NadA can elicit bactericidal antibodies and protect in the infant rat model, but the gene is found in 50% of disease-associated serogroups B, C, Y, W-135 (Comanducci *et al.*, 2002) and 50% of serogroup B isolates (Beernink *et al.*, 2007). NspA can elicit protective antibody responses against serogroups A, B and C in mice and is expressed in all isolates (Martin *et al.*, 1997; Martin *et al.*, 2000a) although levels

of expression may vary in serogroup B isolates (Moe *et al.*, 1999). Despite its variability (Rokbi *et al.*, 2000), TbpB has also been considered as a vaccine candidate due to its expression in all meningococci, its surface-exposure and its ability to elicit bactericidal antibodies (Ala'Aldeen & Borriello, 1996; West *et al.*, 2001).

1.5.3.1.1 The 'reverse vaccinology' approach to vaccine discovery

The advent of whole-genome sequencing of bacterial species has provided a wealth of information to be mined (Tettelin *et al.*, 2000). One of the ways in which this information has been exploited is in the discovery of novel vaccine candidates by 'reverse vaccinology' (Rappuoli, 2000). Reverse vaccinology is a method whereby a bacterial pathogen's whole genome is analysed computationally to predict *in silico* previously unidentified vaccine candidates (Figure 1.14).

It was first applied to the meningococcal serogroup B MC58 genome (Pizza *et al.*, 2000). Bioinformatic algorithms were used to screen for open reading frames (ORFs) potentially encoding surface-exposed or excreted proteins. Six hundred novel genes were identified of which 350 were successfully expressed in *E. coli*. They were then purified and used to immunise mice. The antisera were then used for *in vitro* assays such as fluorescence-activated cell sorting analysis (FACS) and ELISA to assess surface localisation of the antigens. SBAs were also carried out to measure the ability of specific antibodies to kill the pathogen. Of these genome-derived antigens (GNA), 28 elicited complement-mediated bactericidal antibody response. The selected candidates were screened across a diverse range of *N. meningitidis* strains and other *Neisseria* species to check for sequence conservation. This yielded a small number of antigens which were both conserved in sequence and able to elicit a cross-bactericidal response against all the strains in the panel. This demonstrated that they could potentially confer general protection against meningococcus and have since been subject to further investigation. One antigen discovered by this means is fHbp. It was also independently identified through biochemical fractionation and shown to elicit bactericidal response against diverse meningococcal strains (Bernfield *et al.*, 2002; Fletcher *et al.*, 2004). fHbp is a component in two recombinant vaccines in clinical development at the time of writing. One contains a protein from each of the subfamilies (Anderson *et al.*, 2009), and the other contains a fHbp subfamily B/variant 1 protein in combination with a number of other antigens, namely, other genome-derived antigens

GNA2091, GNA1030 and Neisserial Heparin-binding antigen (NHBA, formerly known as GNA2132), NadA, and PorA P1.4 (Biolchi *et al.*, 2009; Giuliani *et al.*, 2006).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.14: The reverse vaccinology approach to vaccine candidate discovery. (a) Firstly, bioinformatics methods are used to screen for potential ORFs. (b) Candidate antigens are, exported proteins, outer membrane and porin-like proteins, periplasmic proteins or integral membrane proteins. (c) Genes are expressed in *E. coli*, purified and used to immunise mice. (d) Antisera are then used for *in vitro* assays such as FACS and ELISA to assess localisation in the cell surface and SBA for bactericidal activity of antibodies against the pathogen. (e) The selected vaccine candidates are then screened for sequence conservation on a panel of strains including diverse members of *N. meningitidis* and other *Neisseria* species. From (Muzzi *et al.*, 2007).

1.5.3.2 Other vaccines

Meningococcal LOS is a major immunogen and potential vaccine candidate (Plested *et al.*, 2003). There was concern however about the toxicity of the lipid A moiety (Verheul *et al.*, 1993) but it has been possible to produce a detoxified form of the molecule. Concern over the variability of the outer-core structure has led to the investigation of the more conserved inner-core region although its surface accessibility may be an issue (Plested *et al.*, 2003).

The closely related commensal species *N. lactamica* is also being considered as a vaccine candidate for a number of reasons. Firstly, since carriage of *N. lactamica* peaks in young children at a time when meningococcal carriage is rare (Figure 1.2), the implication is that the carriage of the commensal induces protective immunity against meningococcal colonisation (Bennett *et al.*, 2005; Gold *et al.*, 1978). Secondly, although *N. lactamica* does not possess a capsule or PorA, it does share many antigens with the meningococcus, and these demonstrate antigenic cross-reactivity (Troncoso *et al.*, 2000). A study of mouse immunisation with *N. lactamica* killed whole cells, OMVs or OMP pools showed protection against a lethal challenge against diverse meningococcal strains (Oliver *et al.*, 2002). However, there was a lack of a bactericidal response which may indicate other mechanisms involved in protection. A *N. lactamica*-based OMV vaccine is being developed using similar methods to those for the well-studied meningococcal OMV vaccines. Pre-clinical results have again shown protection against lethal challenge with diverse meningococci following OMV immunisation (Gorringe *et al.*, 2005). Also, sera obtained from mice and rabbits against *N. lactamica* or *N. meningitidis* OMVs showed similar cross-reactivity against OMVs from a panel of diverse strains (Finney *et al.*, 2008). Recent data from phase 1 tests in adult male volunteers demonstrated a good safety profile and immunogenicity and an increase in opsonophagocytic activity but a modest increase in SBA (Gorringe *et al.*, 2009).

1.6 Population genetics

Bacteria are considered asexual organisms as they reproduce by binary fission with the haploid mother (progenitor) cell giving rise to two identical haploid daughter cells. It was long believed that the only way in which variation could be introduced into a bacterial population was through point mutation, insertions and deletions. However, bacterial life and reproduction has been revealed to be not so straightforward (Levin & Bergstrom, 2000). Recent developments in bacterial typing techniques and nucleotide sequencing have exposed more complexity in bacterial populations than first thought. This complexity was revealed to be due to genetic exchange events and it precipitated studies of the impact of recombination on bacterial populations.

1.6.1 How clonal are bacteria? – models of bacterial population structure

The clonal model of bacterial population structure (Figure 1.15 (a)) assumes no horizontal gene transfer between bacteria, with vertical transmission of genetic material from the progenitor cell to its subsequent identical progeny. Diversity is brought about by *de novo* mutations and this variation can only be passed on vertically to the following generations. Accumulation of selectively favourable mutations over time that confer fitness to the organism may lead to the emergence of new lineages, resulting in a loose assembly of related lineages which can represented with a bifurcating tree-like phylogeny (Maynard Smith, 1989). Stochastic events such as mutation, periodic selection and bottle-necks act on the population to increase/reduce diversity. Periodic selection occurs when a fit genotype emerges and multiplies at a cost to other genotypes which then become reduced in frequency (Levin, 1981). Bottlenecking is another diversity-reducing event where a great reduction in population size means only a small number of bacteria go on to reproduce thus narrowing genetic variability (Achtman, 1997; Morelli *et al.* , 1997). In the absence of recombination, clonal populations typically display 'linkage disequilibrium' which means the non-random association of alleles at one or more loci (Maynard Smith *et al.*, 1993). Very few examples of truly clonal bacteria exist with *Salmonella enterica* being a prime example.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.15: Models of bacterial population structure based on descent. (a) The clonal model. There is no recombination with the only variation being brought about by point mutation. Accumulation of mutations may lead to diversification and expansion of new genotypes. However, diversity may be reduced from time to time by periodic selection and bottlenecking. (b) Panmictic/non-clonal model. There is high diversity and recombination with low linkage disequilibrium. (c) The epidemic clone/clonal expansion model. A successful genotype may emerge and dominate as there is not enough recombination to prevent it. The epidemic clone dominates the population and may persist for a time or disappear and be replaced by another. From (Gupta & Maiden, 2001).

The relative contributions of recombination and mutation determine the population structure of a bacterium and vary widely among species (Spratt & Maiden, 1999). At one extreme is the clonal population (Figure 1.15 (a)) with no recombination and linkage equilibrium, and, at the other end is the freely recombining non-clonal, sometimes referred to as panmictic, population (Figure (1.15 (b)). A non-clonal population is characterised by high amounts of recombination, linkage equilibrium and high levels of diversity and would be best represented as a net-like phylogeny. Examples of bacteria with this kind of population structure are *Helicobacter pylori* and *Neisseria gonorrhoeae* (Go *et al.* , 1996; O'Rourke & Spratt, 1994). Most bacteria occupy the middle-ground of the spectrum, ranging from clonal to non-clonal populations, where there is some recombination but not enough to deter the emergence of clonal lineages (Figure 1.15 (c)). From time to time a fit strain may emerge due to favourable conditions and expand dominating a population ('epidemic clone') (Maynard Smith *et al.*, 1993). The clone may persist and diversify over time and eventually die out and be replaced.

Recombination in bacteria occurs as a result of horizontal genetic transfer also referred to as 'localised sex' which arises by the parasexual mechanisms of conjugation, transduction and transformation (Maynard Smith *et al.* , 1991). Conjugation allows transfer of genetic material by direct cell-to-cell contact, transduction is a process of transfer of DNA *via* bacteriophages and natural transformation is the uptake by a competent recipient cell of 'foreign' DNA which is free in the environment. Homologous recombination requires as little as 70% nucleotide sequence identity to take place (Lorenz & Wackernagel, 1994) and can occur between closely related species (Frosch & Meyer, 1992). The result of horizontal genetic exchange is a genome pock-marked by small DNA fragments, which can be up to 9.9 kb in length, from other bacteria (Linz *et al.*, 2000; Maynard Smith *et al.*, 1991; Maynard Smith, 1992). These so-called 'mosaic genes' are composites consisting of DNA from the recipient cell and the genetically distinct donor strain (Figure 1.16) (Maiden, 1993).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.16: The formation of mosaic genes. A bacterium can obtain novel genetic material by horizontal gene transfer. This can occur *via* conjugation, transduction or transformation. Where there is at least 70% nucleotide sequence identity, homologous recombination can occur. A gene (gene A) from one strain can be introduced into the chromosomal genes of another strain as a result of homologous recombination. This produces a chromosomal mosaic gene. This type of event can occur numerous times (*e.g.* introduction of gene B) to produce a gene pock-marked by insertions of various foreign DNA fragments. Adapted from (Maiden, 1993).

1.6.1.1 Neisseria meningitidis – 'an epidemic clone' paradigm?

Two important findings have helped place *N. meningitidis* in the spectrum of bacterial population structure; first, the discovery of clonal lineages by MLEE (Caugant *et a l.*, 1986b; Caugant *et al.* , 1987b), and second, evidence of recombination (Feil & Spratt, 2001). The support for recombination is the presence of mosaic genes (Feil *et al.*, 1995; Zhou *et al.*, 1997), a net-like phylogeny of house-keeping loci as demonstrated by split decomposition (Holmes *et al.*, 1999) and incongruence among trees based on the house-keeping loci (Figure 1.17) (Feil *et al.*, 2001; Holmes *et al.*, 1999).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.17: Evidence of recombination – Incongruence. Congruence testing of seven *N. meningitidis* house-keeping loci using Maximum Likelihood (ML) analysis. A ML tree is drawn for each locus and is compared with the other 6 loci. The difference between the likelihoods of the loci (*pink* squares) and 200 randomly drawn trees (*navy* diamonds) is given as Δ-lnL. The dotted line is the 99th percentile of the difference in likelihood between the ML tree for each locus and the random trees. If the Δ-lnL values of the ML tree comparisons fall within this percentile then they are significantly different and hence incongruent. From (Feil *et al.*, 2001).

MLST has been used to estimate the relative contributions of recombination and mutation to meningococcal diversity and it has been shown that a single nucleotide site in a meningococcal house-keeping gene is at least 80 times more likely to change as a consequence of recombination than mutation (Feil *et al.* , 1999). The fact that the meningococcus is a naturally transformable bacterium means that it is competent for uptake of foreign DNA from the environment (Maiden, 1993).

While representative samples of meningococcal disease isolates cases may give the impression of a clonal population, carriage samples show quite a large amount of diversity. Therefore, in order to get a picture of complete population diversity, it is important to sample not just the invasive disease population but also the carriage population. The background population freely recombines and is quite diverse exhibiting a net-like structure (Figure 1.18). From time to time clones emerge and multiply in the population. In *N. meningitidis* these clonal lineages are the genotypes most associated with disease.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.18: The emergence of clones from the background diverse population of *N. meningitidis*. The background population which exhibits a net-like rather than a tree-like phylogeny due to high levels of recombination is composed of numerous relatively rare and unrelated genotypes. From this population a few clones (*grey* cones) can emerge and over time grow and diversify (outward arrows) forming clonal complexes with a central ancestral genotype (*black* circles). From (Feil & Spratt, 2001).

The persistence of clonal lineages over long time periods and in diverse regions cannot completely be explained by a microepidemic model however, since the amount of recombination that occurs in the meningococcus would be expected to erase the clonal structure over time. Instead, models of selection have demonstrated that the longevity of these lineages is due to their slightly enhanced transmission efficiency (Figure 1.19 (a)) which also allows them to tolerate the penalty of increased virulence which is of no benefit for further transmission (Buckee *et al.*, 2008).

Host immune selection has an important role in the structure of the meningococcal population and is demonstrated clearly in the non-overlapping distribution of PorA VR associations (Gupta *et al.*, 1996). Various studies have shown this non-random structuring (Russell *et al.*, 2008; Urwin *et al.*, 2004). This phenomenon cannot be explained by neutral processes or the microepidemic model because of recombination and the existence of the same antigen combinations in otherwise unrelated STs. The model of strain structuring postulates that a dominant host immune response against a polymorphic determinant (PorA, FetA) which has moderate to strong cross-protection among genotypes sharing the same alleles at strain loci will create a non-overlapping population dominated by subset of strains (Figure 1.19 (b)). Recombination may continue to create strains that share alleles with the dominant strains but they will be at a disadvantage due to host immune response against them and may exist in relatively small numbers.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.19: (a) Emergence of lineages. The effects of competition on the coexistence of strains, characterised by very small differences in transmission efficiency, within a hypothetical six-locus, two-allele system. Blue bars indicate the range of competition over which each strain can survive. On these in red is the smaller range of competition over which the strain can afford to carry excess virulence. The number of cocirculating strains decreases as competition for available hosts increases, and fewer of these strains are able to carry the burden of excess virulence. From (Gupta & Maiden, 2001) and (Buckee *et al.*, 2008) respectively. (b) Schematic representation of the effect of immune selection on diverse pathogen populations and models of strain abundance with time.

1.7 Epidemiology of meningococcal disease

1.7.1 Worldwide epidemiology

The meningococcus is capable of causing sporadic, endemic, epidemic and pandemic disease, significantly contributing to annual worldwide morbidity and mortality. Outbreaks of so-called 'spotted fever' were documented in the United States and Europe as far back as the 19th century (Cartwright, 1995). The onset of World War One precipitated outbreaks of meningococcal disease in army camps — particularly in new recruits — of combatant countries such as the USA, UK, Canada and Germany. Serogroup A organisms were responsible for these outbreaks and continued to cause further epidemics on both sides of the Atlantic in the inter-war period. The Second World War triggered further and even larger outbreaks in combatant countries with most disease, although still associated with the military, found in the civilian population. In the UK in 1940 for example there were 12,000 notified cases (Cartwright, 1995). Other European and involved countries such as Australia and New Zealand suffered comparably. Fortunately, around this time, the discovery of antimicrobials — sulphonamides and subsequently penicillin — greatly reduced case-fatality rates to below 20%. In the 1950s there was a gradual change in the serogroups associated with disease in developed countries with serogroup A being replaced with the more diverse serogroups B and C. Serogroup A is now very rare in these places. There were little or no epidemics in developed countries in the post-war era, but developing parts of the world such as Africa continued to endure them.

Today, there is a relatively large degree of variation in worldwide invasive meningococcal serogroup distribution, but it can roughly divided by region (Harrison *et al.*, 2009) (Figure 1.20). In Western Europe, most disease is caused by serogroup B followed by serogroup C with serogroup A disease being very rare. Similarly, disease in North and South America and other developed countries such as New Zealand and Australia are dominated by the B and C serogroups with a scarcity of serogroup A (Harrison *et al.*, 2009; Muros-Le Rouzic *et al.*, 2006). A recent development however, is the emergence of serogroup Y (ST-23 complex) disease in the USA and Canada (Harrison *et al.*, 2006; McEllistrem *et al.*, 2004). The proportion of disease due to this serogroup in the United States rose from 2% in 1989–1991 to 37% in 1997–2002 (Bilukha & Rosenstein, 2005).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.20 (a) and (b): Worldwide distribution of major meningococcal serogroups. ****N* is total known serogrouped isolates.** ***Other*** **includes other serogroups, non grouped or non groupable isolates. From (Harrison** ***et al.*****, 2009) and (Muros-Le Rouzic** ***et al.*****, 2006) respectively.**

Developing countries continue to have the highest burden of morbidity and mortality due to meningococcal disease and also seem to show a different epidemiology. Africa is dominated by serogroup A which is associated with epidemic waves of disease, particularly in the 'Meningitis Belt' region. In Russia, serogroup A disease dominates, followed by B and C. The more recently emerged serogroup W-135 is associated strongly with disease in the Middle-East and Asia and emerged from Saudi Arabia during the Hajj pilgrimage in 2000. It then was spread worldwide by returning pilgrims prompting increased cases in some European countries, the USA and Asia (Aguilera *et al.* , 2002; Taha *et al.* , 2000) but epidemics in countries such as Burkina Faso in Africa (Decosas & Koama, 2002). It has recently increased significantly in Argentina increasing from about 4% in 2002 to 28% in 2008 (Efron *et al.* , 2008). Relatively scant information is available on meningococcal disease in Asia, but data from China indicate predominance of serogroup A-related disease (Zhang *et a l.*, 2008) and in Japan serogroup B and Y disease predominate (Takahashi *et al.*, 2004).

1.7.1.1 Major disease-associated meningococcal lineages

Serogroup A is associated with higher disease incidence compared to the other serogroups. Within the group, three clonal complexes, which are exclusively associated with this polysaccharide capsule type, have been associated with pandemic/epidemic spread of disease. These are the ST-1 complex, ST-4 complex and ST-5 complex (Achtman, 1990; Olyhoek *et al.*, 1987; Wang *et al.*, 1992). Organisms of ST-1 clonal complex were the cause of the disease outbreaks in the USA and UK during WWII. Since the early 1960s, this complex has been responsible for epidemics in numerous locations including North and West Africa, the USA, Canada, New Zealand and Australia (Achtman, 1995). The ST-5 complex has been responsible for at least two pandemics since the 1960s (Achtman, 1997) (Figure 1.21). The first originated in China, spreading to Eastern Europe, Russia, Scandinavia and then Brazil. The second, which began in the early 1980s also originated in China moving on to India and Nepal before being introduced into Saudi Arabia by South Asians travelling to Mecca for the Hajj pilgrimage (Moore *et al.* , 1989). Returning pilgrims introduced the organism to Africa instigating major epidemics in countries such as Sudan and Chad. The more recent and closely associated clone ST-7 has emerged and is gradually replacing ST-5 as the predominant complex associated with disease in Africa (Nicolas *et al.*, 2001).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.21: ST-5 complex pandemic spread. The first pandemic began in China spreading to Russia, Scandinavia and Brazil (arrows indicate movement of the pandemic waves). The second pandemic spread from China and Nepal to Mecca and on to Africa. The dotted lines indicate carriage of the organism by pilgrims to the US and some European countries. Dark and light shading indicate reported and untested epidemics respectively. From (Achtman, 1997).

Unlike the ST-1, ST-4 and ST-5 complexes, ST-11 complex is associated with more than one serogroup, *i.e.* B, C, W-135, Y (Trotter *et al.*, 2007; Wang *et al.*, 1993). Outbreaks and epidemics linked to the ST-11 complex have been documented since the 1960s when there were high numbers of cases in US army recruits during the Vietnam War (Caugant, 2001). A new variant clone associated with increased virulence identified by MLEE as ET-15 (it bears a *fumC* allele infrequent in other serogroup C strains) emerged in Canada in the late 1980s causing elevated levels of serogroup C disease in the country, prompting mass immunisation (Ashton *et al.*, 1991; Whalen *et al.*, 1995). The clone then spread worldwide (Jelfs *et al.*, 2000a) increasing disease incidence and causing outbreaks of serogroup C disease especially in teenagers and young adults. This lead to the eventual implementation of MCC vaccination campaigns in countries such as the UK, Spain and the Netherlands. The ST-11 complex has been found to be significantly associated with sepsis and poor clinical outcome (Heckenberg *et al.*, 2008). The W-135 outbreaks which occurred in and emanated from the Hajj

pilgrimage in Saudi Arabia in 2000 were ST-11 complex strains (Taha *et al.*, 2000). Pilgrims wishing to travel to Mecca are now required by the Saudi government to have the quadrivalent (ACYW-135) vaccine.

The ST-32 complex, which is associated with serogroup B, appears to have been uncommon prior to the 1970s and is thought to be a recently emerged clone (Achtman, 1995; Caugant, 2001). From the mid-1970s it appeared and began to cause elevated levels of disease in Norway and Spain (Caugant *et al.*, 1987a). It then spread intercontinentally initiating epidemics in South Africa in the 1970s and Cuba, Chile and Brazil in the 1980s (Caugant *et al.*, 1986b). Particular clones within the ST-32 complex were responsible for the Norwegian and Cuban epidemics (B:15:P1.7,16 and B:4:P1.19,15 respectively) and vaccines (MenBvac and VA-MENGOC-BC®) were developed and implemented with some success to target these strains. Disease returned to endemic levels in Norway in the early 2000s after 25 years. Disease levels have also been greatly reduced in Cuba since the introduction of its vaccination programme. The USA which has previously not experienced epidemics involving ST-32 complex-serogroup B disease has shown increases in levels in parts of the country (Oregon, Washington State) of strains with the same phenotype as that of the Norwegian epidemic strain (Diermayer *et al.*, 1999).

The ST-41/44 complex is highly diverse and is the only one for which it is necessary to have two central genotypes; ST-41 is disease-associated while ST-44 is rarely associated with disease. This clone was first identified associated with disease in 1980 in Netherlands and is now found globally (Caugant, 2001). A serogroup B ST-41/44 complex strain (B:P1.7-2,4) was the cause of a prolonged disease epidemic in New Zealand which began in 1991 (Baker *et al.*, 2001). A 'tailor-made' OMV vaccine was developed based on the Norwegian Men-B vaccine targeting the causative strain (Dyet *et al.*, 2005; Oster *et al.*, 2005). The ST-8 complex is related to the ST-11 complex and disease-related strains of this type have been reported since the 1970s and 1980s in the USA, Canada, South Africa and Europe (Achtman, 1995). New complexes that are disease-associated such as ST-269, ST-162 and ST-213 have emerged more recently (Law *et a l.*, 2006; Yazdankhah *et al.*, 2005a). The USA and Canada have experienced increases in the serogroup Y disease in recent times and this is linked to the ST-23 clonal complex (Harrison *et al.*, 2006).

1.7.2 Africa and 'The Meningitis Belt'

> *Contrairement à ce qui se passe dans les zones tempérées d'Europe et d'Amérique du Nord, où elle ne sévit plus que sous forme sporadique, la méningite cérébro-spinale à méningocoques (MCS) constitue en Afrique un problème de santé publique. Ce problème est surtout préoccupant dans une étroite bande de terrain courant de l'Atlantique à la Mer Rouge et comprise entre le 4e et 16e degré de latitude Nord.*
>
> *Lapeyssonnie, 1963*

Epidemics have been recorded in the African continent since the mid-19th century (Greenwood, 2006). The term 'Meningitis Belt' was coined by Lapeyssonnie in 1963 (Lapeyssonnie, 1963) and refers a band of the African continent that runs from the west to east coast from Senegal to Ethiopia in the semi-arid sub-Saharan region (Figure 1.22 (a) and (b)). It suffers from high endemic levels of meningococcal disease and very large epidemics which occur every 5 to 12 years although the intervals between epidemics have become more irregular in the last couple of decades (LaForce *et al.* , 2007) (Figure 1.22 (c)). The area of the belt includes 21 countries and a population of ~420 million. More recently, there is evidence that the boundaries of the area may be extending southwards, most likely due to increasing desertification, as reports of epidemics in these regions are increasing (Molesworth *et al.*, 2002).

Africa has the highest burden of meningococcal disease in the world with annual attack rates in non-epidemic years between 10–150 per 100,000 and in epidemic years 250–1000 per 100,000 (Perea, 2007). The highest attack rates are found in the under fives and the 5–14 years age group with a case-fatality ratio of 10–20%. The worst epidemic season of recent times was in 1995–1996 with 201,000 cases and 14,500 deaths (WHO, 2003). Epidemics, when they occur, are in conjunction with the dry season when there is low humidity and the hot dusty *Harmattan* wind blows from the Sahara (Sultan *et al.*, 2005). This 'epidemic season' which lasts about 26 weeks, begins in late December, peaks at the end of the dry season (~13/14 weeks) and stops abruptly when the rains begin in May and June (LaForce *et al.* , 2007). Environmental factors such as absolute humidity, land-cover type, population density and dust are implicated in increased risk of epidemics (Figure 1.22 (b)) (Molesworth *et al.*, 2003). It is thought that the harsh dusty winds may aggravate the nasopharyngeal mucosa thus making individuals more susceptible to bacterial invasion.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.22: (a) Countries in the African meningitis belt including those outside it which have had epidemics (*red* dots) as at 2004. (b) Fitted model of observed meningitis epidemics based on environmental variables (humidity, rainfall, aerosol index, land-cover type, population density). (c) Trends in epidemic meningococcal meningitis in the African meningitis belt 1970–2006. From www.ncid.cdc.gov, (Molesworth *et al.*, 2003) and (WHO, 2007) respectively.

Serogroup A has been the predominant serogroup associated with epidemics in Africa since at least the 1960s (Achtman, 1995). The central genotype of the ST-5 complex is currently the most common clone involved (Nicolas *et al.*, 2005) although more recently other members of this complex, ST-7 and ST-2859, have emerged and may replace ST-5 (Nicolas *et al.*, 2001; Norheim *et al.*, 2006; Teyssou & Muros-Le Rouzic, 2007; WHO, 2007). The ST-2859 strain has been isolated in West Africa since 2003 and was responsible for a large epidemic of 19,000 cases in Burkina Faso. It subsequently caused outbreaks of 18,000 and 26,000 cases respectively in the years 2006 and 2007 and has since been isolated in Togo, Niger and Mali (Nicolas *et a l.*, 2008; Norheim *et al.*, 2008). The first epidemic of serogroup W-135 (ST-11 complex) disease occurred in Burkina Faso in 2002 following the return of pilgrims from Saudi Arabia despite strains of this group being in circulation in West Africa for many years (Ahmad, 2004; Greenwood, 2006). No further epidemics involving this serogroup have occurred although the situation is under surveillance. Serogroup C and the more recently emerged serogroup X (ST-181) (Boisier *et al.*, 2007; Njanpop-Lafourcade *et al.*, 2008) cause a small number of occasional outbreaks in the belt.

The current means of tackling epidemic situations involves surveillance of disease levels and, when a certain threshold of disease cases is passed, implementation of a mass vaccination strategy using bivalent or trivalent polysaccharide vaccines (WHO, 2007). There are several limitations to the use of polysaccharide vaccines in this scenario: immunity provided is short-lived; the vaccines have minimal effectiveness in infants and young children; there is neither reduction in carriage nor increase in herd immunity. However, to date, the polysaccharide vaccine has been the only affordable option in Africa to tackle epidemics. The MVP project is currently under way and its target is to produce a vaccine (MenAfriVac^{TM}) for Africa at \$0.40 a dose (LaForce *et al.*, 2007; Roberts, 2008). It is in Phase II and II/III trials at the time of writing in various African countries and is hoped to be rolled by 2009/2010 firstly in the most at-risk countries, *i.e.* Burkina Faso and Niger.

1.7.3 European epidemiology

The epidemiological pattern in Europe shows seasonal peaks of meningococcal infection in the first quarter of the year during winter/early spring. Possible reasons for this include increased susceptibility due to pulmonary infections such as influenza, decreased ventilation indoors and closer personal contact (Jones, 1995).

There is considerable variation in the incidence of meningococcal disease across Europe with no apparent geographical pattern. The lowest incidence is 0.30 per 100,000 in Italy and the highest 4.96 per 100,000 in the Republic of Ireland as of 2006 (Figure 1.23) (EU-IBIS, 2007).

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.23: Incidence of invasive meningococcal disease in Europe, 2005. From (EU-IBIS *et al.*, 2007).

There has been a decline in overall meningococcal disease incidence in Europe since 1999 dropping from 1.9 per 100,000 to 1.1 per 100,000 in 2006. The ranking of the countries in terms of incidence has remained relatively stable over this time (EU-IBIS *et al.*, 2007). The decline in incidence is most probably due to the implementation of MCC vaccination strategies in different European countries (Table 1.1) and its associated herd immunity, although other factors may be important such as the natural fluctuations in incidence that occur in countries over time (Jones, 1995). The first country to implement routine MCC immunisations was the UK in 1999. Countries with more than a two-fold drop in incidence between their peak incidence year (in brackets) and 2006 include several that implemented the vaccine: Belgium (2001), Denmark (1999), Greece (2000), Iceland (1999), Ireland (1999), Netherlands (2001), Norway (2000), Switzerland (2000) and UK (1999). Although the Republic of Ireland still has the highest incidence in Europe there has been a large overall decrease since 1999 from 11.89 per 100,000 to 4.96 in 2006; it implemented the MCC vaccine in 2000 (EU-IBIS, 2007). Unsurprisingly, there has been a marked decrease (10-fold) in serogroup C disease in countries which implemented MCC vaccine campaigns (Figure 1.24). The level of C disease in countries which *did not* implement the MCC vaccine has remained relatively stable in this time.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.24: Incidence of serogroup C invasive meningococcal disease in European countries with and without routine vaccine MCC, 1999–2006 (countries with consistent data only). Adapted from (EU-IBIS *et al.*, 2007).

Table 1.1: MCC vaccination programmes in European countries as at October 2007 (Reproduced from (EU-IBIS, 2007)).

The table originally presented here cannot be made freely available via ORA for copyright reasons.

Serogroup B remains the predominant cause of disease in Europe, followed by serogroup C, and together they account for 90% of disease cases (Figure 1.25) (EU-IBIS, 2007). Meningococcal disease is most prevalent in the under five years age group; 50% of serogroup B disease and 25% serogroup C disease are in this age group. Serogroup B disease tends to have a lower age profile than serogroup C. Serogroup B disease declined overall by 40% in MCC countries and in individual countries remained stable or declined. This indicates that there has been no capsule-replacement with serogroup B strains filling the niche left by the reduction of serogroup C.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.25: Percentage distribution of serogroups causing laboratory-diagnosed, confirmed and probable meningococcal disease in Europe 2006. From (EU-IBIS, 2007).

In 2006 serogroup Y was the major non-B non-C serogroup contributing to meningococcal disease followed by W-135 which increased in importance in 2001/2002 following the Hajj. Notably, the three Scandinavian countries Denmark, Norway and Sweden had relatively higher proportions of disease due to serogroup Y and W-135 (10–15%). The highest case-fatality ratios are found with serogroups C, W135 and Y (EU-IBIS, 2006). Disease associated with the epidemic A serogroup is very rarely found across Western Europe but is notably found in Russia and former Soviet states.

Certain clonal complexes are more frequently associated with particular serogroups (Figure 1.26). In a study of European disease isolates from 1999–2004, 94% of isolates belonging to the ST-41/44 clonal complex, 94% of ST-32 complex and 96% of ST-269 complex were found to be serogroup B (Trotter *et al.*, 2007). 83% of isolates of the ST-11 clonal complex and 86% of the ST-8 complex were serogroup C. Serogroup Y was associated with 90% of ST-23 complex and serogroup W135 accounted for 76% of ST-22 complex isolates. A carriage study comparing disease and carriage isolates from three European countries (Norway, Greece and the Czech Republic) from 1991–2000 showed a more diverse carriage population which had relatively low representation of disease-associated hyperinvasive lineages (Yazdankhah *et al.* , 2004). The ST-11 complex and serogroup C were found to be positively associated with disease while ST-23 complex was positively associated with carriage. There was also shown to be significant differentiation in the distribution of meningococcal genotypes in disease and carrier isolates among countries.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 1.26: Clonal complexes by serogroup from European disease isolates 1999–2004. From (Trotter *et al.*, 2007).

1.8 Aims and outline of study

The aim of the study was to investigate the temporal and geographical distribution and structuring of invasive meningococcal genotypes within Europe over a three year period. The isolates chosen were representative of invasive disease in each of the participating 18 countries for the time period and the data describe the recent circulating disease-associated strains in the continent. Exploiting high throughput sequencing technology, it was possible to carry out sequence typing of the PorA and FetA antigens and of the MLST house-keeping genes for a large European dataset from these countries.

The main question was to ascertain the distribution of meningococcal lineages, both clonal complex and antigenic types, associated with disease in Europe. Within this, the dynamics in distribution of lineages over the three years across Europe as a whole were identified and quantified. Also, the differences in predominating types and levels of diversity among countries, association of particular types with particular countries, and presence of population structuring in Europe in terms of genotypes were quantified. Identification of the emergence of (new) types and the disappearance of types was gauged. Within countries themselves, the presence of any trends or changes over time and if any major public health interventions such as vaccine strategies may have affected the distribution of types was assessed. Relationships between the antigenic types themselves and also with clonal complexes were investigated for any correlation. A major use of data such as that obtained in this study is to apply it in the design of vaccine recipes.

Chapter 1 presents a general introduction and background to the study and sets out the aims of the thesis. Chapter 2 details the experimental methodology and techniques used during the study. Chapter 3 explores the geographic and temporal distribution of meningococcal invasive genotypes in Europe from 2000–2002. Chapter 4 assesses the geographic and temporal distribution and strain structuring of the two vaccine candidate antigens PorA and FetA. Chapter 5 outlines the rationale for the design of a novel PorA/FetA vaccine using data described in the previous chapters to formulate a potential vaccine recipe with assessment of the theoretical impact of such a formulation. Chapter 6 investigates a relatively newly identified genome-derived vaccine candidate antigen fHbp and its diversity in a representative collection of

meningococcal isolates and examines the recombination and selection acting on it. An improved understanding of the dispersion of the genotypes associated with invasive disease in Europe will greatly help disease surveillance and prevention and also the advancement of vaccine design.

CHAPTER 2: MATERIALS AND METHODS

2.1 Materials

Unless otherwise stated all chemicals and reagents were obtained from Sigma (Molecular Biology grade) or BDH (Analar grade). All water used was purified using a Milli-Q water purification system (Millipore). Microcentrifuge tubes and filtered micropipette tips were from Axygen; all other micropipette tips were from Alpha Laboratories. Micropipettes were from Gilson. Electronic micropipettes were from Biohit. Thermocyclers used were from MJ Research Inc. All glassware used was from Fisher Scientific.

2.2 Meningococcal isolates

2.2.1 European meningococcal disease isolates

The European Meningococcal Multilocus Sequence Typing Centre (EMMC) received 2672 isolates (killed cell suspensions or purified DNA) from the 18 participant countries of the EU-MenNet project (Table 2.1). Data were also submitted for a further 1507 isolates from a number of participants. The total representative collection (4179) of isolates from each of the three years (2000, 2001, 2002) were chosen on the basis of an algorithm devised by the EU-MenNet management committee. For national reference laboratories receiving 80 meningococcal disease isolates per year or fewer, *all* isolates were submitted to the EMMC. For laboratories receiving more than 80 isolates, every third isolate received was submitted, with the exception of the England and Wales Meningococcal Reference Unit, which received more than a thousand cases per year and submitted every tenth isolate.

Table 2.1: Summary of isolates received from 18 participant countries of EU-MenNet.

sender	affiliation	boilate	DNA	data	total
Lene Berthelsen	Statens Serum Institut, Copenhagen, Denmark	175			175
Sigrid Heuberger	National Reference Centre for Meningococci, Graz, Austria	188			188
Hjördís Harðardóttir	Dept. of Microbiology, Landspitali University Hospital, Reykjavik, Iceland	17			17
Paola Stefanelli	Istituto Superiore di Sanità, Italy	81			81
Maija Toropainen	National Public Health Institute, Helsinki, Finland	139			139
Manuela Caniça	Antibiotic Resistance Unit, National Institute of Health Dr. Ricardo Jorge, Lisbon, Portugal	36			36
Ingrid Ehrhard & Heike Claus	Institut für Hygiene und Mikrobiologie, Würzburg, Germany	537			537
Françoise Carion	Meningococcal Reference Laboratory, Scientific Institute of Public Health, Brussels, Belgium	283			283
Mathew Diggle	Scottish Meningococcus and Pneumococcus Reference Laboratory, Glasgow, UK			236	236
Muhamed-Kheir Taha	National Reference Centre for Meningococci, Pasteur Institute, Paris, France	20		360	380
Georgina Tzanakaki	National Meningococcal Reference Laboratory, National School of Public Health, Athens, Greece			119	119
Paula Kriz	National Institute of Public Health, Prague, Czech Republic	89		105	194
Julio Vazquez	Meningococcal Reference Laboratory, Madrid, Spain	393		31	424
Arie van der Ende	Academic Medical Centre, Department of Medical Microbiology, Amsterdam, Netherlands		568	72	640
Karen Murphy	Irish Meningococcal and Meningitis Reference Laboratory, Dublin, Ireland	113			113
Dominique Caugant	Norwegian Institute of Public Health, Oslo, Norway			188	188
Susanne Jacobsson	Dept. of Clinical Microbiology, Örebro University Hospital, Örebro, Sweden			93	93
Steve Gray	Meningococcal Reference Unit, UK	33		303	336
total		2104	568	1507	4179

2.2.2 Globally representative set of 107 meningococcal isolates

Purified DNA from a globally representative sample of 107 meningococcal isolates (Maiden *et al.*, 1998) was used in the study to ascertain the diversity of the *fHbp* gene (see table 6.1). The isolates were obtained from both patients and carriers and represent all known invasive genotypes and serogroups (details available at http://pubmlst.org/neisseria/). They were originally collected for the development and assessment of multilocus sequence typing (MLST) (Maiden *et al.*, 1998) and contain 75 meningococci from major lineages (60 invasive, 4 non-invasive and 11 unspecified) and 32 isolates defined as 'other' by MLEE (19 invasive and 13 non-invasive). The collection contains approximately ten isolates from each of the seven recognised hyper-virulent MLST clonal complexes: ST-1 complex, ST-5 complex, ST-4 complex, ST-11 complex, ST-32 complex, ST-8 complex, ST-41/44 complex.

2.2.3 GenBank fHbp peptide sequence.

A previously published *N. gonorrhoeae* protein sequence: EEH61327 from GenBank was also used as part of the analysis.

2.3 Microbiological methods

At the various participant laboratories bacterial isolates were grown on *N. meningitidis* specific media such as Mueller-Hinton agar and grown overnight (for at least 12 hours) at 37°C in a 5% CO_2 atmosphere. Owing to the fact that isolates came from various source laboratories (Table 2.1), there may be some slight variation in microbiological protocols used.

2.4 Molecular biology techniques

2.4.1 DNA preparation

Genomic DNA was extracted from each isolate by means of preparing killed cell suspensions (see Appendix for protocol). Thick bacterial suspensions were made by using sterile swabs/loops to sweep colonies from culture plates and mix in 1.5–2.0 ml screw-capped and clearly-labelled Eppendorf tubes containing 0.5 ml phosphate buffered saline. Immediately, the tubes were placed in a boiling water bath and left for 20 minutes. Boiling cells releases the bacterial DNA while simultaneously inactivating nucleases which could destroy it. Sealing film was wrapped around the lids of the tubes and the samples were sent to the EMMC at room temperature *via* normal post. The majority of isolate genomic DNA from Arie van der Ende at the Academic Medical Centre, Department of Medical Microbiology, Amsterdam, the Netherlands was purified and sent by post. The isolates from the 107 collection were cultured as previously described (Jolley *et al.* , 2000) with genomic DNA extracted using a Promega Wizard Genomic DNA purification kit (Promega, Southampton, United Kingdom) according to the manufacturer's instructions.

2.4.2 Polymerase Chain Reaction (PCR) amplification

PCR amplifications of the internal fragments of the seven MLST housekeeping genes *abcZ* (putative ABC transporter), *adk* (adenylate kinase), *aroE* (shikimate dehydrogenase), *fumC* (fumarate hydratase), *gdh* (glucose-6-phosphate dehydrogenase), *pdhC* (pyruvate dehydrogenase subunit) and *pgm* (phosphoglucomutase) (Figure 2.1, Table 2.2) were carried out for all isolates (boiled cell suspensions or purified DNA). Also, amplifications of the variable regions (VR1, VR2) of the *porA* gene, the VR of the *fetA* gene and the *fHbp* gene were carried out (Figure 2.1, Tables 2.3, 2.4 and 2.5). Details of all oligonucleotide primers used are in Tables 2.2, 2.3, 2.4 and 2.5. PCR reactions were set up using 48-well plates with adhesive film to prevent evaporation and spillage (Abgene). For each 49 µl PCR the following reaction components were used: 29.75 µl sterile Milli-Q water, 10 µl 5× solution Q (Qiagen), 5 µl 10× buffer (Qiagen), 1 µl of each primer (10 µM stock solution, Table 2.2 and 2.3), 1 µl premixed 10 mM

dNTP (2.5 mM each dATP, dCTP, dGTP, dTTP, Applied Biosystems), 0.25 µl Taq polymerase (5 units/µl, Qiagen) and 1 µl template DNA (approximately 50 ng/µl). One well was set up as a negative control with mastermix but no template DNA. MLST cycling conditions were as follows: initial denaturation at 95°C for 3 minutes followed by 35 cycles of denaturation (95°C for 30 secs), annealing (55°C for 30 secs), and extension (72°C for 1 min). There was then a final extension step of 72°C for 10 minutes.

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 2.1: Circular representation of the *N. meningitidis* Z2471 genome (Parkhill *et al.*, 2000) showing the locations of the seven housekeeping loci used for multilocus sequence typing. The origin of replication is marked (origin) and the reference point for nucleotide positions (0). From: (Maiden, 2007).

PorA cycling conditions were as follows: initial denaturation at 94°C for 3 minutes followed by 30 cycles of denaturation (94°C for 1 min), annealing (60°C for 1 min), and extension (72°C for 2 min). There was then a final extension step of 72°C for 3 minutes. FetA cycling conditions were as follows: initial denaturation at 94°C for 2 minutes followed by 40 cycles of denaturation (94°C for 1 min), annealing (55°C for 1 min), and extension (72°C for 2 min 30 secs). There was then a final extension step of 72°C for 5 minutes. For *fHbp* (Figure 2.3), the 50 µl PCR amplification reactions contained: 5 µl 10x PCR Buffer, 4 µl dNTP mix, 1 µl (0.4 µM) of each primer, 0.25 µl

Taq polymerase and 38.75 µl dH_2O. The cycling conditions were as follows: initial denaturation at 95°C for 5 minutes followed by 33 cycles of denaturation (95°C for 50 secs), annealing (59°C for 50 secs), and extension (72°C for 50 sec). There was then a final extension step of 72°C for 7 minutes.

Table 2.2: Oligonucleotide primers used for the PCR amplification and sequencing of MLST loci.

function	primer	gene	nucleotide sequence 5' to 3'
amplification	abcZ-P1C	*abcZ*	TGTTCCGCTTCGACTGCCAAC
	abcZ-P2C	*abcZ*	TCCCCGTCGTAAAAAACAATC
	adk-P1B	*adk*	CCAAGCCGTGTAGAATCGTAAACC
	adk-P2B	*adk*	TGCCCAATGCGCCCAATAC
	aroE-P1B	*aroE*	TTTGAAACAGGCGGTTGCGG
	aroE-P2B	*aroE*	CAGCGGTAATCCAGTGCGAC
	fumC-P1B	*fumC*	TCCCCGCCGTAAAAGCCCTG
	fumC-P2B	*fumC*	GCCCGTCAGCAAGCCCAAC
	gdh-P1B	*gdh*	CTGCCCCCGGGGTTTTCATCT
	gdh-P2B	*gdh*	TGTTGCGCGTTATTTCAAAGAAGG
	pdhC-PB1	*pdhC*	CCGGCCGTACGACGCTGAAC
	pdhC-PB2	*pdhC*	GATGTCGGAATGGGGCAAACA
	pgm-PB1	*pgm*	CGCCTCAAAACGCAACACCAG
	pgm-PB2	*pgm*	ACGGCACTTTCCCCAACCACC
sequencing	abcZ-S1A	*abcZ*	AATCGTTTATGTACCGCAGR
	abcZ-S2	*abcZ*	GAGAACGAGCCGGGATAGGA
	adk-S1A	*adk*	AGGCWGGCACGCCCTTGG
	adk-S2	*adk*	CAATACTTCGGCTTTCACGG
	aroE-S1A	*aroE*	GCGGTCAAYACGCTGRTK
	aroE-S2	*aroE*	ATGATGTTGCCGTACACATA
	fumC-S1	*fumC*	TCCGGCTTGCCGTTTGTCAG
	fumC-S2	*fumC*	TTGTAGGCGGTTTTGGCGAC
	gdh-S3	*gdh*	CCTTGGCAAAGAAAGCCTGC
	gdh-S4C	*gdh*	RCGCACGGATTCATRYGG
	pdhC-S1	*pdhC*	TCTACTACATCACCCTGATG

pdhC-S2	*pdhC*	ATCGGCTTTGATGCCGTATTT
pgm-S1	*pgm*	CGGCGATGCCGACCGCTTGG
pgm-S2A	*pgm*	GGTGATGATTTCGGTYGCRCC

oligonucleotide primers previously published (Jolley *et al.* , 2002; Maiden *et al.* , 1998) http://pubmlst.org/neisseria/mlst-info/nmeningitidis/primers.shtml

Table 2.3: Oligonucleotide primers used for the sequencing of the Variable Regions (VR1, VR2) of the *porA* gene.

primer	gene and location	nucleotide sequence 5' to 3'	function
210	*porA* VR1 loop I	ATGCGAAAAAAACTTACCGCCCTC	amplification and sequencing
211	*porA* VR2 loop IV	AATGAAGGCAAGCCGTCAAAAACA	amplification
8L	*porA* VR1 loop I	GGAGAATCGTAGCGTACGGA	sequencing
103U	*porA* VR2 loop IV	GAGCAAGACGTATCCGTT	sequencing
122L	*porA* VR2 loop IV	GGCGAGATTCAAGCCGCC	sequencing

oligonucleotide primers previously published (Feavers & Maiden, 1998; Suker *et al.*, 1994)

The figure originally presented here cannot be made freely available via ORA for copyright reasons.

Figure 2.2: Locations of oligonucleotide primers used for the PCR amplification and sequencing of the *porA* gene. This figure depicts the location and orientation of primers (Table 2.3) used to amplify and sequence the *porA* gene. Arrows indicate orientation of the primers. The variable regions VR1 and VR2 are located in the white boxes. The scale bar below the diagram shows the approximate distances in base pairs (bp). Adapted from (Russell, 2001).

Table 2.4: Oligonucleotide primers used for the sequencing of the VR of the *fetA* gene.

primer	nucleotide sequence 5' to 3'	function
S1	CGGCGCAAGCGTATTCGG	amplification
S8	CGCGCCCAATTCGTAACCGTG	amplification
S12	TTCAACTTCGACAGCCGCCTT	sequencing
S15	TTGCAGCGCGTCRTACAGGCG	sequencing

oligonucleotide primers previously published (Thompson *et al.*, 2003)

Table 2.5: Oligonucleotide primers used for the sequencing of the *fHbp* gene.

primer	nucleotide sequence 5' to 3'	function
Long 5UNI 2086	CTATTCTGCGTATGACTAGGAG	amplification
3UNI	GTCCGAACGGTAAATTATCGTG	amplification
5'2086forseq	TATGACTAGGAGCAAACCTG	sequencing (Subfamily A)
3'2086forseq	TACTGTTTGCCGGCGATG	sequencing (Subfamily A)
2086interforseq5’LA	AGCTCATTACCTTGGAGAGCGGA	sequencing (Subfamily A)
5'2086forseq	TATGACTAGGAGCAAACCTG	sequencing (Subfamily B)
2086seq3’BLA	TTCGGACGGCATTTTCACAATGG	sequencing (Subfamily B)
2086interforseq5’LA	AGCTCATTACCTTGGAGAGCGGA	sequencing (Subfamily B)
2086seqBinternal	GGCGATTTCAAATGTTCGATTT	sequencing (Subfamily B)

oligonucleotide primers previously published (Fletcher *et al.*, 2004; Murphy *et al.*, 2009)

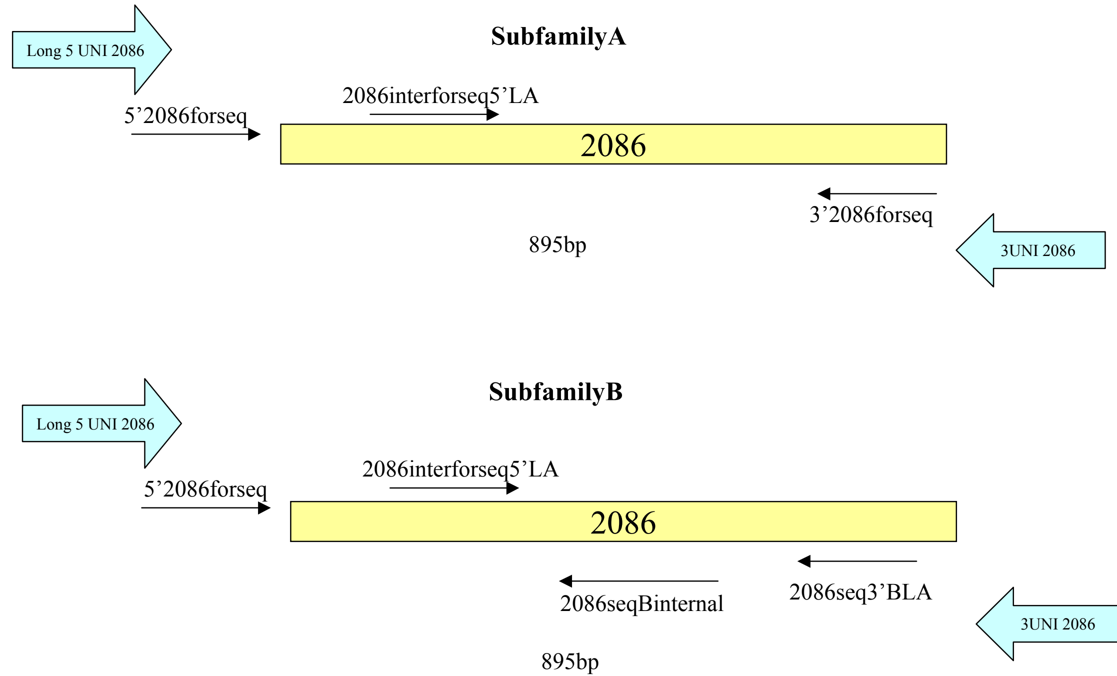


Figure 2.3: Locations of oligonucleotide primers used for the PCR amplification and sequencing of the *fHbp* gene. This figure depicts the location and orientation of primers (given in Table 2.5) used to amplify and sequence the *fHbp* gene. Arrows indicate orientation of the primers. Oligonucleotide primers specific for each of the subfamilies A and B were used to amplify internal fragments of the purified amplified gene products (~900 bp).

2.4.3 Agarose gel electrophoresis

Amplification products were analysed using agarose gel electrophoresis. Gels were made by heat dissolving 1% (w/v) agarose (Severn Biotech) in 1× TBE buffer (0.089 M Tris-borate, 0.089 M Boric acid, 0.002 M EDTA). Ethidium bromide (0.5 µg/ml) was added to the cooled gel mixture to visualise the DNA following electrophoresis. The gel was cast and set at room temperature and submerged in 1× TBE Buffer in a gel tank (Scotlab). 5 µl PCR products were mixed with 1 µl of loading dye (6× concentration, 0.25% bromophenol blue, 40% sucrose in TBE buffer). Following electrophoresis at a current of 100–140 V for 15–20 minutes, gels were placed in an Ultra-violet transilluminator (UVIpro) to enable visualisation of the DNA.

2.4.4 Purification of PCR products

PCR products were precipitated and separated from unincorporated dNTPs and oligonucleotide primers by polyethylene glycol (PEG) purification. 60 µl of 20% PEG_{8000}/2.5 M NaCl was added to each 49 µl reaction and mixed using a vortex mixer. Plates were incubated at room temperature for 30 minutes before being centrifuged for one hour at 2750× *g* (4°C) to pellet the PCR products. Following this, plates were gently inverted onto tissue paper, then placed onto Whatman 3 MM chromatography paper and spun at 500× *g* for one minute to remove any residual liquid from the wells. To wash the pellets, 150 µl 70% (v/v) ice-cold ethanol was added to each well and plates were centrifuged for five minutes at 2750× *g*. Ethanol was removed by inversion of plates onto tissue and again inverting onto Whatman paper and spinning for one minute at 500× *g*. The ethanol washing procedure was then repeated. Plates could then be stored at -20˚C. 20 µl distilled water was added to each well just before the sequencing reaction.

2.4.5 Automated nucleotide sequence determination

The chain termination method of DNA sequencing was developed by Sanger (Sanger *et al.*, 1977). Subsequent adaptations (Smith *et al.* , 1986) have increased ease of automation and safety by use of fluorescently labelled dideoxynucleotides (ddNTPS) instead of radioactive reagents. Sequencing reactions (modified from the PE ABI BigDye Ready Reaction Termination Mix protocol) were set up per well as follows in 96-well sequencing plates (Abgene): 0.5 µl Big Dye Ready Reaction Termination Mix (PE-Biosystems), 1.75 µl 5× solution (400 mM Trizma-HCl pH 9.0, 10 mM $MgCl_2$), 4 µl forward or reverse primer (1:15 dilution of 10 µM solution, Table 2.2 and 2.3), 1.75 µl distilled H_2O and 2 µl template DNA. Plates were then sealed with adhesive film and placed on thermocyclers with the following cycling conditions: 30 cycles of 96°C for 10 sec, 50°C for 5 sec and 60°C for 4 min.

2.4.6 Purification of sequencing reactions

The following procedure was carried out on sequencing reaction samples to remove excess dye terminator and ready them for DNA sequencing. Following cycling, plates were briefly centrifuged to ensure all liquid was at the bottom of the wells. The adhesive film was carefully peeled back and 30 µl distilled water was added to each well to help reduce dye-blobs in the sequence. 52 µl of a 7 ml absolute ethanol/280 µl 3 M sodium acetate mix was added to each well. After the adhesive lid was placed back on, the plates were vortexed briefly to ensure mixing, centrifuged briefly at 500× *g* and then incubated at room temperature for 45 minutes. Following this incubation, plates were spun at 2750× *g* (4°C) for one hour. Immediately after spinning, the plates were gently inverted onto tissue and then spun inverted onto Whatman paper at 500× *g* for one minute to discard the supernatant. DNA pellets were washed by adding 150 µl of ice-cold 70% ethanol and spinning at 2750× *g* for a further ten minutes. The supernatants were discarded as before. The pellets were then dry and the plates could be sealed and stored at -20°C or run immediately on the sequencer after they were resuspended in formamide loading buffer. Separation of the labelled extension products was carried out on a 3700 or 3730 capillary DNA Analyzer (Applied Biosystems) at the Department of Zoology Sequencing Facility, University of Oxford.

2.5 Data analysis

2.5.1 Assembly and analysis of automated nucleotide sequence data

Assembly and editing of MLST, PorA, FetA and fHbp sequence data generated was carried out using Sequence Typing Analysis and Retrieval System (STARS) (http://www.cbrg.ox.ac.uk/~mchan/stars/) and the STADEN suite of software using programs PREGAP 4 Version 1.3 and GAP Version 4.7 (Staden, 1996). For each isolate (for MLST), sequences for each of the seven loci were assigned allele numbers upon interrogation of the *Neisseria* MLST database (http://pubmlst.org/neisseria/). This gave a string of seven integers — the allelic profile — which was then assigned an arbitrary ST number based on the order of submission of the allelic profile to the database.

Previously devised nomenclature systems for the PorA and FetA VRs (Russell *et al.* , 2004; Thompson, 2001) were used to assign alleles using the PorA and FetA databases http://neisseria.org/nm/typing/. The isolate data were stored in a proprietary database based on MLSTDBNET software (Jolley *et al.* , 2004). In the case of fHbp, reformatted nucleotide sequences were visualised, aligned and translated manually using SEQLAB, part of the GCG Wisconsin Package suite of programs (Womble, 2000) (Version 10.3 for Unix (Accelrys Inc., San Diego, CA)). Each nucleotide sequence was grouped with its corresponding translated peptide sequence to maintain codon integrity and alignments based on amino acid sequence similarity. Alignment gaps were minimised, in order to preserve maximum positional homology and structural and biochemical properties. A web-based front end to the NRDB program (written by Warren Gish, Washington University) was used to compare the nucleotide and amino acid sequences to find ones that were identical to each other (http://pubmlst.org/perl/mlstanalyse/mlstanalyse.pl?site=pubmlst&page=nrdb&referer=pubmlst.org).

2.5.2 fHbp Database

Unique nucleotide and peptide sequences were arbitrarily assigned allele numbers in order of discovery. A database was established containing these allele sequences obtained as part of this study, from direct submissions from collaborators or by interrogation of the GenBank database. AGDBNET antigen sequence software for web-based bacterial typing was used to do this (http://pubmlst.org/software/database/agdbnet/). It allows simultaneous BLAST querying of multiple loci using either nucleotide or peptide sequences (Jolley & Maiden, 2006).

2.5.3 Genetic analyses and software programs used

2.5.3.1 Analysis of selection

Nucleotide substitutions in a gene may either be synonymous ('silent' – encoded amino acid unchanged) or nonsynonymous (encoded amino acid changed). Most nonsynonymous changes are purged by purifying selection. However, if they offer a selective advantage, Darwinian natural selection may retain them. Investigation of the levels of synonymous and nonsynonymous nucleotide substitutions that take place may give an indication of the levels and type of selection occurring. The Nei and Gojobori (Nei & Gojobori, 1986) method of estimating nucleotide substitution was used to do this. This method computes the number of synonymous and nonsynonymous substitutions and the numbers of potential synonymous and nonsynonymous sites. Firstly, the total number of synonymous (S) and nonsynonymous (N) sites in a sequence with n nucleotides (where $S + N = n$) is enumerated. The next step is to calculate the number of synonymous (S_d) and nonsynonymous (N_d) changes (codon-by-codon) between each pair of aligned sequences. When comparing two sequences, all possible pathways and mutational steps by which differences between the two are generated are considered. The total number of S_d and N_d can thus be summed across all codons. Therefore, the proportion of synonymous and nonsynonymous sites can be calculated as:

$$p_S = S_d/S$$

and

$$p_N = N_d/N$$

Using these proportions, the number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous substitutions per nonsynonymous site (d_N) can be calculated correcting for multiple substitutions by the Jukes-Cantor correction method:

$$d_S = \frac{-3\ln(1 - \frac{4}{3} p_S)}{4}$$

and

$$d_N = \frac{-3\ln\left(1 - \frac{4}{3} p_N\right)}{4}$$

Thus, the ratio nonsynonymous to synonymous substitutions is:

$$d_N/d_S$$

A d_N/d_S ratio of one indicates neutral evolution of the sequences, *i.e.* both synonymous and nonsynonymous substitutions do not affect fitness (Jolley & Urwin, 2001). A ratio value of greater than one implies that positive selection is acting to fix nonsynonymous changes. A value of less than one indicates that deleterious nonsynonymous substitutions are under purifying selection and are removed.

2.5.3.2 Analysis of recombination

The maximum Chi (χ^2) squared test (Maynard Smith, 1992), uses the distribution of polymorphic sites to locate potential points of recombination between two sequences by comparing it with those that would be expected to occur by chance (Jolley & Urwin, 2001). To test the significance of a putative recombination point, trials are run where random sequence pairs of the same length and the same number of polymorphic sites are created. The level of significance is established as the proportion of trial pairs with greater maximum χ^2 values than the observed data.

The index of association (I_A) (Maynard Smith *et al.* , 1993) is a statistical test which determines the level of linkage equilibrium (the random association of alleles at various loci) in sequence data. In the presence of frequent recombination, sequences typically display linkage equilibrium and the expected value of I_A is 0. Conversely, in the absence of recombination there is non-random association of alleles and therefore an expected I_A value that differs significantly from 0. The index of association is calculated as follows:

$$I_A = V_O/V_E - 1$$

V_O = Observed variance of K (number of loci the two individuals differ, *i.e.* “distance”)

V_E = Expected variance of K

The 'standardised' index of association accounts for the fact that the I_A statistic scales with l-1 where l is the number of loci analysed (Hudson, 1994). Therefore, the standardised index was used in this study and is calculated as:

$$I^S_A = \frac{1}{l-1}(V_O/V_E - 1)$$

2.5.3.3 Analysis of population structuring

F-statistics or fixation indices were developed by Wright to describe the amount of heterozygosity in a population (Wright, 1950; Wright, 1965). Three hierarchical *F*-statistics are used to do this: F_{IS} (inbreeding coefficient – the mean reduction in heterozygosity of an individual due to non-random mating within a subpopulation), F_{IT} (overall fixation index – mean reduction of heterozygosity in an individual relative to the total population) and the fixation index F_{ST}. The relationship between the three F statistics is:

$$(1 - F_{\mathrm{IT}}) = (1 - F_{\mathrm{IS}})\ (1 - F_{\mathrm{ST}})$$

F_{ST} is a measure of population differentiation and gene flow. Subdivision in a population causes a decline in heterozygosity, *i.e.* there is less heterozygosity than there would be if the population was undivided. The F_{ST} statistic is a measure of the difference between the mean heterozygosity among the subdivisions in a population, and the potential frequency of heterozygotes if all members of the population mixed freely (gene flow) and non-assortatively. A common definition is given as (Hudson *et al.*, 1992):

$$F_{\mathrm{ST}} = \frac{\Pi_{\mathrm{between}} - \Pi_{\mathrm{within}}}{\Pi_{\mathrm{between}}}$$

where Π_{between} is the average number of pairwise differences between two individuals sampled from different populations and Π_{within} is the average number of pairwise

differences between two individuals sampled from the same population. The F_{ST} value can range from zero meaning no differentiation and presence of gene flow, to one meaning complete differentiation and presence of structuring in the population.

Analysis of molecular variance (AMOVA) is an analogue of the statistical technique analysis of variance (ANOVA) which is used to test the hypothesis that the means among two or more groups are equal, under the assumption that the sampled populations are normally distributed. AMOVA incorporates information on molecular data and about the mutational pathway likely to connect sequences rather than just the number of differences between them (Excoffier *et a l.*, 1992). It uses a statistic for nucleotide diversity analogous to the F_{ST} for allelic diversity: the Φ_{ST}. It is calculated as follows:

$$\Phi_{ST} = \frac{\pi_T - \overline{\pi_S}}{\pi_T}$$

π_T = nucleotide diversity in the total population

$\overline{\pi_S}$ = average nucleotide diversity in the subpopulations

2.5.3.4 Software programs

2.5.3.4.1 Sequence Type Analysis and Recombinational Tests – START

START was specifically designed for the analysis of MLST allele sequence and profile data (Jolley *et al.* , 2001) (available for free download at http://pubmlst.org/). START2 Version 0.5.14 beta was used to carry out data summary analysis such as allele and profile frequencies, identification and quantification of polymorphic sites, codon usage and GC content. It was also used for tests of recombination such as Sawyer's Runs test, maximum χ^2 test, the homoplasy test and the 'standardised' index of association. The test for selection used was the ratio of nonsynonymous to synonymous nucleotide substitutions (d_N/d_S ratio). The program was also used to assess for selection in the *fHbp* gene.

2.5.3.4.2 Molecular Evolutionary Genetics Analysis – MEGA

The MEGA 3.1 program which is freely available for download at http://www.megasoftware.net/ (Kumar *et al.* , 2004) was used to calculate the overall mean distances as well as the within and between group distances for sequences using the Kimura 2-parameter model for nucleotide sequences and *p*-distances for amino acid sequences. It was also used to produce distance matrix-based Neighbour-Joining (NJ) trees. NJ (Saitou & Nei, 1987) is a distance based minimum evolution method which uses successive clustering to estimate the minimum length of the tree which is seen as the best estimate of that tree. The reliability of the inferred trees was assessed by the bootstrap test with 2000 replications. MEGA 3.1 implements Felsenstein's bootstrap test evaluated using Efron's bootstrap resampling method. In the bootstrap test of phylogeny, a matrix of *m* sequences x *n* (nucleotides/peptides) are sampled with replacement (bootstrapping). These new sequences are reconstructed into a tree using the previously used phylogenetic method and the topology is compared to the original tree. This procedure is repeated 2000 times and the percentage of times a particular interior branch is the same between the original tree and the bootstrap tree is given. Bootstrapping is a means of assessing confidence in a particular phylogeny and values are interpreted as the probability of interior branches being "correct" (generally 95% or higher).

2.5.3.4.3 PHYML

The maximum likelihood (ML) method is used for the inference of phylogeny. It evaluates a hypothesis about evolutionary history in terms of the probability that the proposed model and the hypothesised history would give rise to the observed data set. The notion is that a history with a higher probability of reaching the observed state is preferred to a history with a lower probability. The method searches for the tree with the highest probability or likelihood. The PHYML program (Guindon & Gascuel, 2003) which is freely available at http://www.phylogeny.fr (Dereeper *et al.*, 2008) was used to construct the phylogeny of fHbp nucleotide sequences using the ML method. Support for inferred branches was assessed by the bootstrap test with 100 replications. MEGA 3.1 was used to display and annotate the phylogeny.

2.5.3.4.4 SPLITSTREE

The SPLITSTREE program (Huson, 1998) (freely available for download at http://www.splitstree.org/) uses the split decomposition method (Bandelt & Dress, 1992) to visualise and analyse distance data from sequences. Split decomposition is applicable to data from organisms where horizontal gene transfer is an important evolutionary force and where ordinary tree-drawing methods may force a linear bifurcating tree-like structure on the data when it may not be always appropriate. Analysis depicts all the shortest pathways linking sequences, including those that produce an interconnected network. SPLITSTREE Version 4.6 was used to visualise relationships among fHbp sequences and also to produce UPGMA (unweighted pair group method with arithmetic mean) and split decomposition trees from distance matrices of F_{ST} values produced by the ARLEQUIN program. Splits graphs were drawn to scale using the Hamming method. Fit index values were calculated for each tree. Trees with values above 90% are considered robust, values above 80% give reasonable confidence and values below 60–70% should be treated with caution (Moulton, 2003).

2.5.3.4.5 ARLEQUIN

The population genetics program ARLEQUIN Version 3.1 (Excoffier *et al.*, 2005) (freely available for download at http://lgb.unige.ch/arlequin/) was used to assess the levels of geographic and temporal structuring in the European meningococcal disease population. Significant genetic differentiation among groups of isolates (years/populations) and the contribution of temporal and geographic factors to genetic differentiation were assessed by AMOVA (Excoffier *et al.*, 1992). F_{ST} values and corresponding permutation tests to test for statistical significance were implemented in ARLEQUIN using concatenated MLST sequences, allele sequences, allele designations, allelic profiles and STs. F_{ST} values of pairwise comparisons of concatenated MLST, PorA and FetA sequences of the participating EU-MenNet countries were also calculated. Corresponding statistical tests were run alongside the F_{ST} test giving a p-value to determine if the value obtained was significantly different from the null hypothesis of no difference between populations.

2.5.3.4.6 ClonalFrame

The software package ClonalFrame Version 1.1, which implements a statistical model for inferring bacterial microevolution, was used for phylogenetic analysis of the 107 fHbp nucleotide sequences and to identify regions likely to have undergone homologous recombination (Didelot & Falush, 2007). ClonalFrame performs inference in a Bayesian framework which assumes a standard neutral coalescent model whereby the bacteria in the sample come from a constant-sized population where each bacterium is equally likely to reproduce, irrespective of its previous history. The key assumption is that recombination events introduce a constant rate of substitutions to a contiguous region of sequence. Six independent runs, each with 250,000 iterations, 100,000 burn-in iterations and with every hundredth tree sampled, were used to derive a 75% majority-rule consensus tree. ClonalFrame tree output was imported and further annotated in MEGA 3.1.

2.5.3.4.7 OmegaMap

OmegaMap is a program which uses a Bayesian method to estimate the selection parameter ω (d_N/d_S) and the recombination rate ρ from gene sequences by use of reversible jump Monte Carlo Markov Chain (Wilson & McVean, 2006). The program was used to detect selection acting on the *fHbp* gene by inferring the posterior distributions of omega and rho along the gene. The means of the posterior distributions were used as point estimates for omega and rho. The per-site posterior probability of positive selection was also used to summarise the posterior distribution of omega. Three independent runs each with 1,000,000 iterations and a thinning interval of 100 were compared to assess convergence and combined. Output from the OmegaMap runs was used to visualise possible selection acting on the sequence by means of fireplots and graphs indicating the posterior probability of positive selection along the sequence. Fireplots visualise the posterior on log(ω) or ω along the sequence using a colour gradient where a higher posterior density is represented by more intense colour (closer to white), and lower posterior density is represented by less intense colour (closer to red). These plots were produced using the R program version 2.7.1 (http://www.r-project.org/). The point estimate of ω, was used to colour a three-dimensional pdb file of the solution structure of a complex between a subfamily B/variant 1 GNA1870/fHbp

protein and a region of the fH protein (Schneider *et al.*, 2009) (http://www.rcsb.org/pdb/home/home.do Protein Data Bank code 2W81).

2.5.3.4.8 SNAP

The Synonymous Nonsynonymous Analysis Program (SNAP) (http://www.hiv.lanl.gov/) was used to analyse pairwise synonymous and nonsynonymous differences in the fHbp nucleotide sequences according to the method of Nei and Gojobori (Nei & Gojobori, 1986). Output from the program was imported into Microsoft Excel to produce graphs indicating levels of synonymous and non-synonymous substitutions for each codon of the *fHbp* gene.

2.5.3.4.9 R program

The R program is a language and environment for statistical computing and graphics which is freely downloadable from http://www.r-project.org/ (R Core Team & Team, 2008). It was used for calculations of χ^2 and Fisher's exact tests, odds ratios and their associated 95% confidence intervals. It was also used for analysis and visualisation of OMEGAMAP run outputs.

2.5.3.8.10 STATA

Intercooled STATA 10.0 for Windows (Statacorp, 2007) was used to carry out logistic regression analyses.

2.5.3.8.11 eBURST

eBURST (Based Upon Related Sequence Types) is an algorithm related to BURST which attempts to show the relationships amongst a dataset of MLST profiles and displays them graphically (Feil *et al.*, 2004). STs are grouped according to the user-defined setting of the number of shared loci and a putative central founder genotype is found. Members of the same clonal complex should share identical alleles at six of seven loci with at least one other member of the group. This is the default and most stringent

setting for the program and the one used to illustrate the distribution of clonal complexes with the EU-MenNet disease isolate collection.

2.5.4 Statistical analysis

2.5.4.1 Pearson's Chi-squared and Fisher's exact tests

The χ^2 test for independence examines whether the distribution of categorical variables in a contingency table (2 x 2 or row by column tables) are independent of each other or not and gives a rough estimate of confidence in this (p-value). That is, it examines whether the distribution of individuals among the categories of one variable is independent of their distribution among the categories of the other (Kirkwood & Sterne, 2003). The null hypothesis is of no association between the two distributions. The value of χ^2 is calculated by the formula:

$$\chi^2 = \sum \frac{(Observed - Expected)^2}{Expected}$$

Observed: Observed frequency in a particular cell of contingency table
Expected: Expected frequency in a particular cell of contingency table
Null hypothesis (H_0): observed group mean - expected group mean equals zero (there is no difference between the two groups)
Alternative hypothesis (H_A): observed group mean - expected group mean does not equal zero (there is a difference between the two groups)

The Fisher's exact test is similar to the χ^2 test and is used when use of the χ^2 test is inappropriate: for small sample sizes (<40), if any value in the 2 x 2 cross table is less than ten and/or if the values across the cells of the table are very uneven. It calculates an exact probability value for the relationship between the variables. Scripts for the Pearson's χ^2 and Fisher's exact tests were written and run in the R system for statistical computation and graphics version 2.7.1 (R Core Team & Team, 2008). The tests were employed to compare rates as appropriate.

2.5.4.2 Odds ratios

An odds ratio (OR) is the ratio of the odds of an outcome/occurrence of an event in one group compared to the odds of an outcome/event in another group. The odds of an event happening is the probability that the event will happen divided by the probability that the event will not happen. Therefore:

$$\text{OR} = \frac{\text{Odds of event in group 1}}{\text{Odds of event in group 2}} \qquad \textit{(Equation 1)}$$

An OR of one indicates that the event is equally likely to occur in each group. An OR less than one indicates that the event is less likely in group 1 than group 2, whereas an OR more than one indicates the event is more likely in group 1. An OR is significant at the p=0.05 level if the 95% confidence intervals do not include one. ORs and their corresponding 95% confidence intervals were calculated using scripts written and implemented in the R program.

When performing k multiple independent significance tests each at the α, *e.g.* α=0.05 level, the probability of making at least one Type I error (rejecting the null hypothesis inappropriately) is $1 - (1-\alpha)^k$. To correct for this, the Bonferroni method can be applied which divides the test significance level by the number of tests to give a corrected α level (α_β):

$$\alpha_\beta = \alpha / k$$

The Bonferroni correction was applied in the study analysis where appropriate to account for multiple comparisons.

2.5.4.3 Logistic regression

Regression analysis is used to assess the relationship between two or more variables and has three main uses: (i) To measure cause and effect between the variables; (ii) Investigate associations between the variables; (iii) Use the value of one variable to estimate the value of another. The aim of linear regression is to achieve the best fitting line to a set of data points by calculation of the equation of the line with its slope and y-

intercept. The null hypothesis is that the slope is zero indicating no linear relationship between the two variables. Logistic regression is a further extension of this, whereby there is a dichotomous dependent variable, *i.e.* has two possible values, *e.g.* male/female and one or more independent variables. The equation that best predicts the value of the dependent variable is found given the values of the independent variable(s). The independent variable(s) are not directly measured and instead are probabilities of obtaining a particular value of the nominal variable. The odds of the dichotomous variable occurring (equation 1, section 2.5.4.2) are then natural log transformed (logit transformation) to a straight line. Unlike linear regression which uses the least-squares method, the maximum likelihood method is used to estimate the slope and intercept of the line. The overall aim of the test is obtain the probability of a certain value of the dichotomous variable given the independent predictor variables. For example, in this study it was used to assess the odds of an isolate being a particular clonal complex or not, independent of serogroup.

2.5.4.4 Simpson's index of diversity

Simpson's index of diversity (*D*) is a mathematical formula often used in ecology to determine the species diversity in a particular habitat. It gives the probability that any two randomly selected individuals drawn from an infinitely large community belong to different species, or in the case of this study, isolates drawn from a population belong to different genetic types (Hunter & Gaston, 1988; Simpson, 1949). The value of the index ranges from zero to one, meaning that the nearer to one the greater the diversity and the nearer to zero the less the diversity. The bias-corrected form of the formula is as follows:

$$D = 1 - \frac{\sum n(n-1)}{N(N-1)}$$

N = total number of isolates

n = total number of isolates of a particular genetic type

Σ = sum of

It was used to assess the level of diversity of each country/year in terms of clonal complex and PorA and FetA type, each complex in terms of ST and each

subfamily/variant in terms of fHbp types. The 95% confidence intervals (CIs) for these indices were calculated as described by Grundmann *et al.* (Grundmann *et al.*, 2001):

$$\text{CIs} = (D - 2\sqrt{\sigma^2}, D + 2\sqrt{\sigma^2})$$

Where:

$$\sigma^2 = \frac{4}{n} [\Sigma\pi_j^3 - (\Sigma\pi_j^2)^2]$$

σ^2 = variance

$2\sqrt{\sigma^2}$ = standard deviation

π_j = frequency of type (no of types 1 to j)

Σ = sum of

The evenness value (E) is a measure of the relative abundance of the different genotypes making up the richness, *i.e.* number of genotypes of a population sample, *e.g.* country. It ranges from zero to one, meaning the nearer to one the more evenly each genotype contributes to the overall sample. It is calculated as per the method of Robinson *et al.* (Robinson *et al.*, 2006) as the ratio of the effective number of clones (S_e) to the total number of clones (genotypes) S, *i.e.* the richness of the sample. S_e and E increase as the number of isolates of each clone becomes more equal. E is calculated as follows:

$$S_e = \frac{1}{1-D}$$

Therefore,

$$E = \frac{S_e}{S}$$

CHAPTER 3: Geographic and temporal distribution and structuring of disease-associated *Neisseria meningitidis* genetic types in Europe

3.1 Introduction

Population samples of meningococcal disease isolates typically display a predominance of particular lineages (Caugant *et al.*, 1987b; Caugant, 1998) while carriage populations show much more diversity (Jolley *et al.* , 2000; Yazdankhah *et al.* , 2004). Despite a large overall diversity of meningococcal genetic types, most disease is caused by a small number of (hyperinvasive) clonal groupings (Figure 3.1) which can persist over time and across location (Caugant, 1998; Yazdankhah *et al.*, 2004). The reason for the apparent differences in population structures are that carriage in itself, while essential, is not a direct predictor of disease, but transmission and acquisition of the particular (hyperinvasive) clones that are associated with higher probability of disease are (Stephens, 2007). These more virulent genetic types may cause an increase in disease, particularly when first introduced, in a population if factors such as transmission, acquisition, population and host resistance allow it. The persistence of these lineages over time in the face of much recombination however, cannot be explained solely by neutral processes and instead is consistent with selection for variants that can out-compete others with slightly improved transmission efficiency and that can tolerate the penalty of increased virulence (Buckee *et al.*, 2008). Meningococcal disease genotypes are a subset of combinations of polymorphisms of the overall gene pool shared by the carriage and disease populations (Jolley *et al.*, 2005).

MLST is now becoming a routine part of disease surveillance and population biology of an increasing number of microorganisms and is accepted as a gold standard typing technique (Maiden *et al.* , 1998). It offers many advantages over traditional serological typing methods such as portability, unambiguity of results, ability of data to be shared and disseminated over the web, ease of comparison of data amongst laboratories and further use of the sequence data such as for evolutionary and phylogenetic analyses. Representative and well-sampled isolate collections can provide

insights into the dynamics of the population biology of the meningococcus and potentially aid the planning and implementation of interventions such as immunisation programmes and antibiotic use. This collection of samples from participating members of the EU-MenNet programme is the largest European disease dataset MLST typed to date (Brehony *et al.* , 2007). The aim of this part of the study was to determine the distribution and structuring of *N. meningitidis* hyperinvasive genetic types both geographically (across Europe) and temporally (over the three years 2000–2002). Sequence typing by MLST was used to examine the genetic diversity of the sample of 4048 representative disease isolates for the three years from the 18 countries involved.

3.2 Results

3.2.1. Analysis of MLST house-keeping genes

The most diverse of the house-keeping loci was the *aroE* locus with 111 alleles (Table 3.1). The least diverse locus was *adk* with 43 alleles. The range in number of variable sites for the loci was 62 for *gdh* and 203 for *aroE*. The test for selection, the d_N/d_S ratio, was calculated for each locus and produced low values (<1). Therefore, there was a higher proportion of synonymous to nonsynonymous substitutions indicating stabilising selection, as would be expected for house-keeping genes. The highest ratio value was found in the *aroE* locus and the lowest was in the *adk* locus.

Table 3.1: Genetic variation in *N. mening itidis* gene fragments. Table includes information on fragment size, number of allele types, number of variable nucleotide sites, percentage variable sites, d_N/d_S ratio and mean GC% content.

locus	fragment size (bp)	no. of alleles	no. of variable sites	% variable sites	d_N/d_S ratio	GC %
abcZ	433	73	103	23.8	0.0620	50.9
adk	465	43	64	13.8	0.0275	52.2
aroE	490	111	203	41.4	0.2726	55.6
fumC	465	90	79	17.0	0.0291	57.2
gdh	501	83	62	12.4	0.0587	52.4
pdhC	480	84	105	21.9	0.0690	56.2
pgm	450	81	110	24.4	0.1071	53.9

3.2.2 Genetic diversity of European disease isolates

Analysis of 4048 European disease isolates from the 18 countries involved in the EU-MenNet project for the 3 years 2000–2002 inclusive (Table 1, Appendix) revealed the predominance of hyperinvasive complexes despite the large overall number of genetic types (Figure 3.1).

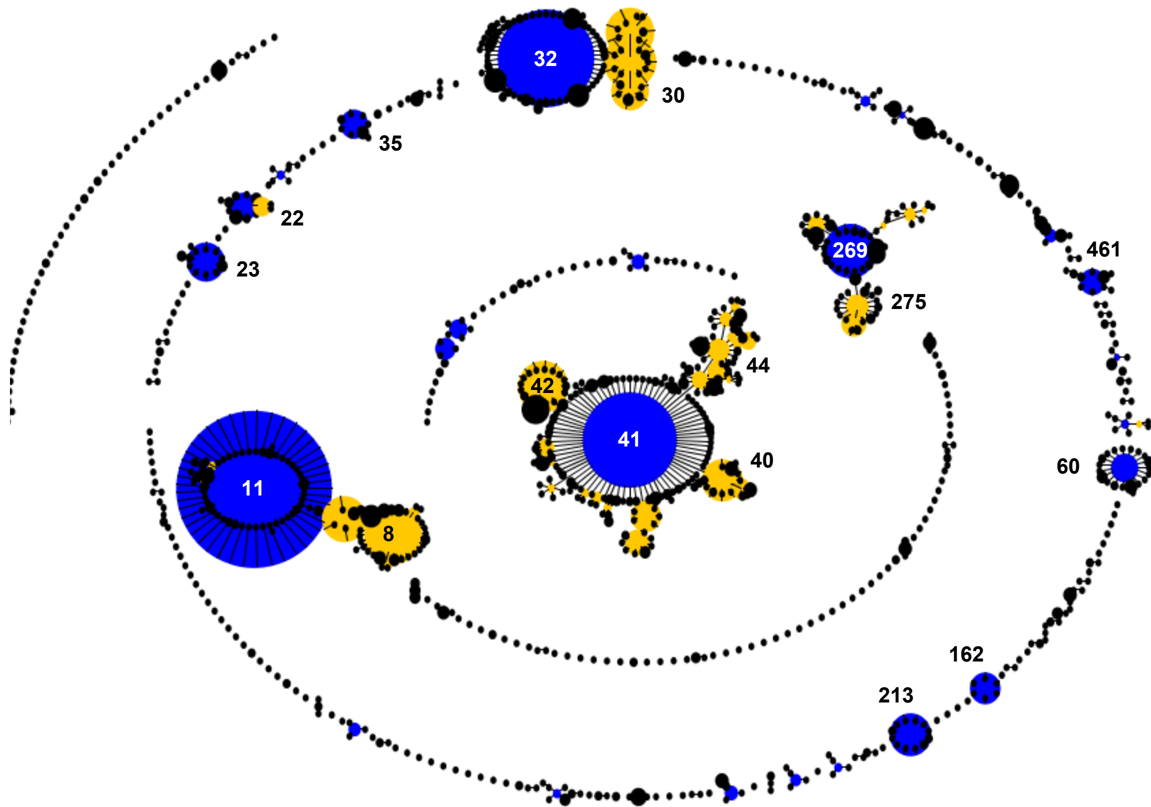


Figure 3.1: Snapshot of the European disease isolate dataset using eBURST (Feil *et al.* , 2004) illustrating the predominance of a few major clones despite a large number of genetic types. *Blue* circles correspond to the central founder ST of a complex. *Yellow* circles correspond to the central genotypes of the related subgroups. *Black* circles correspond to all the other unlinked STs. The area of the circles is proportional to the abundance of the ST in the dataset. For example, ST-11 has the highest frequency with 812 isolates and is therefore the largest of the circles. By the algorithm, members of a clonal complex must share six out of seven loci with at least one other member of the group.

While there was diversity in the sequence types (STs) found (978 types) (Figure 3.2), only ten accounted for half of the isolates: ST-11 (20%), ST-32 (8%), ST-41 (7%), ST-8 (3%), ST-269 (2%), ST-33 (2%), ST-42 (2%), ST-34 (2%), ST-66 (2%) and ST-213 (1%). 724 STs, accounting for 14% of the isolates, were observed only once. The STs resolved into 33 distinct clonal complexes (Table 3.2), the most prevalent being the ST-41/44 complex (1014 isolates, 25%), ST-11 complex (901 isolates, 22%), ST-32 complex (706 isolates, 17%), ST-8 complex (273 isolates, 7%) and ST-269 complex

(256 isolates, 6%). These five complexes together accounted for 77% (3150 isolates) of all isolates. Isolates unassigned to a clonal complex accounted for 7% (274 isolates) of the dataset.

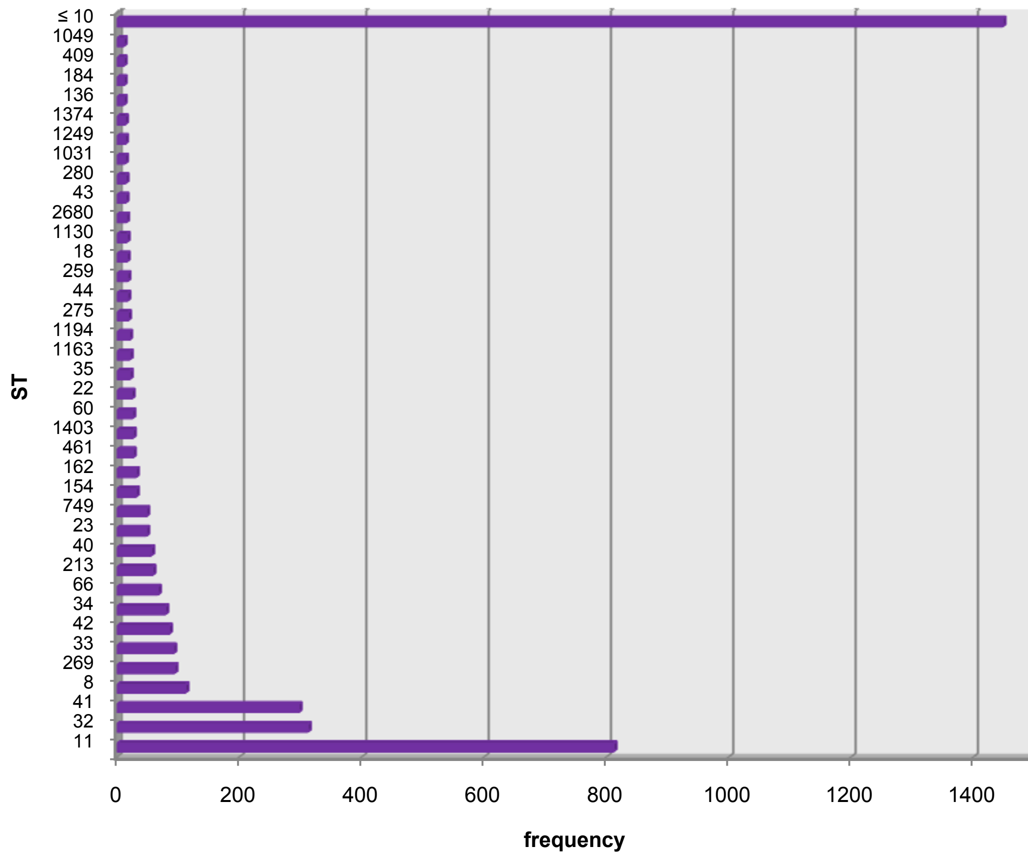


Figure 3.2: Rank-abundance plot illustrating ST diversity in the European disease population. A total of 978 STs were found in the dataset. STs with a frequency less than or equal to 10 are grouped.

Using Simpson's index of diversity (*D*) (Simpson, 1949), the diversity of each clonal complex found in terms of ST was calculated (see Materials and Methods) (Figure 3.3; Table 2 Appendix). 95% confidence intervals (CIs) were calculated as described by Grundmann *et al.* (Grundmann *et al.* , 2001). Non-overlapping CIs indicate a significant difference in *D*. A value of *D* near one indicates a high level of diversity, while a value near zero indicates a low level of diversity. Differences in the degree of diversity among clonal complexes were evident. The ST-254 and ST-364 complexes were found to be the most diverse with very high *D* values of 0.989 – with almost a unique ST for each isolate found. They were significantly more diverse than 25 out of the remaining 30 complexes. The least diverse was the ST-11 complex with a *D* value of 0.188. Its central genotype ST-11 accounted for 90% of all isolates. It was significantly less diverse than 26 out of the remaining 31 complexes. The other major hyperinvasive complexes, ST-41/44, ST-32, ST-8 and ST-269 complex, were much more diverse in comparison. The ST-41/44 complex was the most diverse of the major hyperinvasive clonal complexes, with 259 STs and an index of diversity of 0.901.

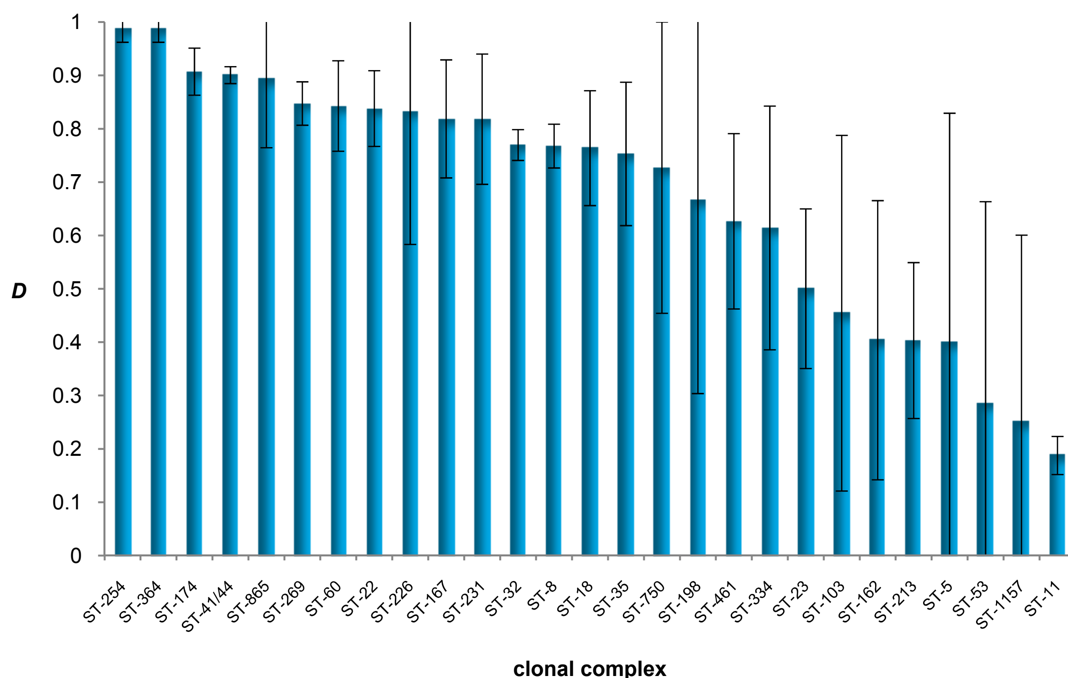


Figure 3.3: Diversity of clonal complexes in terms of ST using Simpson's index of diversity (*D*) arranged in order of decreasing *D* value (Simpson, 1949). Error bars indicate 95% CIs. Note: some complexes had relatively low frequencies, *i.e.* n<15 and therefore gave very wide 95% CIs when calculated. Non-overlapping CIs indicates significant difference in *D*.

3.2.3 Distribution of genetic types in Europe

The geographical distribution of genotypes across Europe showed the main hyperinvasive complexes accounting for most of the disease isolates over the three year period of the study. There was some variance in the prevalence of each complex among countries (Figure 3.4) and some clonal complexes were found to be significantly associated with certain countries by the χ^2 test (Table 3.2).

For example, the range of prevalence of the ST-8 complex varied from 0% (Czech Republic, Iceland, Sweden, and Finland) to 39% (Portugal). Unlike the other major complexes, *i.e.* ST-11, ST-32 and ST-41/44, the ST-8 complex was not present in all 18 countries in the study. It was not found in the Czech Republic, Finland, Iceland and Sweden. This complex was significantly associated with Spain, Portugal and Germany (Table 3.2, $p<0.00008$, after Bonferroni correction). The ST-32 complex was much more prevalent in Denmark than most other countries and it accounted for almost half of all isolates in this country (χ^2 test, $p<0.00008$). In 2002, serogroup B, which the ST-32 complex is associated with (Trotter *et al.*, 2007), was shown to account for 75% of Danish meningococcal disease. The ST-11 complex was found to be very prevalent in Icelandic isolates (71%), although it must be noted that there were just 17 isolates from one year (2001) in the dataset. It was the second most common complex after ST-32 complex in a study of invasive disease in Iceland from 1977 and 2004 (Gottfredsson *et al.*, 2006). The ST-32 and ST-11 complexes were significantly associated with five out of 18 countries. The ST-269 complex appeared to be particularly associated with countries of the North-West region, *i.e.* England and Wales, Scotland (χ^2 test, $p<0.00008$) and the Republic of Ireland (χ^2 test, $p<0.05$). This complex has emerged in these countries in recent years and is strongly linked with serogroup B disease in Europe (Trotter *et al.*, 2007). The UK and Republic of Ireland implemented MCC vaccine strategies which greatly reduced serogroup C disease (Miller *et al.*, 2001) and this is likely to account for the increased proportion of serogroup B-related complexes although there is no evidence of vaccine-induced capsule replacement. Also noteworthy is the serogroup Y-associated ST-23 complex (Yazdankhah *et al.*, 2004) which was found at a much higher frequency in Sweden (11%) and Finland (8%) than in most other countries (Fisher's exact test, both $p<0.00008$).

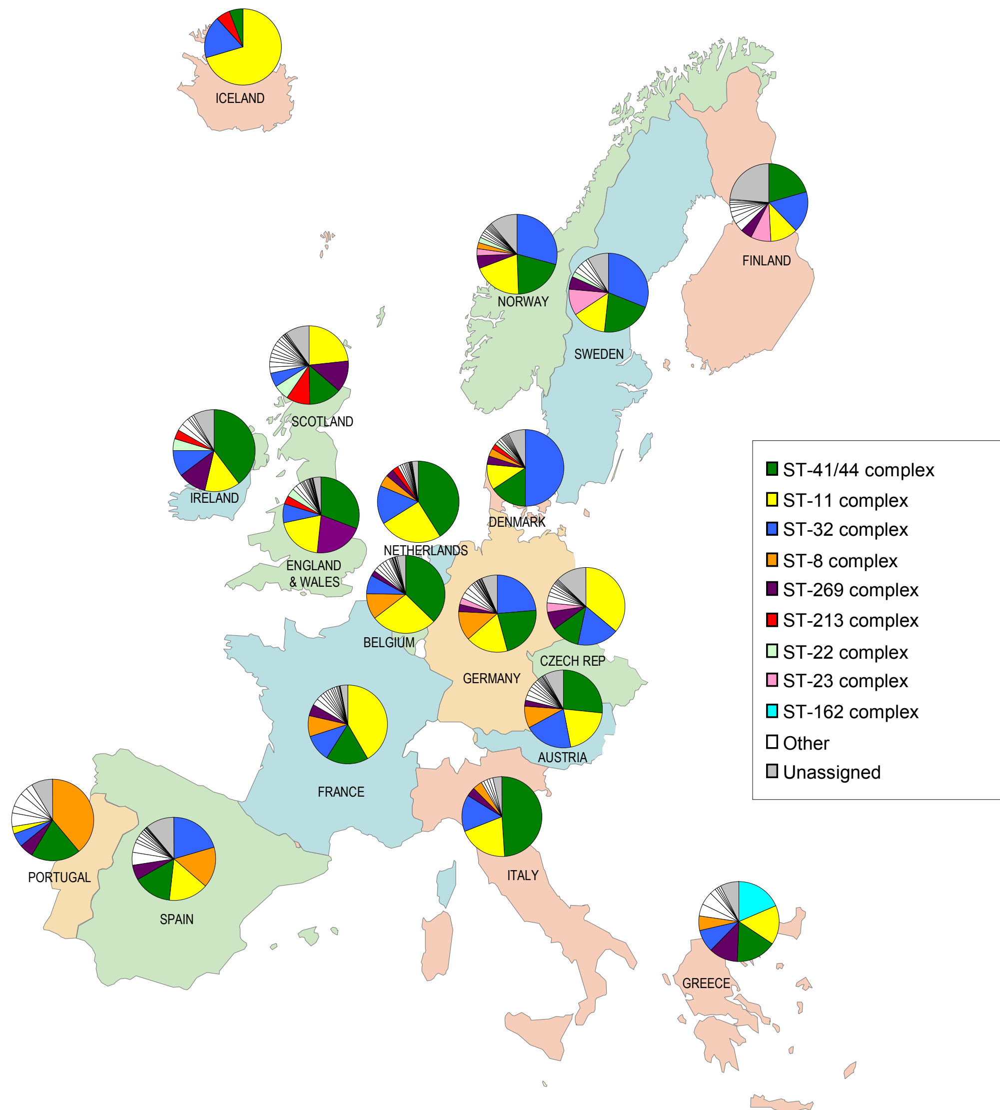


Figure 3.4: Distribution of clonal complexes in Europe 2000–2002. Note: Isolates were not submitted for: Iceland 2000 and 2002; Portugal 2000 and 2002; Sweden 2000.

Some clonal complexes were found to be significantly associated with certain countries and were rarely found elsewhere (Table 3.2). For example, in Greece the ST-162 complex had a significant association (Fisher's exact test, p<0.00008) and a much higher prevalence (18%) in comparison to other countries where it is found much less frequently. The importance of this complex in Greece has previously been noted in studies of carriage and disease isolates where it was first identified (Yazdankhah *et al.*, 2004; Yazdankhah *et al.* , 2005b). Also significant (Fisher's exact test, at the p<0.005 level) was the association of ST-231 and ST-60 complexes with Greece. In Scotland the ST-213 complex was present at a higher prevalence (10%) than in most other countries and found to be significantly associated (Fisher's exact test, p<0.00008). Scotland was noteworthy in the significant associations of a number complexes rarely found elsewhere (Fisher's exact test, p<0.00008). Two of these complexes, the ST-53 and ST-198 complex, have previously been found to possess the *cnl* locus, *i.e.* are acapsulate (Claus *et al.* , 2002). Scotland also accounted for half of all the ST-1157 complex isolates found in Europe (Fisher's exact test, p<0.00008). There was also a significant association here with the ST-22 complex (Fisher's exact test, p<0.00008) which is predominantly associated in Europe with the W-135 complex (Trotter *et al.* , 2007). This serogroup was introduced into the UK and the Republic of Ireland following the Hajj in 2000. Another significant association (Fisher's exact test, at the p<0.005 level) with Scotland was the ST-750 complex

Other notable significant associations of complex and country were the ST-174 complex in France and the ST-18 complex in the Czech Republic. There was a relatively large proportion of unassigned isolates in Finland and these are dominated by the ST ST-2793 which accounts for 30% of isolates. It was found over the three years in Finland and was not found in any other country in Europe so could possibly be an emerging clone so far just found in this country. Interrogation of the PubMLST database (http://pubmlst.org/neisseria/) has not shown this ST in any other country worldwide further pointing to a possible new clone. It shares four loci with members of the ST-334 complex but not the central genotype so cannot be considered as part of this complex. There were no ST-334 complex isolates found in the Finnish sample. ST-2793 also has two single-locus variants ST-2997 and ST-2690 within the Finnish collection, which both vary at the *adk* locus. They differed from the ST-2793 allele by five and two polymorphisms respectively and therefore could have arisen through mutation.

Table 3.2: Association of clonal complex and country. Numbers correspond to frequencies of each complex in each country.

clonal complex	Austria	Belgium	Czech Rep	Denmark	E & W	Finland	France	Germany	Greece	Iceland	Ireland	Italy	Netherlands	Norway	Portugal	Scotland	Spain	Sweden	Total
ST-41/44 complex	50	***101****	22	27	***100****	27	58	116	19	1	***43****	***39****	***254****	38	7	31	62	19	1014
ST-11 complex	38	***74****	***67****	19	65	15	***142****	95	21	***12****	15	16	***155****	37	1	55	63	13	903
ST-32 complex	38	21	32	***87****	26	23	38	***125****	12	3	11	12	94	***55****	2	14	***85****	***29****	707
Unassigned	14	9	23	13	11	***31****	10	34	9		8	3	15	21	3	23	41	8	275
ST-8 complex	17	29		6	2		29	***64****	6		1	3	30	5	***14****	2	***65****		273
ST-269 complex	5	6	14	6	***68****	6	16	16	14		12	3	18	10	2	***31****	24	5	256
ST-213 complex		4	2	4	12	2	2	2		1	4	1	15	2		***23****			74
ST-23 complex	3	1	7		2	***11****	4	14	1		1		3	5		1	5	***10****	68
ST-22 complex	1	4		2	12	2	6	2			5	1	5	4		***15****	6	2	67
ST-60 complex	4	2		1	6	1	5	14	7		3	1	8			5	6	2	65
ST-35 complex	4	1	1		1		3	1	6				4	2	2	1	***16****	1	43
ST-461 complex		5			2		4	2					4		1	5	***20****		43
ST-162 complex	2		1		1		2	4	***23****				1	2		3	2		41
ST-18 complex			***11****		6	5		10				1				4			37
ST-174 complex	3	1	2	1	2	3	***10****	3	1				3				3		32
ST-334 complex		4			1		1	***12****					1	1			1		21
ST-167 complex	1	1			1		1	2			3		1	2	1	4		2	19
ST-364 complex				1	1	***4****	3	2								1	2		14
ST-254 complex		3	1					1					2		2	1	1	2	13
ST-103 complex	1		1	1	3	1	3						1		1				12
ST-865 complex		1	3			1		2	1								4		12
ST-231 complex	2	1		1				4	3										11
ST-750 complex		1		2	1								1	1		4	1		11
ST-1157 complex					2								1	1		***4****			8
ST-53 complex											1					***6****			7
ST-5 complex				1			3							1					5
ST-226 complex	1		3																4
ST-198 complex																***3****			3
ST-92 complex				2										1					3
ST-37 complex																	2		2
ST-376 complex	1												1						2
ST-116 complex	1																		1
ST-549 complex	1																		1
ST-175 complex								1											1
Total	**187**	**269**	**190**	**174**	**325**	**132**	**340**	**525**	**123**	**17**	**107**	**80**	**617**	**188**	**36**	**236**	**409**	**93**	**4048**

**bold italics* correspond to a significant association p<0.00008 after Bonferroni correction. *Purple* corresponds to complexes unique to country and found just once, *green* found only two times in one country and *blue* found only three times in one country. Note: E & W is England and Wales.

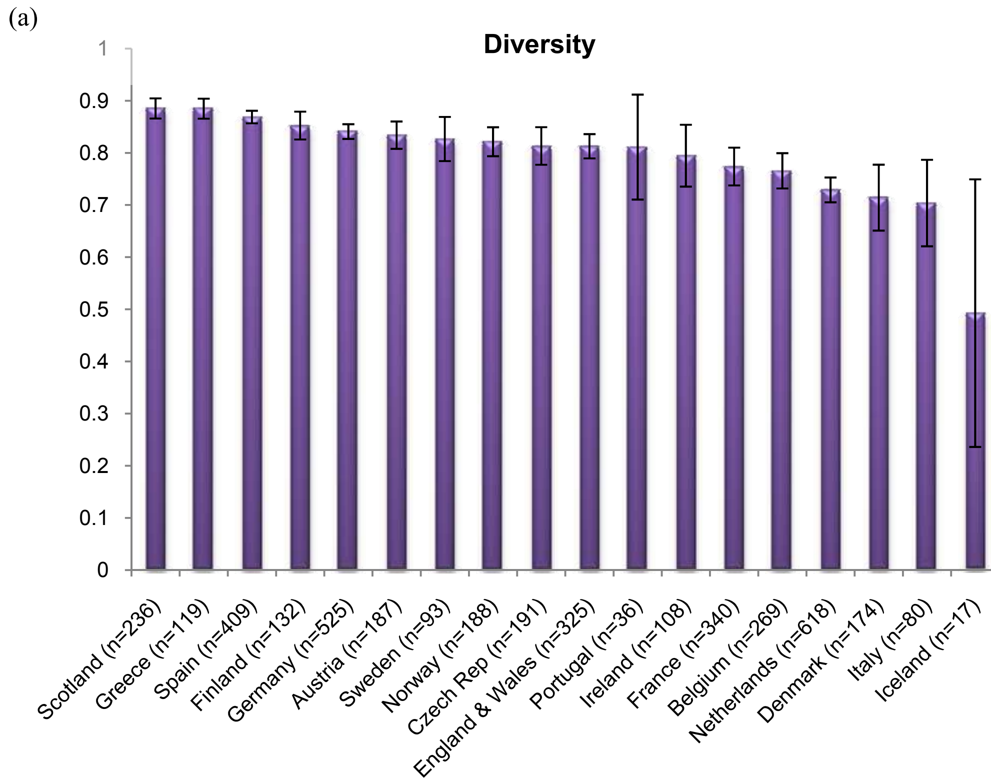
3.2.3.1 Genetic diversity of invasive meningococcal populations by country

The genetic diversity of each country in terms of clonal complexes present was calculated using Simpson's index of diversity. 95% confidence intervals (CIs) were calculated as described by Grundmann *et al.* (Grundmann *et al.* , 2001). A significant difference in *D* is indicated by non-over-lapping CIs. The evenness (*E*) value for each country was also calculated. *E* gives a measure of the relative abundance of the different genotypes making up the richness, *i.e.* number of genotypes in a sample. Like *D* it ranges from zero to one: as *E* increases the more equal the contribution of each clone in terms of frequency becomes.

There was a range in diversity across Europe (Figure 3.5 (a)). The most diverse countries with a *D* value of 0.885 were Scotland and Greece. Greece was also the country with the greatest evenness (0.667) meaning that it had the most even contribution of clonal complexes to its genotype distribution and therefore not dominated by a particular clone. Greece and Scotland were significantly more diverse than all countries except Austria, Finland, Portugal, Spain, and Sweden. The least diverse country was Iceland (0.493), although it must be noted that there were just data from a single year (2001). The next lowest diversity was found in Italy (0.704). The country with the lowest *E* value was the Netherlands (0.176), therefore displaying an unequal distribution in the abundance of each clonal complex. It was also significantly less diverse than all countries except Portugal, Ireland, Iceland, France and Belgium.

The range of diversity for each country in terms of ST was similar, although the *D* value in each country was higher, as would be expected (Table 3, Appendix). All countries had a *D* value of at least 0.8 except for Iceland (0.515). This time Finland was the most diverse country with a *D* of 0.966. This is probably due to the fact that it has quite a high proportion of STs unassigned to a clonal complex. Evenness values for each country were lower than for clonal complexes indicating more uneven distributions of STs in each country than for complexes. This is because many of the hyperinvasive clonal complexes are not evenly made up of STs. Rather, one or two (in the case of ST-41/44), presumptive founder, STs dominate, *e.g.* ST-11 complex. France had the lowest evenness value of just 0.064. This is most likely due to the fact that ST-11 dominates disease isolates in this country, accounting for 39% of them, while the next most prevalent STs contribute much less in comparison (ST-41, 7%; ST-8, 5%, ST-32, 4%).

(a)



(b)

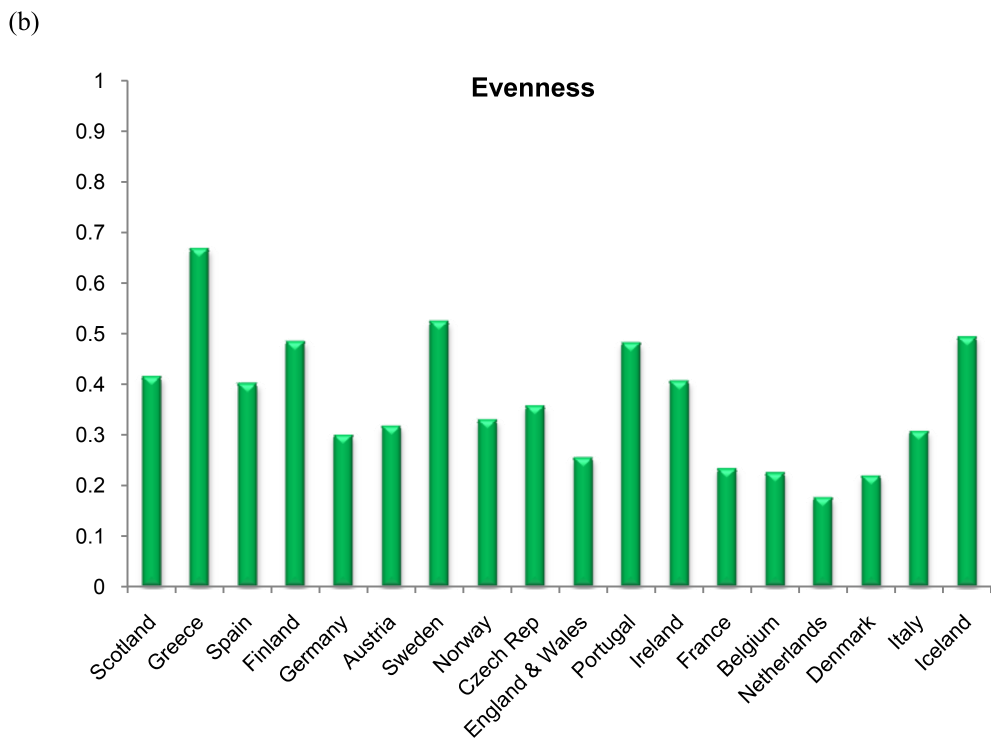


Figure 3.5: (a) Genetic diversity (*D*) of each country in terms of clonal complex. A value of *D* near one indicates a high level of diversity, while a value near zero indicates a low level of diversity. Error bars indicate 95% CIs. Non-overlapping CIs indicates significant difference in *D*. (b) Evenness (*E*) is a measure of the relative abundance of each of the genotypes making up the population (in this case, country) sample. As *E* increases the frequency of each clone becomes more equal.

There was a wide range of proportions of isolates accounted for by presumably transient unique STs not found in any other country (Figure 3.6). The highest proportion was in Spain where over a quarter (27.1%) of all isolates had an ST that was found just once in that country and also not found elsewhere in Europe. Iceland had the fewest with 6% although there were just isolates from one year from this country.

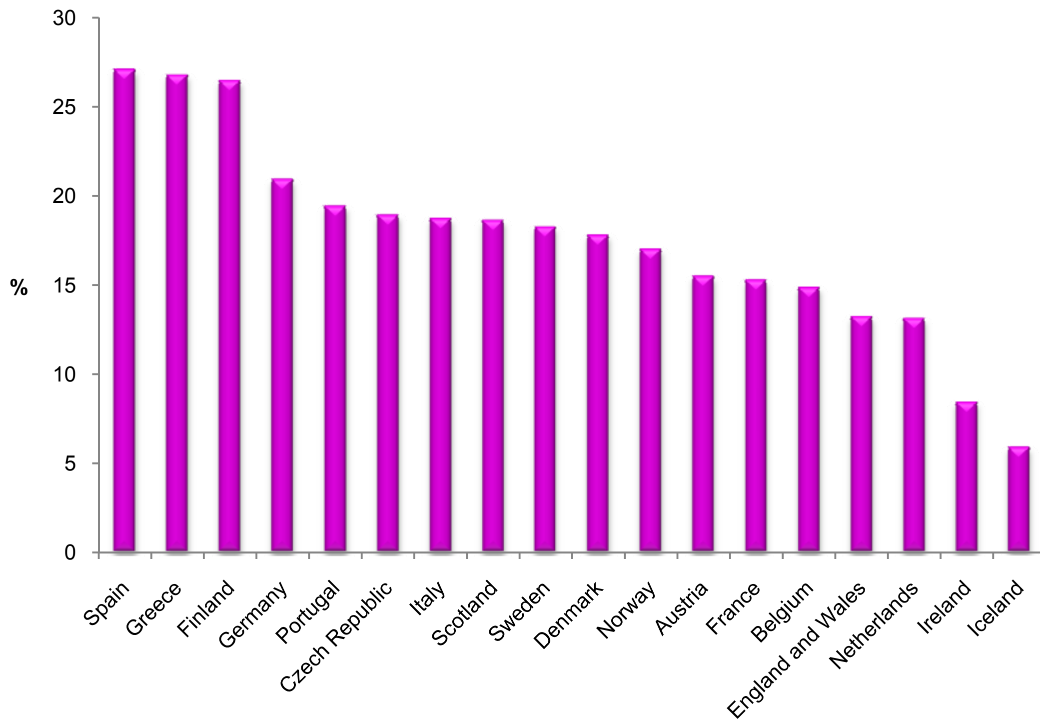


Figure 3.6: Percentage of unique STs only found in the respective country.

3.2.4 Changes in genetic types over time

Overall there were no major changes in the distribution of genetic types over the three years. The main hyperinvasive clonal complexes were found to account for most disease isolates for each year of the study. However, there was a very significant decrease in ST-8 complex from 9% to 5% over this period ($p<0.0005$ after Bonferroni correction) (Table 3.3). The odds ratio (OR) of 0.49 (which was significant at the 0.05 level, 95% CIs 0.358–0.678) comparing 2002 to 2000 means that it was almost half as likely for an isolate to be ST-8 complex in 2002 than it was in 2000. There was also a significant increase, albeit at the $p=0.005$ level, in ST-174 complex from 0.3% to 1.5% from 2000 to 2002. The OR comparing 2002 and 2000 was 5.5 and significant for this complex (1.867–24.349). There were significant changes at the $p<0.05$ level over the three years in the ST-11 complex, increasing 2000–2001 from 19.5% to 24.8% and then decreasing 2001/2002 from 24.8% to 21.9%. However, the difference between the prevalence in 2000 and 2002 was not significant as was the OR comparing 2000 and 2002 (OR 1.15; 95% CIs 0.952–1.40). There were changes in the ST-213 complex first decreasing (1.5 to 1.3%) then increasing (1.3 to 2.6%); the ST-162 complex decreasing (0.9 to 0.5%) then increasing (0.5 to 1.5%); the ST-18 complex increasing (0.9 to 1.5%) then decreasing (1.5 to 0.4%) and the ST-334 complex first decreasing (0.4 to 0.2%) then increasing (0.2 to 0.9%).

Table 3.3: Change in clonal complex over time (2000–2002).

clonal complex	year (frequency and percentage)				χ^2 p-value		
	2000	**2001**	**2002**	**total**	**overall change**	**2000/2002 change**	**2002/2000 OR and 95% CIs**
ST-103 complex	3 (0.3%)	8 (0.5%)	1 (0.1%)	12 (0.3%)	0.075	0.340	0.309 (0.011–2.666)
ST-11 complex	230 (19.6%)	372 (24.8%)	301 (21.9%)	903 (22.3%)	0.005†	0.145	1.154 (0.952–1.40)
ST-1157 complex	1 (0.1%)	2 (0.1%)	5 (0.4%)	8 (0.2%)	0.266	0.227	3.846 (0.591–101.784)
ST-116 complex	0	1 (0.1%)	0	1 (0.0%)	n/a	n/a	n/a
ST-162 complex	11 (0.9%)	9 (0.6%)	21 (1.5%)	41 (1.0%)	0.014†	0.184	1.626 (0.792–3.534)
ST-167 complex	4 (0.3%)	7 (0.5%)	8 (0.6%)	19 (0.5%)	0.676	0.564	1.676 (0.515–6.20)
ST-174 complex	3 (0.3%)	9 (0.6%)	20 (1.5%)	32 (0.8%)	0.002†	0.001	5.505 (1.867–24.350)
ST-18 complex	10 (0.9%)	22 (1.5%)	5 (0.4%)	37 (0.9%)	0.008†	0.124	0.431 (0.1307–1.236)
ST-198 complex	0	0	3 (0.2%)	3 (0.1%)	0.064	n/a	n/a
ST-213 complex	18 (1.5%)	20 (1.3%)	36 (2.6%)	74 (1.8%)	0.025†	0.059	1.716 (0.980–3.113)
ST-22 complex	17 (1.4%)	29 (1.9%)	21 (1.5%)	67 (1.7%)	0.555	0.873	1.052 (0.551–2.038)
ST-226 complex	2 (0.2%)	2 (0.1%)	0	4 (0.1%)	0.467	n/a	n/a
ST-23 complex	17 (1.4%)	28 (1.9%)	23 (1.7%)	68 (1.7%)	0.702	0.659	1.153 (0.614–2.211)
ST-231 complex	3 (0.3%)	3 (0.2%)	5 (0.4%)	11 (0.3%)	0.698	0.733	1.394 (0.328–7.241)
ST-254 complex	1 (0.1%)	7 (0.5%)	5 (0.4%)	13 (0.3%)	0.191	0.227	3.846 (0.591–101.784)
ST-269 complex	71 (6.1%)	93 (6.2%)	92 (6.7%)	256 (6.3%)	0.789	0.515	1.163 (0.848–1.603)
ST-32 complex	222 (18.9%)	252 (16.8%)	234 (17.0%)	707 (17.4%)	0.269	0.207	0.878 (0.717–1.076)
ST-334 complex	5 (0.4%)	3 (0.2%)	13 (0.9%)	21 (0.5%)	0.022†	0.095	2.185 (0.812–6.9784)
ST-35 complex	12 (1.0%)	11 (0.7%)	20 (1.5%)	43 (1.1%)	0.170	0.212	1.419 (0.697–3.021)
ST-364 complex	4 (0.3%)	2 (0.1%)	8 (0.6%)	14 (0.3%)	0.120	0.387	1.676 (0.515–6.520)
ST-37 complex	0	0	2 (0.1%)	2 (0.0%)	0.207	n/a	n/a
ST-376 complex	0	2 (0.1%)	0	2 (0.0%)	0.337	n/a	n/a
ST-41/44 complex	307 (26.2%)	365 (24.4%)	342 (24.9%)	1014 (25.1%)	0.539	0.447	0.933 (0.780–1.116)
ST-5 complex	2 (0.2%)	3 (0.2%)	0	5 (0.1%)	0.274	n/a	n/a
ST-549 complex	1 (0.1%)	0	0	1 (0.0%)	0.283	n/a	n/a
ST-60 complex	21 (1.8%)	24 (1.6%)	20 (1.5%)	65 (1.6%)	0.797	0.499	0.810 (0.432–1.511)
ST-750 complex	4 (0.3%)	3 (0.2%)	4 (0.3%)	11 (0.3%)	0.757	1	0.852 (0.192–3.785)

ST-8 complex	107 (9.1%)	101 (6.7%)	65 (4.7%)	273 (6.7%)	**0.0001***	**0.00001***	0.495 (0.358–0.678)
ST-865 complex	7 (0.6%)	3 (0.2%)	2 (0.1%)	12 (0.3%)	0.097	0.090	0.256 (0.035–1.092)
ST-92 complex	0	1 (0.1%)	2 (0.1%)	3 (0.1%)	0.642	n/a	n/a
Unassigned	75 (6.4%)	99 (6.6%)	102 (7.4%)	278 (6.9%)	0.619	0.232	1.139 (0.839–1.55)
total	1172	1500	1376	4048			

* **significant** after the Bonferroni correction p=0.0005; † significant at the p=0.05 level

At the individual country level, there were a number of changes in clonal complex prevalence over the three year timeframe (Figure 3.7), some of which were significant (Table 3.4). There were overall increases in the ST-11 complex in Austria, Czech Republic, Denmark, Germany, Netherlands, France and Spain with the changes being significant in Austria, Germany, the Netherlands and Spain ($p<0.05$, $p<0.05$ $p<0.005$ and $p<0.05$ respectively). The ORs comparing 2002 to 2000 for Germany, the Netherlands and Spain were also significant at 1.938 (95% CIs 1.10–3.538), 3.126 (95% CIs 1.835–5.510) and 2.054 (95% CIs 1.075–4.057) respectively. There was a decline in the overall prevalence of this complex in England and Wales, Finland, Greece, Republic of Ireland, Scotland and Sweden. The decrease was highly significant in the Republic of Ireland and Scotland ($p<0.005$ and ORs 2002/2000, 0.109 (95% CIs 0.004–0.602) and 0.079 (95% CIs 0.022–0.217) respectively).

Other trends of note were a sizeable and very significant decrease in ST-8 complex in Spain from 27% to 7% over the 3 years ($p<0.005$; OR 0.202; 95% CIs 0.091–0.409). In Spain, the predominant ST in this complex was ST-8. It accounted for 59% of the complex in 2000 and 75% in 2001 but fell to 20% in 2002. The complex seemed to disappear in England and Wales in this time although the change in prevalence from 2% in 2000 to 0% in 2001 and 2002 was of borderline significance. It also disappeared by 2002 in Greece, Norway and the Republic of Ireland but these were not significant. The ST-32 complex decreased in Belgium, Finland, Republic of Ireland, the Netherlands and Norway, but only significantly in Austria (significant; $p<0.05$; OR 0.355; 95% CIs 0.131–0.894)) and Germany ($p<0.05$; OR 0.594; 95% CIs 0.360–0.979). It increased in England and Wales and Scotland, although not significantly. The ST-41/44 complex increased in the three years in England and Wales, France, Greece, Italy, Spain, Sweden and the Republic of Ireland though not significantly. It decreased in Austria, Germany the Czech Republic but only being

significant in the Netherlands ($p<0.05$; OR 0.648; 95% CIs 0.429–0.974)). There were significant changes in ST-269 complex prevalence in the Netherlands and Spain ($p<0.05$). In the Netherlands ST-479 became the predominant ST in this complex accounting for 0% in 2000 and 2001 rising to 80% of isolates in 2002. Previously, the central ST, ST-269, accounted for most isolates, *i.e.* 50% in 2000. However, this dropped to 20% in 2002. In Spain ST-1163 went from accounting for 67% of isolates in the ST-269 complex in 2000, to 91% in 2002. There was an overall increase in this ST from 2000 to 2002 ($p<0.05$) when it went from 1 to 7%. The OR comparing 2002 to 2000 for ST-269 complex in Spain was 3.679 (95% CIs 1.101–17.346). The ST-269 complex which is serogroup B associated has expanded in the UK in recent years. Interesting to note is the increase in proportion of the ST-275 of the ST-269 complex in England and Wales while the proportion of the central genotype itself reduced (however neither significantly). This has been noted elsewhere (Lucidarme *et al.*, 2009).

The ST-213 complex emerged in Belgium in 2002 having being absent in the previous two years ($p<0.005$). Another possible emerging complex was the ST-174 complex in France which went from a prevalence of 1.1% to 6.4% from 2000 to 2002 ($p<0.05$). 60% of the ST-174 complex in France was ST-2495 and this ST appeared here in 2002. It went from accounting for 0% of isolates in 2000 to 5% in 2002 ($p<0.05$). The ST-174 complex also appeared in the Netherlands in 2002 having not been present the previous two years although this change was of borderline significance. The capsule-null-associated ST-53 complex increased significantly in Scotland over the three years from 0 to 6% ($p<0.05$) although numbers were small (0 to 6 isolates). There was a highly significant increase in the ST-334 complex in Germany from 0.7% to 5.2% from 2000 to 2002 ($p<0.005$; OR 6.83; 95% CIs 1.30–169.1), while the ST-18 complex seemed to disappear (2.1% to 0%; $p<0.005$). There was also a highly significant reduction in the ST-66 from 2000 to 2002 from 7 to 0.5% ($p<0.005$). ST-66 is a member of the ST-8 complex and it is noteworthy that while it went from accounting for 50% of the complex in 2000 to 4% in 2002, ST-8 itself went from accounting for 30% of the complex in 2000 to 87% in 2002. So it appears there has been a switch in the dominant ST of this complex over the three years. The complex itself decreased slightly in prevalence from 14 to 11%.

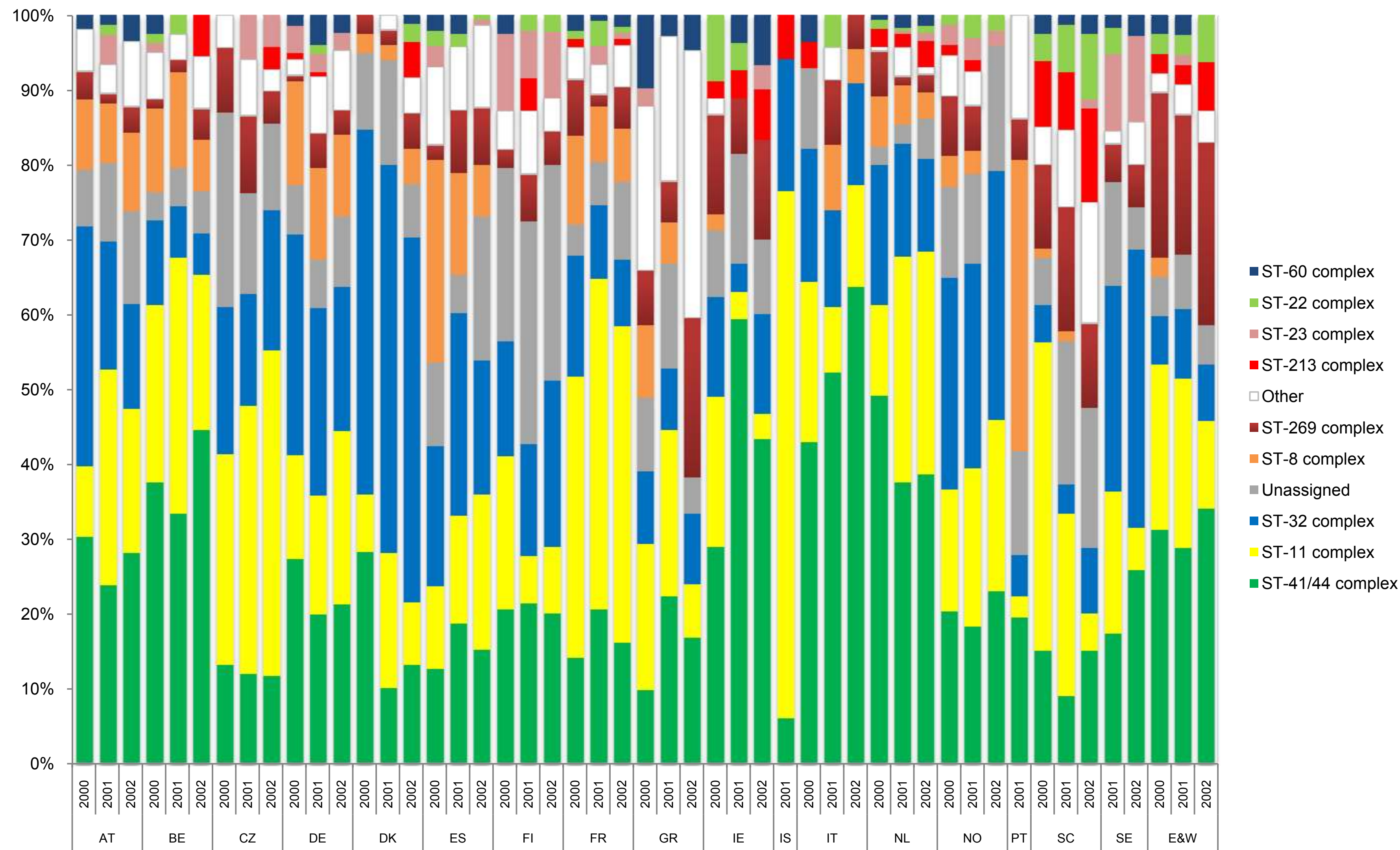


Figure 3.7: Clonal complex distribution seen in participating countries over the three year period 2000–2002.

Table 3.4: Significant clonal complex change from 2000–2002 by country by frequency (percentage in brackets).

country	clonal complex	year (frequency and percentage)			χ^2 p-value		2002/2000 OR and 95% CIs
		2000	2001	2002	overall change	2000/2002 change	
Austria	ST-11	5 (9.3)	22 (28.9)	11 (19)	0.021	0.181	2.244 (0.743–7.770)
	ST-32 complex	17 (31.5)	13 (17.1)	8 (14)	0.045	0.040	0.355 (0.136–0.894)
Belgium	ST-213 complex	0 (0)	0 (0)	4 (5.5)	0.005	0.048	n/a
Denmark	ST-41/44 complex	11 (27.5)	5 (10.0%)	11 (13.1)	0.007	0.050	0.401 (0.153–1.044)
England & Wales	ST-8 complex	2 (2.6)	0 (0)	0 (0)	0.0556*	0.192	n/a
France	ST-174 complex	1 (1.1)	1 (0.8)	8 (6.4)	0.030	0.082	5.576 (0.975–142.1124)
Germany	ST-11 complex	19 (13.3)	27 (15.8)	49 (23)	0.042	0.022	1.938 (1.10–3.538)
	ST-18 complex	3 (2.1)	8 (4.7)	0 (0)	0.003	0.064	n/a
	ST-334 complex	1 (0.7)	0 (0)	11 (5.2)	0.001	0.032	6.829 (1.293–169.055)
	ST-32 complex	41 (28.7)	43 (25.1%)	41 (19.2)	0.106	0.039	0.594 (0.360–0.979)
Ireland	ST-11 complex	13 (26)	1 (3.7)	1 (3.2)	0.003	0.013	0.109 (0.0043–0.602)
	ST-41/44 complex	13 (26)	16 (59.3)	14 (45.2)	0.014	0.093	2.313 (0.891–6.124)
Netherlands	ST-11 complex	21 (11.9)	73 (30.5)	61 (29.9)	0.00001	2.17E-05	3.126 (1.835–5.510)
	ST-174 complex	0 (0)	0 (0)	3 (1.5)	0.0584*	0.252	n/a
	ST-269 complex	10 (5.7)	3 (1.3)	5 (2.5)	0.032	0.120	0.424 (0.1268–1.2386)
	ST-41/44 complex	87 (49.4)	88 (36.8)	79 (38.7)	0.026	0.036	0.648 (0.429–0.974)
Scotland	ST-11 complex	32 (41)	19 (24.4)	4 (5)	5.72E-07	3.15E-08	0.079 (0.022–0.217)
	ST-53 complex	0 (0)	1 (1.3)	5 (6.3)	0.050	0.0587*	n/a
	ST-461 complex	1 (1.3)	4 (5.1)	0 (0)	0.049	0.245	n/a
Spain	ST-8 complex	39 (26.9)	16 (13.3)	10 (6.8)	0.00001	4.35E-06	0.202 (0.091–0.409)
	ST-269 complex	3 (2.1)	10 (8.3)	11 (7.5)	0.038	0.0516*	3.679 (1.101–17.346)
	ST-11 complex	16 (11)	17 (14.4%)	30 (20.4)	0.0837	0.028	2.054 (1.075–4.057)

*borderline significance at the p=0.05 level

3.2.4.1 Changes in genetic diversity over time

To assess any change in the levels of clonal complex diversity, which indicates richness in terms of STs and their respective abundance, the index of diversity was calculated for each country per year (Figure 3.8). 95% confidence intervals (CIs) were calculated according to the method of Grundmann *et a l.* (Grundmann *et al.* , 2001). Non-overlapping CIs indicate significant differences in the diversity index.

In 2000 the country with the highest genetic diversity was Greece (0.895) and the country with the lowest diversity was Denmark (0.681). In 2001 Finland had the highest diversity (0.888) while the Republic of Ireland had the lowest diversity (0.593). In 2002, the highest diversity was in Scotland (0.915) and the lowest was in Italy (0.655). Over the course of the three years the only country which had a significant change in diversity was Scotland. In 2000 the value of D was 0.767 rising to 0.915 in 2002. The implementation of the MCC vaccine in 1999 is likely to be a causal factor in this increase in diversity. This may be due to the reduction of the prevalent hypervirulent and serogroup C-associated ST-11 complex. The MCC vaccine campaign reduced the prevalence of this highly clonal lineage thus increasing overall diversity. To test this hypothesis, the vaccine-targeted ST-11 complex was omitted from the Scottish dataset and the index calculated again. While there was still a trend in increasing diversity, there was no significant difference between 2000 and 2002. This indicates that the vaccination programme impacted mainly on the target genotype. Other countries which showed upward trends in diversity, although non-significant, were Austria and Denmark while there were downward trends in the Czech Republic and Italy.

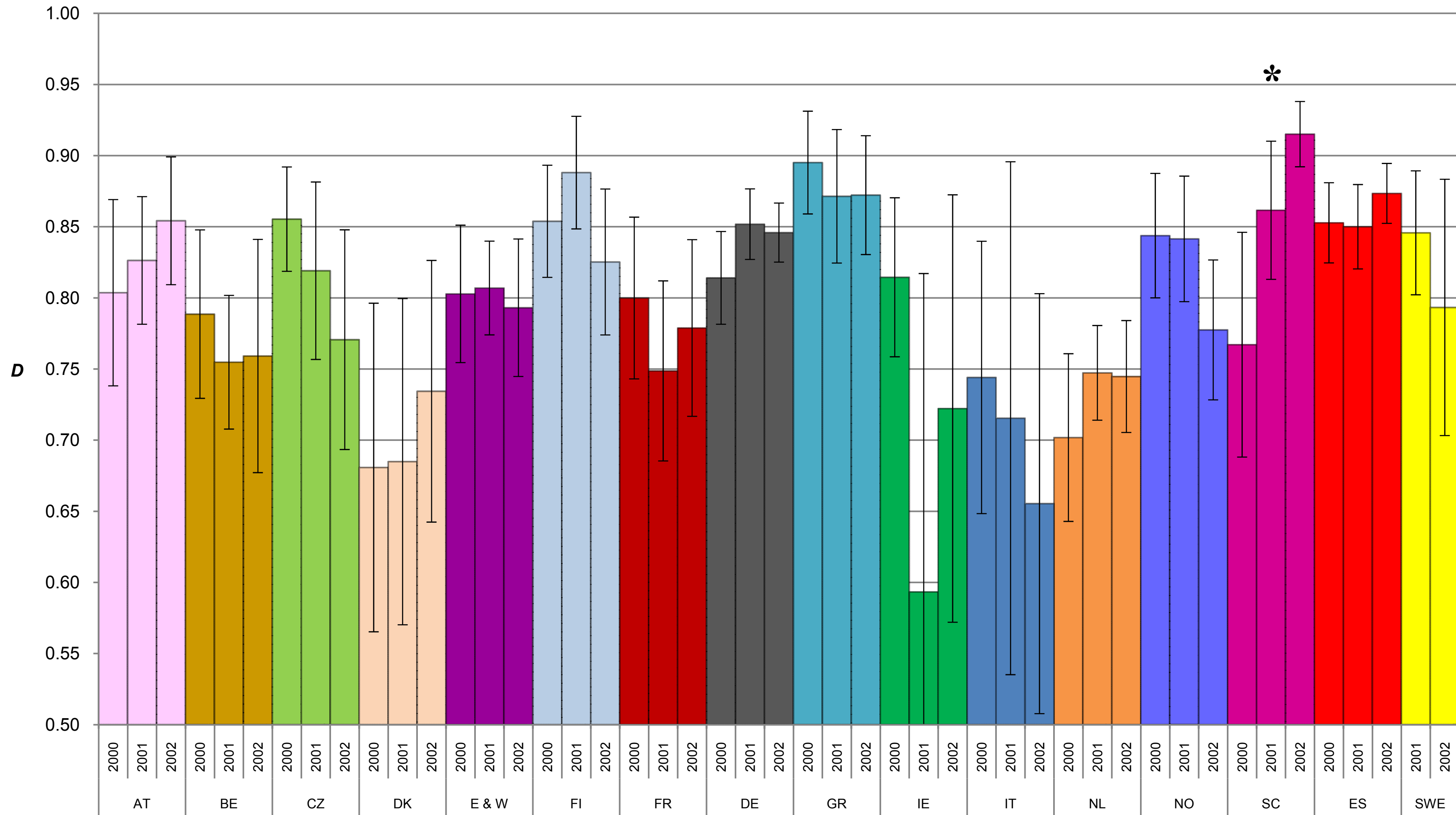


Figure 3.8: Change in index of diversity (*D*) in terms of ST for each country over the time period 2000–2002 in terms of clonal complex. Note: Since there were isolates only for a single year for Iceland and Portugal these data are omitted. *A significant change in *D* is indicated by non-overlapping CIs.

Similarly, the standardised index of association (I^S_A) gives an indication of levels of clonality and therefore diversity as it measures the effect of recombination events on a population. An I^S_A value of zero indicates frequent recombination and linkage equilibrium (random association of alleles) and therefore a non-clonal population. In the absence of recombination there is non-random association of alleles (linkage disequilibrium) and therefore an expected I^S_A value that differs significantly from zero. An I^S_A value towards one indicates increasing clonality in a population. The overall I^S_A value in Europe remained relatively constant over the three years (Table 3.5). It differed significantly from zero indicating a degree of clonality in the population. This would be expected from a disease population with a few dominant clones in the midst of a number of diverse genotypes.

The country with the highest I^S_A value and therefore most clonal was Iceland (0.989). It must be noted however that there was data only for the year 2001 for this country. The country with data for the three years with the highest I^S_A value was Denmark (0.771). The lowest overall I^S_A value and therefore the least clonal was Spain (0.470). Scotland was the only country which showed a trend of decrease in I^S_A value over the three years (0.763 in 2000, 0.487 in 2001 and 0.463 in 2002). This has been noted previously (Diggle & Clarke, 2005) and indicates an increase in diversity in the population. This concurs with the significant increase in the index of diversity from 2000 to 2002. The increase in diversity/reduction in clonality is speculated to be due to the implementation of the MCC vaccine in 1999 and the replacement in the niche occupied by the ST-11 clone by numerous other STs. As was done in the diversity index analysis, the test was run again with the ST-11 complex isolates omitted to test the effect on the I^S_A values. The result was that there was no downward trend in I^S_A and again points to the reduction in ST-11 complex in Scotland increasing the diversity of genetic types. Other countries (Czech Republic, France, Netherlands and Norway) had trends of increases in I^S_A over the three year period. Overall, for each country for each year the value of I^S_A was significantly different to zero and therefore showed significant linkage disequilibrium and therefore levels of clonality. This is in keeping with the concept of the meningococcal disease landscape being populated by a diverse range of genotypes but dominated by a number of highly prevalent clones.

Table 3.5: 'Standardised' Index of association (I^S_A) for each country per year. An I^S_A of zero indicates frequent recombination and linkage equilibrium (random association of alleles) and therefore a non-clonal population. An I^S_A value close to one indicates increasing clonality in a population.

	I^S_A in:			
country	**2000**	**2001**	**2002**	**total**
Austria	0.609	0.616	0.523	0.573
Belgium	0.657	0.694	0.648	0.668
Czech Rep.	0.568	0.704	0.767	0.700
Denmark	0.826	0.818	0.719	0.771
England & Wales	0.556	0.562	0.502	0.542
Finland	0.617	0.439	0.511	0.511
France	0.682	0.702	0.733	0.708
Germany	0.540	0.602	0.589	0.573
Greece	0.484	0.613	0.555	0.548
Iceland	n/a	0.989	n/a	0.989
Ireland	0.632	0.485	0.612	0.564
Italy	0.636	0.425	0.524	0.542
Netherlands	0.600	0.675	0.691	0.650
Norway	0.596	0.642	0.699	0.640
Portugal	n/a	0.595	n/a	0.595
Scotland	0.763	0.487	0.463	0.560
Spain	0.508	0.586	0.398	0.470
Sweden	n/a	0.654	0.669	0.661
total	0.544	0.589	0.548	0.561

note: significant linkage disequilibrium found for all countries and years.

3.2.5 Analysis of temporal and geographic structuring in Europe

While the major disease-associated clonal complexes were found spread across Europe over the three years, there were some differences in prevalence among countries and some complexes were more associated with some countries than others (Figure 3.4). To measure the extent of geographic and temporal genetic structuring in Europe more extensively, Wright's F-statistics were employed. F_{ST} values were calculated by AMOVA among the subpopulations (country/year) by using allele designations, allele sequences, concatenated allele sequences, allelic profiles and STs. AMOVA F_{ST} values give an indication of the contribution of geography/time to genetic/allelic structuring. Pairwise comparisons between subpopulations were also carried out to calculate F_{ST} values measuring gene flow between particular subpopulations (countries/years). These tests were carried out in the ARLEQUIN program. The program also ran statistical tests alongside the F_{ST} tests giving a p-value to determine if the values obtained are significantly different from the null hypothesis of no difference between populations, *i.e.* no structuring. For AMOVA a p-value less than 0.05 (or the p-value after Bonferroni correction) indicates a significant difference among/between populations and therefore presence of population structuring. For a pairwise comparison, a p-value more than 0.05 (or the p-value after Bonferroni correction) indicates no significant difference between populations and therefore indicates presence of gene flow between the two populations, *e.g.* two countries.

3.2.5.1 Geographic structuring

AMOVA provided evidence of geographic structuring as there were statistically significant F_{ST} values for the nucleotide sequences of each of the seven loci, the concatenated loci, allele designations, allelic profiles and STs among the 18 countries (Table 3.6) (all $p<0.00001$; significant at $p<\alpha$ where $\alpha=0.007$ after Bonferroni correction). The observed F_{ST} value for the concatenated MLST loci showed that almost 4.4% of genetic structuring of concatenated MLST sequences could be attributed to differences among geographical locations. Similar values were obtained for each of the house-keeping loci sequences, allele designation and allelic profiles. ST gave an F_{ST} value indicating that 2.8% of ST structuring was due to geographic location.

Table 3.6: AMOVA for isolate data grouped geographically, performed on nucleotide sequences (allele and concatenated alleles), allele designations, allelic profiles and STs.

data	F_{ST}	p-value
Nucleotide		
abcZ	0.04857	<0.00001
adk	0.03880	<0.00001
aroE	0.04831	<0.00001
fumC	0.04913	<0.00001
gdh	0.04094	<0.00001
pdhC	0.03498	<0.00001
pgm	0.03788	<0.00001
Concatenated loci	0.04397	<0.00001
Allele		
abcZ	0.04225	<0.00001
adk	0.03975	<0.00001
aroE	0.04202	<0.00001
fumC	0.04535	<0.00001
gdh	0.04223	<0.00001
pdhC	0.03938	<0.00001
pgm	0.04240	<0.00001
Allelic profile	0.04174	<0.00001
Sequence type	0.02862	<0.00001

Pairwise comparisons of the 18 countries using concatenated MLST locus sequences obtained F_{ST} values (Table 3.7) which were significantly different from the null hypothesis of no difference between the populations, *i.e.* p>0.05. This means that there was significant subdivision among countries. However, the only pair of countries for which there was no significant difference (although borderline), therefore indicating gene flow between them, was Norway and Sweden (p=0.05545). A similar outcome was obtained for pairwise comparisons of STs and allelic profiles (Appendix; Tables 4 and 5; Figure 3.9). It is interesting to note that Norway and Sweden were the only country pair that were indistinguishable for all three pairwise tests.

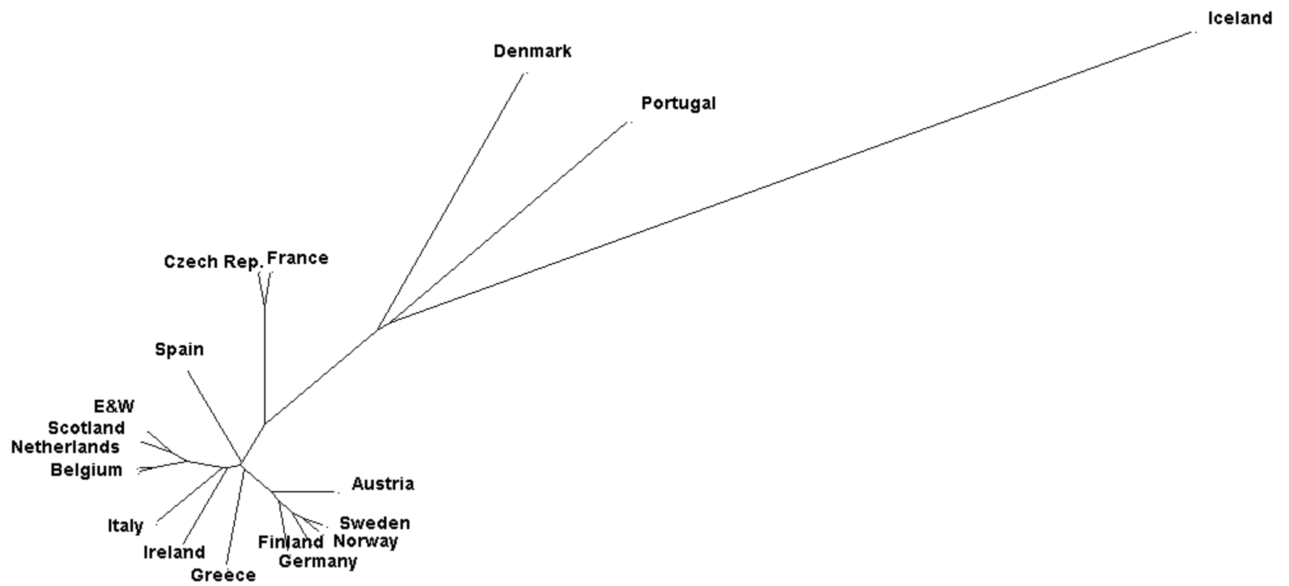


Figure 3.9: UPGMA tree constructed in SPLITSTREE using a distance matrix of F_{ST} values of pairwise comparisons of STs in 18 countries representing population structure in Europe. Note: Iceland and Portugal had data from one year only (2001). Note: E & W is England and Wales.

Table 3.7: F_{ST} values of pairwise comparisons of concatenated MLST nucleotide sequences in 18 European countries.

	Austria	Belgium	Czech Rep.	Denmark	England & Wales	Finland	France	Germany	Greece	Iceland	Ireland	Italy	Netherlands	Norway	Portugal	Scotland	Spain
Belgium	0.02196																
Czech Rep.	0.03311	0.02229															
Denmark	0.05637	0.08906	0.10133														
E & W	0.02198	0.01014	0.02334	0.07912													
Finland	0.01639	0.02650	0.04435	0.05046	0.01695												
France	0.03911	0.01602	0.00829	0.10909	0.02775	0.05255											
Germany	0.00935	0.01918	0.02868	0.04010	0.02069	0.01251	0.03201										
Greece	0.02443	0.03417	0.04434	0.06557	0.02039	0.01923	0.05484	0.02049									
Iceland	0.13043	0.09679	0.05977	0.21529	0.12019	0.15844	0.04455	0.12145	0.15986								
Ireland	0.02645	0.02889	0.03835	0.08212	0.01436	0.02118	0.05015	0.02336	0.02849	0.14790							
Italy	0.01578	0.02529	0.02684	0.07499	0.02071	0.02161	0.03666	0.01680	0.02860	0.13423	0.02249						
Netherlands	0.01751	0.00409	0.02144	0.06885	0.00806	0.01809	0.02198	0.01309	0.02627	0.10331	0.02084	0.01577					
Norway	0.01658	0.03249	0.03464	0.02210	0.02868	0.01792	0.04037	0.00701	0.02664	0.12553	0.03452	0.02777	0.02228				
Portugal	0.06420	0.08219	0.09933	0.11866	0.07104	0.05611	0.10355	0.04611	0.06055	0.24978	0.06778	0.06426	0.07396	0.07605			
Scotland	0.02136	0.01586	0.01455	0.08668	0.00772	0.02209	0.02182	0.01943	0.02504	0.09849	0.01996	0.01862	0.01335	0.02875	0.06800		
Spain	0.02454	0.02962	0.04106	0.07749	0.02039	0.01809	0.04824	0.01888	0.02464	0.15671	0.02555	0.02255	0.02459	0.03537	0.02697	0.02314	
Sweden	0.01825	0.03234	0.04689	0.02474	0.02568	0.00619	0.05204	0.00985	0.02480	0.15425	0.03205	0.02871	0.02188	**0.00551***	0.07252	0.03256	0.03009

bold* p>0.05 can accept null hypothesis of no difference between populations, *i.e.* presence of gene flow between two countries. Note: E & W is England and Wales.

3.2.5.2 Temporal Structuring

AMOVA provided some evidence of temporal structuring in Europe for the years of the study 2000, 2001 and 2002. There were statistically significant F_{ST} values (after Bonferroni correction) for the nucleotide sequences of the seven concatenated loci, *pgm* locus, *aroE*, *pdhC* and *pgm* allele designations, allelic profiles and STs (all p<0.007) among the three years (Table 3.8). The observed F_{ST} value for the concatenated MLST loci showed that almost 0.1% of genetic structuring could be attributed to differences among the three years.

Table 3.8: AMOVA for isolate data grouped temporally (years 2000, 2001, 2002), performed on nucleotide sequences (allele and concatenated alleles), allele designations, allelic profiles and STs.

data	F_{ST}	p-value
Nucleotide		
abcZ	0.00027	0.20594*
adk	0.00092	0.04980*
aroE	0.00112	0.02139*
fumC	0.00113	0.01644*
gdh	0.00059	0.08119*
pdhC	0.00064	0.05386*
pgm	0.00212	0.00099
Concatenated loci	0.00099	0.00396
Allele		
abcZ	0.00033	0.13535*
adk	0.00081	0.04129*
aroE	0.00115	0.00277
fumC	0.00072	0.02218*
gdh	0.00071	0.04198*
pdhC	0.00129	0.00079
pgm	0.00170	0.00030
Allelic profile	0.00096	0.00396
Sequence type	0.00067	0.00297

*non-significant at p>α where α=0.007 after Bonferroni correction

F_{ST} values of pairwise comparisons for each of the three years using concatenated and allele nucleotide sequences, allele designations, allelic profiles and STs showed some genetic structuring. There was a significant difference between the populations of 2000 and 2001 according to the sequence of the *pgm* locus, the allele designations of *aroE*, *pdhC* and *pgm* and STs (Table 3.9). This highlights possible differences in evolutionary constraints acting on each of the different house-keeping genes.

F_{ST} values of pairwise comparisons of each year using concatenated MLST nucleotide sequences were calculated to assess the level of temporal structuring in each country. There was some evidence of temporal structuring in some countries (Table 3.10). Highly significant F_{ST} values were found ($p<0.005$), indicating genetic differentiation, in the Netherlands between 2000 and 2001 and 2000 and 2002, in Scotland between 2000 and 2002 and 2001 and 2002 and in Spain between 2000 and 2002. There was also significant differentiation ($p<0.05$) in Belgium between 2001 and 2002, in the Republic of Ireland between 2000 and 2001 and 2000 and 2002 and in Spain between 2000 and 2001. Similar values were obtained for pairwise comparisons of the years using STs and allelic profiles (Tables 6 and 7, Appendix). Vaccine strategies were implemented to a greater or lesser degree in the aforementioned countries during this time period (Table 1.1). In the Netherlands and Belgium MCC vaccination began in the second half of 2002, in the Republic of Ireland and Spain (in 14 of the 17 autonomous regions) it began late 2000, and in the UK it began at the end of 1999. ST comparisons also showed a highly significant difference in Germany between 2000 and 2002 and 2001 and 2002. A shift in the distribution of STs over the three years such as the significant reduction in ST-66 from 2000 to 2002 ($p<0.005$) may have contributed to this temporal structuring of STs.

Table 3.9: Temporal structuring in Europe. F_{ST} values of pairwise comparisons of each year using concatenated MLST locus nucleotide sequences, allele sequences, allele designations, allelic profiles and STs.

	F_{ST}		
data	**2000–2001**	**2000–2002**	**2001–2002**
Nucleotide			
abcZ	0.00032*	0.00038*	0.00013*
adk	0.00106*	-0.00043*	0.00183*
aroE	0.00096*	0.00216*	0.00042*
fumC	0.00155*	0.00078*	0.00105*
gdh	0.00051*	0.00043*	0.00079*
pdhC	0.00049*	0.00006*	0.00120*
pgm	0.00283	0.00222*	0.00144*
Concatenated loci	0.00107*	0.00118*	0.00076*
Allele			
abcZ	0.00026*	-0.00019*	0.00083*
adk	0.00040*	0.00060*	0.00132*
aroE	0.00197	0.00081*	0.00071*
fumC	0.00100*	0.00040*	0.00074*
gdh	0.00101*	-0.00026*	0.00121*
pdhC	0.00187	0.00131*	0.00078*
pgm	0.00230	0.00181*	0.00111*
Allelic profile	0.00126*	0.00064*	0.00095*
Sequence type	0.00118	0.00030*	0.00054*

*non-significant at $p>\alpha$ where $\alpha=0.007$ after Bonferroni correction *i.e* absence of gene flow.

Table 3.10: Temporal structuring in Europe. F_{ST} values of pairwise comparisons of each year using concatenated MLST nucleotide sequences in 16 European countries.

country		F_{ST} values	
		2000	**2001**
Austria	**2001**	0.00966*	
	2002	0.00257*	-0.00044*
Belgium		0.00044*	
		-0.00040*	0.01888
Czech Rep.		-0.00177*	
		0.01485*	0.00257*
Denmark		0.02061*	
		-0.00505*	0.00845*
England & Wales		-0.00500*	
		-0.00402*	0.0003*
Finland		0.0166*	
		0.00281*	-0.00101*
France		0.00323*	
		0.00383*	0.0005*
Germany		-0.00027*	
		0.00515*	0.00429*
Greece		0.00705*	
		0.00667*	0.01073*
Ireland		0.03561	
		0.04875	-0.01398*
Italy		-0.00527*	
		-0.01502*	-0.00381*
Netherlands		0.02428	
		0.02090	0.00277*
Norway		-0.00881*	
		-0.00924*	-0.00963*
Scotland		0.00722*	
		0.05849	0.02323
Spain		0.01451	
		0.05012	0.01214
Sweden		n/a	
		n/a	-0.00376*

* p>0.05 cannot accept null hypothesis of no difference between populations, *i.e.* presence of gene flow. n/a = data not available for Sweden year 2000. Note: Data not available for 2000 and 2002 for Iceland and Portugal

3.3 Discussion

The aim of this study was, by using MLST, to assess the geographical and temporal distribution of invasive meningococci genetic types in 18 European countries over a three year period and determine the presence, if any, of temporal and geographic structuring. Analysis of a large and representative sample of 4048 disease-associated meningococci demonstrated the large overall diversity of genotypes in Europe (978 STs) but also the over-representation of particular clones. The results have shown the dominance of the major clonal complexes, their wide spread across many countries of Europe and their persistence over a three year period. This is consistent with the results of a study looking at disease and carriage isolates from 1991–2000 in three European countries (Czech Republic, Greece, Norway) which found the predominance of the major invasive clones (ST-11, ST-32, ST-41/44) in disease cases in each country (Yazdankhah *et al.*, 2004). The overall carriage population of these three countries was shown to be much more diverse relative to the disease population. Also in that study, the prevalence of each complex varied according to country, with some complexes being more prevalent in particular countries which reflects the results here. For example, the ST-162 complex in Greece was found a much higher prevalence than in the Czech Republic and Norway. This complex was also found to be relatively more important in Greece than any other European country in this study also.

Geography, more so than time, was a significant contributor to diversity of the European disease dataset, although it must be tempered by the fact that the time period of the study was relatively short so any major shifts in genotype distribution may not be picked up. This indicates that transmission of the meningococcus is more likely to occur within than among countries. Significant geographic structuring has been documented previously for carriage isolates within a single country (Germany) (Claus *et al.*, 2005) and for carriage and disease isolates among three European countries (Yazdankhah *et al.* , 2004). Conversely, a study of Czech carriage isolates showed no evidence of geographically restricted gene flow being the basis of structuring (Jolley *et al.*, 2005). The reasons for differences in prevalence of different genotypes in different countries are likely to be complex. A possibility is that the rate of spread of lineages across the continent is slow enough that geographic structuring is apparent. Different socio-economic and demographic factors such as age-profile, population density,

vaccine schedules, antibiotic usage and carriage risk factors such as rates of respiratory disease and smoking may also affect genotype distribution in the disease and carriage populations. Countries with vaccine programmes may have meningococcal populations which are directly impacted as result of the immunisation itself and indirectly through herd immunity, *e.g.* the UK (Trotter *et al.* , 2004). Despite variation in prevalence in different countries, the major clonal complexes, *i.e.* ST-11, ST-41/44, ST-32, ST-8 and ST-269, still together accounted for the majority of disease isolates (77%), and were very much the foremost complexes in each country in Europe. The ST-11, ST-41/44 and ST-32 complexes were found in all 18 countries, while ST-8 complex was found in 14 of the 18 countries and ST-269 complex was found in 17 of the 18 countries. These results further demonstrate the importance of the unit of the clonal complex in disease-association. These lineages have been shown to transmit successfully despite causing disease and are capable of global spread.

In this study, there was a signal of association between genotypes and geographical location with a significant F_{ST} value for concatenated MLST loci indicating that 4.4% of genetic diversity could be attributed to differences between geographical locations. Noteworthy exceptions to the restricted gene-flow amongst countries were Norway and Sweden which showed no genetic differentiation among them by pairwise comparisons of allelic profile, ST and concatenated MLST sequence. The close physical proximity, similar cultural and socio-economic conditions and free movement of people between these two countries are likely to be factors in this genetic similarity. Intriguingly however, while there was a sharp increase and ongoing epidemic of serogroup B-related disease in Norway from the mid-1970s onwards, there was no such occurrence in Sweden (Peltola *et al.* , 1982). Another Scandinavian country, Finland, is also noteworthy in the significant proportion of unassigned STs found there. Also, a relatively large proportion (26.5%) of STs were unique to this country. Almost one third of the unassigned STs were of the ST-2793 genotype which also had two single-locus variants ST-2690 and ST-2997. This could be an emerging clone that has yet to spread further. Finland is relatively isolated geographically and also shares extensive borders with Russia across which there is free movement. Therefore, Russia may be a potential source of novel genotypes not found in other European countries. Russia has quite a different epidemiology compared to European countries with a large proportion of disease accounted for by serogroup A (Achtman *et al.*, 2001; Koroleva *et al.*, 1998) which is now rare in Europe. In concordance with the

very low occurrence of serogroup A disease in Europe in comparison to other serogroups (EU-IBIS, 2006), clonal complexes associated with this serogroup were rare in the 18 countries of this dataset. Five ST-5 clonal complex isolates found in total: four ST-7; three in France and one in Norway; and one ST-5, found in Denmark. There were no ST-4 complex isolates found in the dataset.

Meningococcal clonal complexes, which are clusters of closely related organisms, were first revealed by population studies using MLEE (Caugant *et al.* , 1987b). Subsequent investigation using MLST has shed light on the population structure of the meningococcus and the presence of dominant genotypes in a highly diverse disease population (Maiden *et al.*, 1998; Yazdankhah *et al.*, 2004). It has been shown that clonal complexes can be found over a wide geographic spread and a relatively long time period (Caugant, 1998). This is in spite of frequent recombination in the meningococcus which would be expected to increase genetic novelty and diversity over time. However, models of strain structure indicate that maintenance of this persistent clonal complex structure is consistent with immune selection as the main driving force rather than simply the oversampling of epidemic clones in a neutral microepidemic model (Buckee *et al.* , 2008; Fraser *et al.* , 2005). Slight differences in transmission fitness among lineages give rise to a small number of persistent genotypes that continue to generate and out-compete less successful variants over time. This gives rise to the clonal complex structure with the founder central genotype and its cloud of related variants. Some complexes are more clonal than others. In this study some complexes were made up mostly of a single ST, often the central genotype of the complex, and others made up of many STs. For example, the ST-11 complex was dominated by its central genotype while the ST-254 complex had almost as many genotypes as there were isolates. ST-11 is strongly associated with hyperinvasiveness and found at low rates in carriage (Yazdankhah *et al.*, 2004). Therefore, it may be able to tolerate the penalty of high invasiveness because of its high transmissibility and may be less diverse having spent relatively less time exposed to host immunity because of shorter duration of carriage.

Certain complexes, the hyperinvasive lineages, while having low carriage prevalence are able to transmit more successfully and may cause disease. While the mechanism of their invasiveness is incompletely understood, although being capsulate is required, the successful spread and establishment of a genotype in a population is down to a combination of various factors such as transmission, population susceptibility

and herd immunity. For example, the introduction of a novel genotype into a new and susceptible population might cause it to increase rapidly in the carriage population and subsequently cause a rise in disease. Since disease is not useful for the spread of the organism, transmission fitness is thought to be the key factor in the spread of the hyperinvasive lineages. It has been speculated that hyperinvasive complexes such as ST-11 are relative recent evolutionarily and have attributes that allow for increased transmission particularly in naïve populations (Jolley *et al.*, 2005). They rise swiftly in frequency and are dominated by a central genotype (the majority of ST-11 complex is made up of its central genotype). With increasing host immunity in the population the selective advantage of the novel variant decreases thus preventing any single variant dominating the entire population, therefore increasing diversity.

The dominant complexes were persistent over the three year time period although this length of time would probably be insufficient to detect any large and long-term shifts in continent-wide genotype distribution. Despite this, some significant changes were noted such as the highly significant decrease in the ST-8 complex. This appears to be driven by the large drop in this complex in Spain and its virtual disappearance since 2000 in countries such as England and Wales, Greece, Republic of Ireland, and Scotland. The implementation of routine MCC vaccination in these countries during this time and, in Greece, an unofficial single dose scheme adopted by most paediatricians (Kafetzis *et al.*, 2007), is a likely explanation for this decrease since ST-8 complex is strongly associated in European disease with serogroup C. A study of meningococcal disease in Europe which links serological data with EU-MenNet genotype data shows that 86% of this complex are serogroup C (Trotter *et a l.*, 2007). The MCC vaccine also appears to have prompted the reduction of the ST-11 complex in countries that implemented it, *i.e.* England and Wales, Greece, Republic of Ireland, and Scotland. The ST-11 complex is also highly associated with serogroup C disease in Europe; 83% of European ST-11 complex isolates are serogroup C (Trotter *et al.* , 2007). Spain on the other hand, despite also having introduced the vaccine, actually had a significant increase in the ST-11 complex between 2000 and 2002 ($p<0.05$). However, this may be due to the effect of capsule switching from serogroup C to B for which there is some evidence in Spain. Strains of B:2a:P1.5, ST-11 complex have been isolated in the Basque region which appear to have replaced C:2a:P1.5, ST-11 complex strains following the immunisation introduction in 2000 (Cano *et a l.*, 2004; Perez-Trallero *et al.* , 2002). In Italy in 2000, in the absence of any vaccine campaign, four

serogroup B ST-11 complex isolates were found that had phenotypes normally associated with serogroup C (Stefanelli *et al.*, 2003). The MCC immunisation campaign was introduced in the UK in 1999 to interrupt the sharp increase in serogroup C 2a:P1.5,2 ST-11 complex strain disease but as yet there has been no evidence of capsule replacement (Trotter *et al.*, 2006b). Therefore, the effect of vaccination campaigns in inducing capsule replacement is still unclear and further surveillance is required to detect it should it occur. Interestingly the only country to show a significant increase in diversity in terms of ST and a decrease in I^S_A indicating a decrease in clonality between 2000 and 2002 was Scotland. This may indicate that in a relatively short period of time the MCC vaccine eliminated the niche occupied by ST-11 complex allowing it to be taken over by a more diverse set of genotypes.

In summary, despite a large range of genotypes found in European disease isolates over three years, the principal hyperinvasive lineages accounted for most disease for the duration. These dominant types were found spread across Europe and, over the time period, remained the most dominant types in each country. Some changes in the distribution of certain clonal complexes were noted over the three years. It is therefore important to maintain continued surveillance across Europe to identify any shifts in the distribution of types, the emergence of any new clones or possible capsule-switching. While there was a predominance of major clonal complexes across Europe there was still evidence of geographical structuring. This shows that location does have an impact on meningococcal diversity in Europe and that transmission is likely more important within countries than among them. Temporal structuring also seemed to play some part in meningococcal diversity although to a lesser degree. However, the three year time period would probably be insufficient to detect any large-scale and lasting shifts in genotype distribution.

CHAPTER 4: Distribution and structuring of potential vaccine antigens PorA and FetA in Europe

4.1 Introduction

Because of the difficulty in producing a vaccine against the serogroup B polysaccharide capsule, a variety of alternative antigens have been assessed. The sub-capsular OMPs in particular are considered prime contenders and have been subject to investigation as vaccine candidates (Jodar *et al.*, 2002). PorA and FetA are two of the leading candidates due to their presence in OMVs and their immunogenicity (Wedege *et al.*, 1998). For this reason, the distribution of these antigens in the population is of particular interest to those formulating vaccines that contain them. The diversity of these two proteins is exploited serologically and genotypically as a means of discriminating between isolates in routine surveillance and epidemiological investigation. PorA and FetA genotyping is now becoming standard in nomenclature (Jolley *et al.*, 2007) and they are included in the classification schemes of the ECDC's TESSy European surveillance system (http://www.ecdc.europa.eu/en/activities/surveillance/Pages/Surveillance_Tessy.aspx) and the real-time EMERT database (http://emgm.eu/emert/). Addition of antigen typing to serogroup and/or MLST information can add a greater degree of discrimination to population biology and epidemiological investigation. For example, finetype data (serogroup: PorA: FetA) of invasive disease isolates collected in Germany over 42 months was used to retrospectively detect clusters of disease cases using automated scan statistics (Elias *et al.*, 2006).

PorA is among the most examined of all meningococcal vaccine candidate antigens because of its abundance in the outer membrane and its high immunogenicity. It has been used as the major component of OMV vaccines targeting single clone epidemics in Cuba, Norway and New Zealand with some success (Bjune *et al.*, 1991; O'Hallahan *et al.*, 2005; Rodriguez *et al.*, 1999). Because of the high variability of both PorA and FetA, such single-component vaccines may have limited coverage in an endemic setting. As a result, vaccines are being developed which contain multiple antigen variants for maximum coverage (Oster *et al.*, 2005; van den Dobbelsteen *et al.*,

2007). Previous surveys of the diversity of these antigens (Russell *et al.*, 2008; Urwin *et al.*, 2004) have shown structuring in their distribution, in that a small number of types dominate and can persist over time and distance. Also notable was the presence of non-random structuring of antigens which is consistent with host immune selection (Gupta *et al.*, 1996) and implies a degree of immunogenicity of the antigens. Associations with hyperinvasive lineages have also been shown. The implications of these phenomena are the potential simplification of vaccine formulation. It is therefore important to know the currently circulating antigen strains in the population before a vaccine is formulated to better estimate its theoretical impact.

This chapter details the distribution of the two vaccine candidates PorA and FetA in the meningococcal disease population in Europe over a three year period. This is the largest known survey of antigen genotypes in a European meningococcal disease collection. There was an appraisal of the levels of diversity of each antigen and their relationship with each other in terms of strain structuring and with clonal complex. Their distribution both geographically and temporally was also assessed. This leads into Chapter 5 which will outline the rationale behind a vaccine formulation based on the antigen distribution found and its theoretical impact.

4.2 Results

4.2.1 Genetic diversity of antigens in Europe

4.2.1.1 PorA

PorA type data were available for 3652 isolates of the EU-MenNet meningococcal disease isolate collection which includes data from 18 countries for the years 2000–2002 inclusive. There were only data for the year 2001 for Iceland and Portugal. There were 54 VR1 types, 130 VR2 types and 273 unique VR1/VR2 combinations, *i.e.* PorA types found. Eight PorA types accounted for 60% of isolates: P1.5,2; P1.7-2,4; P1.7,16; P1.5-1,10-8; P1.19,15; P1.18-1,3, P1.22,14 and P1.5-1,10-4 (Figure 4.1, Table 4.1). There were 146 PorA types found only once (0.04%). Four VR1 types accounted for 61% of isolates: P1.5, P1.7-2, P1.5-1 and P1.7. There were 19 VR1 types that were found only once (0.005%). Six VR2 types accounted for 61% of isolates: P1.2, P1.4, P1.16, P1.10-8, P1.14 and P1.15. There were 54 VR2 types that were found only once (0.01%).

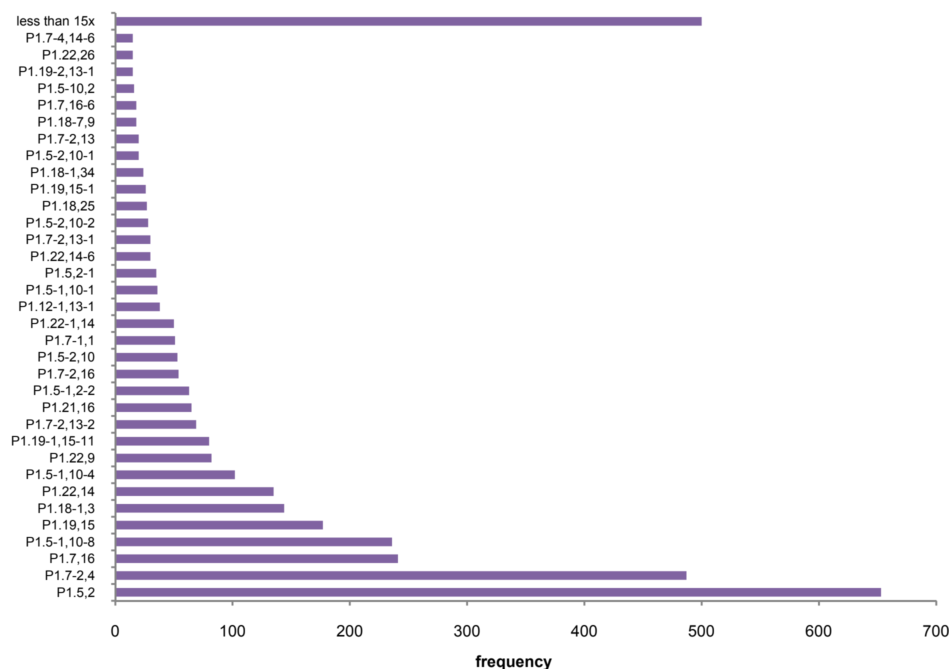


Figure 4.1: Rank-abundance plot illustrating PorA diversity in the European disease population. A total of 273 PorA types were found in the dataset. PorA types with a frequency less than 15 were grouped.

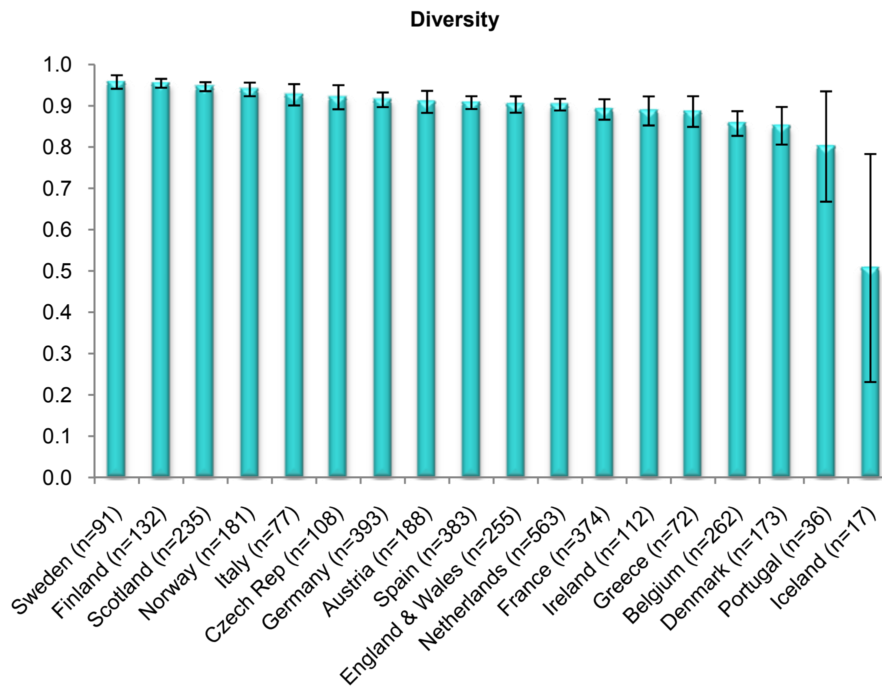
Table 4.1: Breakdown of PorA and VRs in EU-MenNet dataset.

PorA	number	%	VR1	number	%	VR2	number	%
P1.5,2	653	17.9	**5**	719	19.7	**2**	678	18.6
P1.7-2,4	487	13.3	**7-2**	704	19.3	**4**	506	13.8
P1.7,16	241	6.6	**5-1**	467	12.8	**16**	389	10.6
P1.5-1,10-8	236	6.5	**7**	335	9.2	**10-8**	239	6.5
P1.19,15	177	4.8	**22**	271	7.4	**14**	208	5.7
P1.18-1,3	144	3.9	**19**	231	6.3	**15**	192	5.3
P1.22,14	135	3.7	**18-1**	205	5.6	**3**	149	4.1
P1.5-1,10-4	102	2.8	**5-2**	121	3.3	**9**	123	3.4
P1.22,9	82	2.2	**19-1**	88	2.4	**10-4**	108	3.0
P1.19-1,15-11	80	2.2	**21**	83	2.3	**13-1**	104	2.8
P1.7-2,13-2	69	1.9	**12-1**	65	1.8	**15-11**	83	2.3
P1.21,16	65	1.8	**7-1**	63	1.7	**13-2**	81	2.2
P1.5-1,2-2	63	1.7	**22-1**	57	1.6	**1**	69	1.9
P1.7-2,16	54	1.5	**18**	44	1.2	**2-2**	67	1.8
P1.5-2,10	53	1.5	**17**	26	0.7	**10**	60	1.6
P1.22-1,14	51	1.4	**19-2**	25	0.7	**10-1**	59	1.6
P1.7-1,1	51	1.4	**18-7**	22	0.6	**14-6**	46	1.3
P1.12-1,13-1	38	1	**7-4**	20	0.5	**2-1**	37	1.0
P1.5-1,10-1	36	1	**5-10**	17	0.5	**13**	30	0.8
P1.5,2-1	35	1	**17-1**	10	0.3	**10-2**	29	0.8
P1.22,14-6	30	0.8	**21-7**	10	0.3	**15-1**	28	0.8
P1.7-2,13-1	30	0.8	**20**	8	0.2	**25**	28	0.8
P1.5-2,10-2	28	0.8	**18-3**	7	0.2	**34**	26	0.7
P1.18,25	27	0.7	**31**	7	0.2	**16-6**	19	0.5
P1.19,15-1	26	0.7	**<5**	47	1.1	**30**	18	0.5
P1.18-1,34	24	0.7	**total**	3652		**26**	17	0.5
P1.5-2,10-1	20	0.5				**16-29**	13	0.4
P1.7-2,13	20	0.5				**16-32**	12	0.3
P1.18-7,9	18	0.5				**10-7**	9	0.2
P1.7,16-6	18	0.5				**16-3**	9	0.2
P1.5-10,2	16	0.4				**23**	8	0.2
P1.19-2,13-1	15	0.4				**2-3**	7	0.2
P1.22,26	15	0.4				**3-8**	7	0.2
P1.7-4,14-6	15	0.4				**10-28**	7	0.2
P1.19,13-1	14	0.4				**13-9**	7	0.2
P1.7,16-32	12	0.3				**16-26**	7	0.2
P1.7,16-29	11	0.3				**2-12**	6	0.2
P1.7,30	10	0.3				**4-1**	6	0.2
<10	451	10.9				**10-6**	6	0.2
total	3652					**<5**	155	0.1
						total	596	

The genetic diversity of each country in terms of PorA types was calculated using Simpson's index of diversity (*D*) (Figure 4.2) (Simpson, 1949). 95% confidence intervals (CIs) were calculated as described by Grundmann *et al.* (Grundmann *et al.* , 2001). A significant difference in *D* is indicated by non-overlapping CIs. The evenness (*E*) value for each country was also calculated (see Materials and Methods). *E* gives a measure of the relative abundance of the different genotypes making up the richness, *i.e.* number of genotypes in a sample. Like *D* it ranges from zero to one; as *E* increases the more equal the contribution of each clone in terms of frequency becomes. The overall level of diversity of PorA types in Europe was high at 0.932 and there were no significant changes in diversity over the three years. The diversity of VR1 and VR2 were 0.884 and 0.918 respectively. VR2 was significantly more diverse than VR1 due to the larger number of VR2 alleles found. There were no significant temporal changes in diversity for either VR, but there was an upward trend in diversity for VR2 from 0.910 to 0.927.

There was a range in diversity across Europe (Figure 4.2) with Iceland being the least diverse (0.507) and Sweden the most diverse (0.958). It must be noted that there were just data from a single year (2001) for Iceland and Portugal. Of the countries with data for each of the three years, Denmark was the least diverse (0.852). It was significantly less diverse than Italy, Norway, Scotland, Finland and Sweden. The most diverse country Sweden was significantly more diverse than 12 other countries. It also had the highest *E* value (0.656), indicating that it had a more even distribution of types making up its population than other countries. This contrasts with the lowest value which was found in the Netherlands of 0.139 suggesting an uneven distribution of types in the country.

(a)



(b)

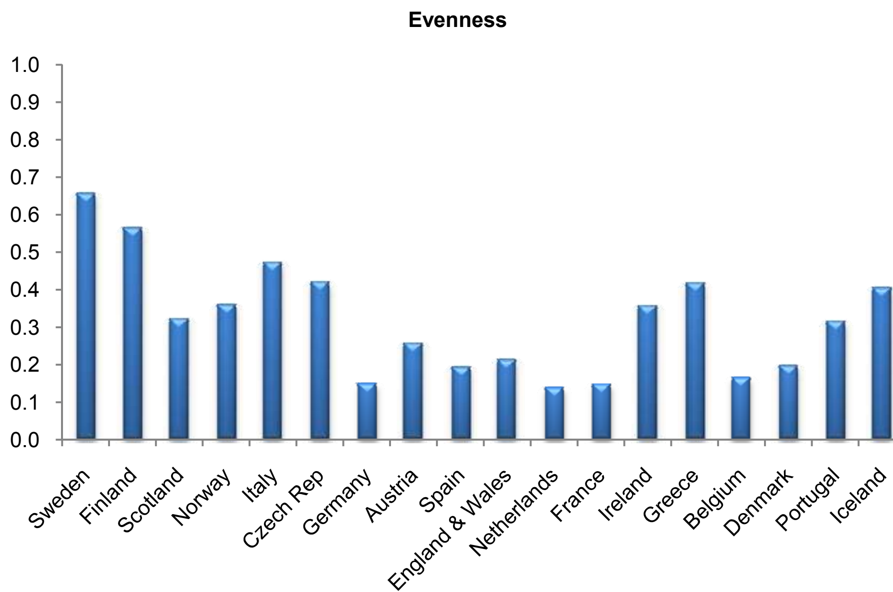


Figure 4.2: (a) Genetic diversity of each country in terms of PorA type ranked in order of decreasing Diversity value (*D*). A value of *D* near one indicates a high level of diversity, while a value near zero indicates a low level of diversity. Error bars indicate 95% CIs. Non-overlapping CIs indicates significant difference in *D*. (b) Evenness (*E*) is a measure of the relative abundance of each of the genotypes making up the population sample (in this case, country). As *E* increases, the frequency of each clone becomes more equal. Note: E & W is England and Wales.

4.2.1.2 FetA

FetA type data were available for 2949 isolates of the EU-MenNet disease collection for 16 countries for the years 2000–2002 inclusive. There were no data available from Scotland and Sweden and only data for the year 2001 for Iceland and Portugal. There were 99 FetA types found (Figure 4.3). Four types accounted for 56% of isolates: F1-5, F3-6, F5-1 and F3-3. There were 36 types which were found only once (0.01%).

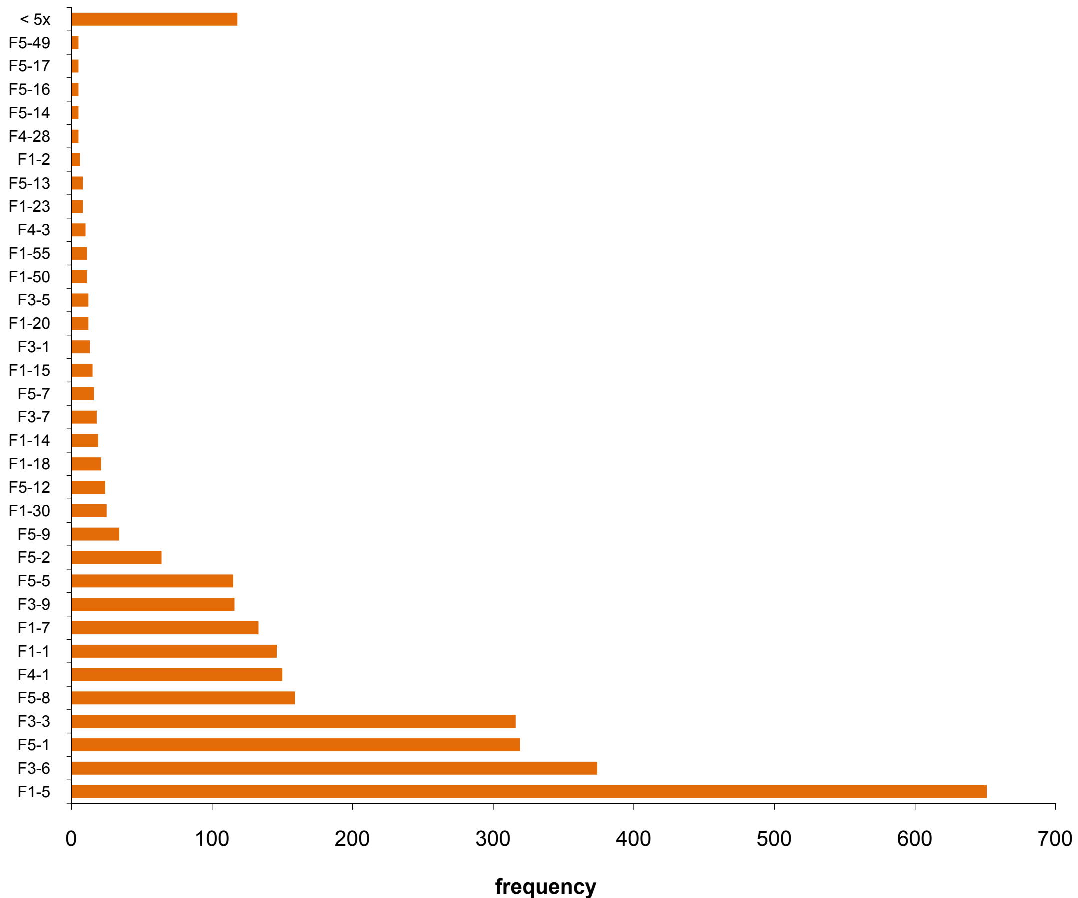
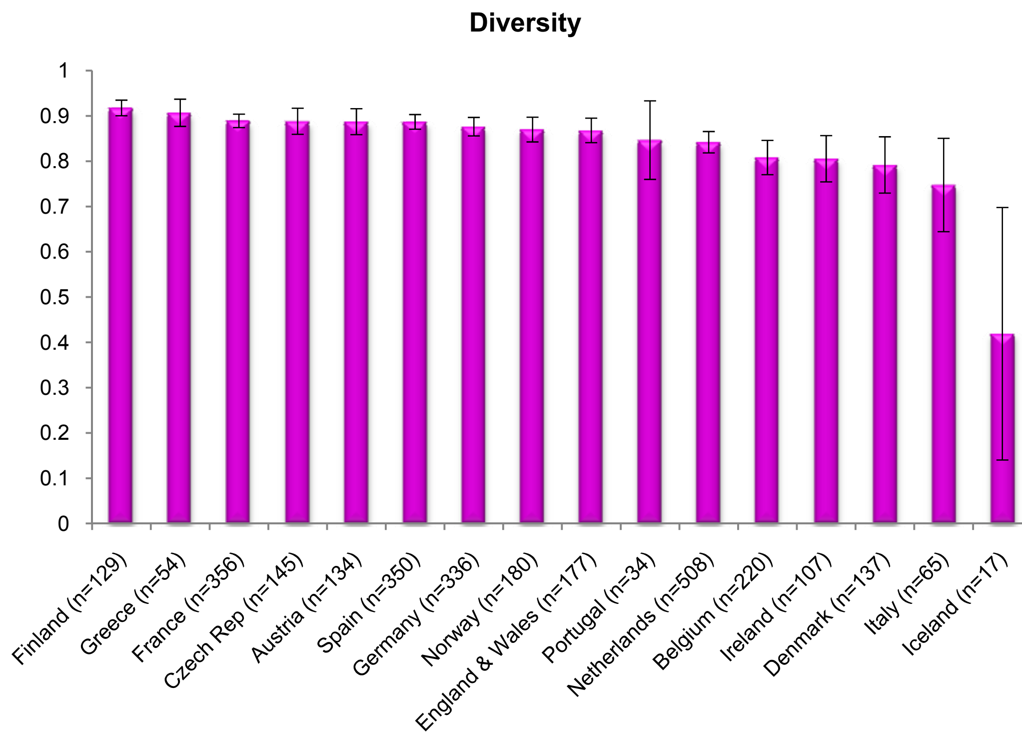


Figure 4.3: Rank-abundance plot illustrating FetA diversity in the European disease population. A total of 99 FetA types were found in the dataset. FetA types with a frequency less than five were grouped.

The genetic diversity of each country in terms of FetA type was calculated using Simpson's index of diversity with 95% CIs (Figure 4.4). The *E* value for each country was also calculated. The overall *D* value (0.898) was significantly less than for PorA and for each country *D* values were less than those for PorA types, which is unsurprising give the larger number of PorA alleles in existence. There were no significant changes in diversity over the three years. There were only data for the year 2001 for Iceland and Portugal. The country with data for each of the three years with the lowest *D* value (0.748) was Italy. It was significantly less diverse than Finland, Greece, France, Czech Republic, Austria, Spain and Germany. The most diverse was Finland (0.918). It was significantly more diverse than all countries except Greece, France, Czech Republic, Austria and Spain. Greece had the highest *E* value of 0.672 and as with PorA, the Netherlands had the lowest *E* (0.186) value again suggesting an uneven make up of antigenic types.

(a)



(b)

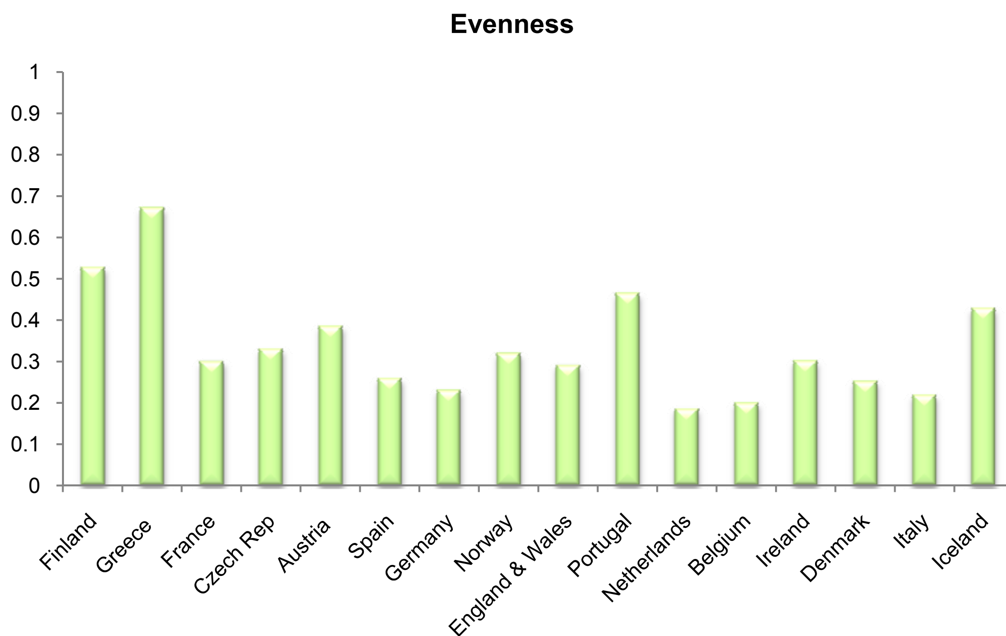


Figure 4.4: (a) Genetic diversity (*D*) of each country in terms of FetA type. A value of *D* near one indicates a high level of diversity, while a value near zero indicates a low level of diversity. Error bars indicate 95% CIs. Non-overlapping CIs indicates significant difference in *D*. (b) Evenness (*E*) is a measure of the relative abundance of each of the genotypes making up the population (in this case, country) sample. As *E* increases, the frequency of each clone becomes more equal.

4.2.2 Antigenic structuring in Europe

Clonal complex distribution across Europe revealed the dominance of the major lineages, if at different prevalences in different countries and a significant level of structuring (Chapter 3). By their very nature, antigens are more variable than house-keeping genes as a result of their exposure to the immune system, so a greater level diversity of types across Europe would be expected. To measure the extent of geographic and temporal genetic/allelic structuring, F_{ST}s were employed. F_{ST} values were calculated by AMOVA among the subpopulations (country/year) by using PorA/FetA allele designations. AMOVA F_{ST} values give an indication of the contribution of geography/time to genetic/allelic structuring. Pairwise comparisons between subpopulations were also carried out to calculate F_{ST} values measuring gene flow between particular subpopulations (countries/years). These tests were carried out in the ARLEQUIN program. The program also ran statistical tests alongside the F_{ST} tests giving a p-value to determine if the values obtained are significantly different from the null hypothesis of no difference between populations, *i.e.* no structuring. For AMOVA a p-value less than 0.05 indicates a significant difference among/between populations and therefore presence of population structuring. For a pairwise comparison, a p-value more than 0.05 indicates no significant difference between populations and therefore indicates presence of gene flow between the two populations, *e.g.* between two countries.

4.2.2.1 Geographic structuring

There was a dominance of particular antigen types across Europe but varying prevalences of particular types in different countries (Figures 4.5 and 4.6). Significant associations between antigen types and particular countries were observed (χ^2 or Fisher's exact test; $p<0.005$) (Tables 4.2 and 4.3). For example, PorA type P1.5,2 was associated with France; P1.7-2,4 with the Netherlands, England and Wales and Belgium (Table 4.2). Notable associations of types that were otherwise of low prevalence elsewhere, included the P1.5-10,2 type, all of which were found in France and the P1.7,16-32 type 11/12 of which were found in the Netherlands. For FetA types, significant associations included F1-5 and F3-6 with Belgium, F5-1 with Czech

Republic and England and Wales and F3-3 with Denmark (Table 4.3). Other notable associations included F3-5, F1-23 and F5-49 which were almost exclusively found in Finland. The F1-55 type was associated with Spain which harboured 10/11 found. F1-5 was significantly under-represented in Spain; F3-6 in Germany; F5-1 in France and F3-3 in Belgium.

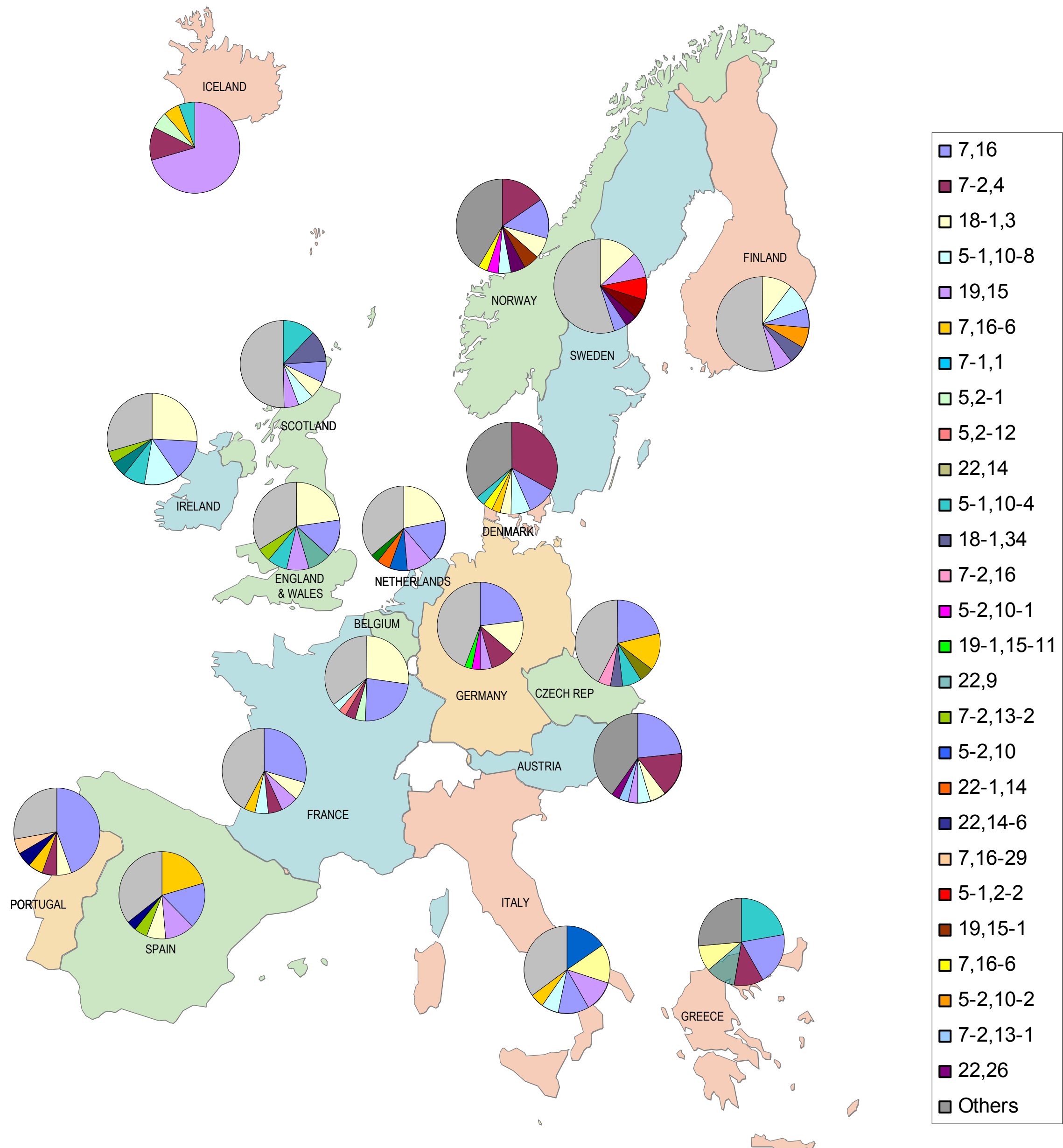


Figure 4.5: Distribution of PorA types in Europe 2000–2002.

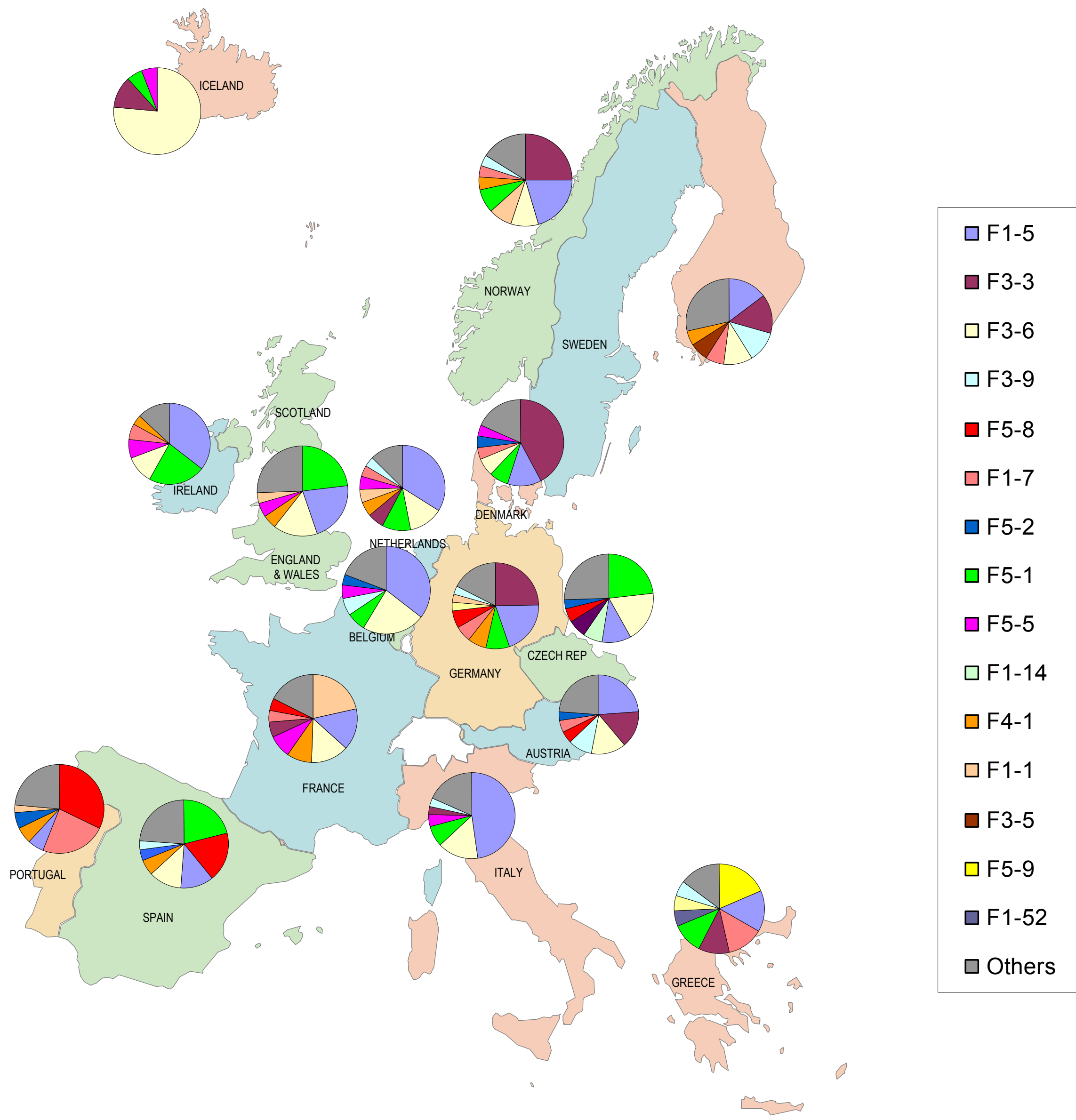


Figure 4.6: Distribution of FetA types in Europe 2000–2002.

Table 4.2: Breakdown of PorA types by frequency and association with country.

PorA	Austria	Belgium	Czech Rep.	Denmark	England & Wales	Finland	France	Germany	Greece	Iceland	Ireland	Italy	Netherlands	Norway	Portugal	Scotland	Spain	Sweden	total
P1.5,2	44	63	23	19	35	9	110*	90	14		16	9	94	25	16	19	63	4	653
P1.7-2,4	11	73*	1	7	58*	14	26	52	7		29	11	123*	13	2	15	33	12	487
P1.7,16	30*	10	4	61*	5	3	20	37	8	2	3	3	5†	28*	2	10	6	4	241
P1.5-1,10-8	7	4	2	1	21	8	25	17	2	12*	3	9	57	5		13	42	8	236
P1.19,15	2	5	15*	6	11	6	16	10	1	1	3	4	3†	5	2	9	77*	1	177
P1.18-1,3	9	8	3	13	10	12	19	10	1		14*	5	6	8	1	14	7	4	144
P1.22,14	4	6	8	6	19	4	4	8	16*	1	9	1	15	5		29*			135
P1.5-1,10-4	3	5	5	5	10	8	13	3			2	1	9	3		27*	4	4	102
P1.22,9	2			4	13	2	2	9			5	1	8	3	1	13	18	1	82
P1.19-1,15-11	1	4	2	1	22*		8	2	8*		6	1	1	5	1	13	3	2	80
P1.7-2,13-2	5	2	1	1	1	1	1	2	1			12*	39*	2			1		69
P1.21,16	4	7	1	6	9	1	1	7			3	1	15			4	6		65
P1.5-1,2-2	2		4	4	2	7	6	8	1		2		4	10		2	5	6	63
P1.7-2,16	1	1		5	2	1	2	12	1			1	16	7		2	3		54
P1.5-2,10	1	3		3	1			5	2		3	2	30*				3		53
P1.22-1,14	4	1	3		1	1	4	6	1				7	5	2		14	2	51
P1.7-1,1	2	11	2		1	1	6	4	1	1	1		14	1		1	3	2	51
P1.12-1,13-1		1	2		2	3	13*	4			1			2		2	6	2	38
P1.5-1,10-1	2	8	1			1	3	4			1		7	4		2	1	2	36
P1.5,2-1	2	9		1			15*	2				1	2	1			2		35
P1.22,14-6	3	4	3			1	3	3				1	3	2	2	2	2	1	30
P1.7-2,13-1	6	3	3		2	2		2				1	5	1		2	1	2	30
P1.5-2,10-2		4				2	2	1					3	9*		1	5	1	28
P1.18,25	1	2	2			1	2	2					4	5	1		6	1	27
P1.19,15-1				6	1	1	2	3						6			6	1	26
P1.18-1,34			5*		1	8*	1	1							1		7		24
P1.5-2,10-1						2	2	11*	1					1				3	20
P1.7-2,13	4	1				6*						1	4			1		3	20
P1.18-7,9	2	1	1	1			2	2	1				7				1		18
P1.7,16-6	1			1		9*		1					2					4	18
P1.5-10,2							16*												16
P1.19-2,13-1		1					2						1		1	1	9		15

P1.22,26	6*		4*				1	3									1		15
P1.7-4,14-6		5			1		1	7						1					15
P1.19,13-1		5			2			1					5			1			14
P1.7,16-32													11*	1					12
P1.7,16-29				3										1				7*	11
P1.7,30				1		1		1	1				2			4			10
P1.17,16-3		1	1				2					1	1						9
P1.17,9								1							1	2	3	1	8
P1.20,9				1			3							2		2			8
P1.7-2,14				1	1	1	2	1								1	1		8
P1.12-1,4		2			1			4											7
P1.17-1,23							2	1			1		3						7
P1.18-1,3-8	5*							2											7
P1.19-2,13-2																	7		7
P1.31,16	1					1							4					1	7
P1.5,2-3				1			1						5						7
P1.5-2,10-28	2					1		3										1	7
P1.7-2,10-7					1	5						1							7
P1.18-1,14	2										2						2		6
P1.18-1,30								2					2			1	1		6
P1.18-3,1				4*			2												6
P1.21-7,16					1								1	1		3			6
P1.5,2-12			6*																6
P1.5-1,10-6								4					1				1		6
P1.7,16-26			1				4									1			6
P1.12-1,13							1										3	1	5
P1.21,16-36		1						1									3		5
P1.5-1,10-10							1	3					1						5
P1.7-2,13-9								2					3						5
<5 counts	16	11	5	11	21	9	28	39	5		8	10	40	19	3	38	27	10	300
total	188	262	108	173	255	132	374	393	72	17	112	77	563	181	36	235	383	91	3652

p<0.00002 after Bonferroni correction: significant association* significant dissociation †

Table 4.3: Breakdown of FetA types by frequency and association with country.

FetA VR	Austria	Belgium	Czech Republic	Denmark	England and Wales	Finland	France	Germany	Greece	Iceland	Ireland	Italy	Netherlands	Norway	Portugal	Spain	total
F1-5	32	78*	15	17	38	19	55	67	8		38	31*	172*	37	2	42†	651
F3-6	20	51*	27	9	28	14	49	12†	3	13*	12	10	67	17		42	374
F5-1	3	15	34*	10	41*	1	6†	30	6	1	24	5	54	15		74*	319
F3-3	19	5†	10	58*	4	19	20	83*	6	2	2	2	32	45*		9†	316
F5-8	6	6	7	5	6	6	16	21	1		2	1	4†	4	11*	63*	159
F4-1	4	4	3	3	9	7	33	23	1		4	1	28	8	2	20	150
F1-1	3	3		1	7	2	76*	10			1		25	15	1	2†	146
F1-7	7	6	3	6	6	9	16	21	7		7	1	22	7	8*	7	133
F3-9	13	14	1	4	6	15*	9	10	3		1	2	19	7	1	11	116
F5-5	2	11	2	6	9	3	29*	5		1	8	3	25	2		9	115
F5-2	5	9	5	6	2	2	3	5	1		1		5	4	2	14	64
F5-9	2	1	1		1		6	5	10*				1	2		5	34
F1-30	2	1		3	2		2	2			1		11	1			25
F5-12			4	2		2	3	1				1	5			6	24
F1-18	2	1	1		1	1	9*	2					4				21
F1-14		1	10*					7	1								19
F3-7		1	1		3	2	1						3	3		4	18
F5-7	2		1			1	2	2	1				2	2		3	16
F1-15		1			1		1	5				1	4	1		1	15
F3-1	1						4	4					1	3			13
F1-20	2		4*	1	1		1					1	2				12
F3-5						9*		1					2				12
F1-50		3					4				1	1			1	1	11
F1-55															1	10*	11
F4-3	1			1				2	1				5				10
F1-23						6*						1			1		8
F5-13		1		1	2			2					2				8
F1-2					2			1			2		1				6
F4-28															1	4*	5
F5-14			1		1						1					2	5
F5-16			4*										1				5
F5-17	1															4*	5

F5-49						5*											5
<5	7	8	11	4	7	6	11	15	5		2	4	11	7	3	17	118
total	134	220	145	137	177	129	356	336	54	17	107	65	508	180	34	350	2949

p<0.0001 after Bonferroni correction: significant association* significant dissociation †

AMOVA provided evidence of geographic structuring as there were statistically significant F_{ST} values for PorA and FetA types (both $p<0.00001$). The observed F_{ST} values showed that 3.3% and 4.8% of structuring of PorA and FetA types respectively could be attributed to differences between geographical locations. These values were similar in magnitude to those obtained for MLST AMOVA analysis. Pairwise comparisons of the 18 countries using PorA types obtained F_{ST} values (Table 4.4) which were significantly different from the null hypothesis of no difference between the populations ($p>0.05$) except for three pairs of countries for which there was evidence of gene flow: England and Wales and the Republic of Ireland, Finland and Sweden and France and Portugal. For FetA, data were available for 16 countries. Pairwise comparisons showed that there was significant structuring of FetA types in Europe except for between Austria and Finland, Belgium and Italy and Germany and Norway (Table 4.5).

Table 4.4: F_{ST} values of pairwise comparisons of PorA types in 18 European countries.

	Austria	Belgium	Czech Rep.	Denmark	England & Wales	Finland	France	Germany	Greece	Iceland	Ireland	Italy	Netherlands	Norway	Portugal	Scotland	Spain
Belgium	0.03505																
Czech Rep.	0.01930	0.05069															
Denmark	0.03055	0.09370	0.06858														
E & W	0.03751	0.01567	0.03947	0.08252													
Finland	0.03119	0.04231	0.02522	0.06538	0.02122												
France	0.01125	0.02885	0.01608	0.06973	0.03381	0.03500											
Germany	0.00553	0.01469	0.02024	0.04869	0.01856	0.02425	0.00817										
Greece	0.02652	0.04567	0.02912	0.06025	0.02634	0.04269	0.03534	0.02417									
Iceland	0.20985	0.25590	0.21995	0.24135	0.19754	0.18606	0.20973	0.20555	0.23244								
Ireland	0.03940	0.01197	0.04735	0.08228	**0.00232***	0.02213	0.03867	0.02096	0.03186	0.23901							
Italy	0.02695	0.03377	0.03458	0.07194	0.01816	0.01633	0.03157	0.02045	0.04250	0.17590	0.02341						
Netherlands	0.03286	0.01273	0.04415	0.08759	0.01068	0.02624	0.03031	0.01513	0.04029	0.18886	0.01508	0.00802					
Norway	0.00744	0.03880	0.02149	0.02323	0.02973	0.01840	0.02232	0.00934	0.02405	0.19816	0.03202	0.02389	0.03196				
Portugal	0.02424	0.04678	0.02890	0.10347	0.06739	0.07488	**0.00996***	0.02328	0.05768	0.31823	0.07203	0.06987	0.06005	0.04860			
Scotland	0.03202	0.05143	0.02305	0.06004	0.01915	0.01316	0.03661	0.02880	0.02080	0.18418	0.02655	0.02649	0.03563	0.02070	0.07462		
Spain	0.03758	0.04829	0.01577	0.08320	0.03120	0.02764	0.02760	0.02614	0.05501	0.17535	0.04367	0.02589	0.03480	0.03195	0.05056	0.03500	
Sweden	0.03275	0.04049	0.03745	0.06381	0.01968	**0.00287***	0.03997	0.02376	0.04541	0.17398	0.02463	0.01710	0.02214	0.01605	0.08464	0.01833	0.03304

bold* $p>0.05$ can accept null hypothesis of no difference between populations, *i.e.* presence of gene flow between two countries. Note: E & W is England and Wales.

Table 4.5: F_{ST} **values of pairwise comparisons of FetA types in 16 European countries.**

	Austria	Belgium	Czech Rep.	Denmark	England & Wales	Finland	France	Germany	Greece	Iceland	Ireland	Italy	Netherlands	Norway	Portugal
Belgium	0.01875														
Czech Rep.	0.04089	0.05531													
Denmark	0.05507	0.12770	0.08926												
E & W	0.03133	0.02789	0.01028	0.10523											
Finland	**0.00534***	0.04739	0.04493	0.05066	0.04330										
France	0.03165	0.05356	0.05779	0.09818	0.04282	0.03397									
Germany	0.01682	0.06420	0.04760	0.02104	0.04558	0.01858	0.04847								
Greece	0.02568	0.06540	0.03809	0.07277	0.03803	0.02625	0.05330	0.02629							
Iceland	0.19408	0.19451	0.16719	0.26969	0.19722	0.20075	0.19833	0.24451	0.25015						
Ireland	0.04073	0.02119	0.04154	0.12778	0.00817	0.06345	0.06595	0.05457	0.05424	0.26525					
Italy	0.03786	**0.00670***	0.08671	0.15009	0.04662	0.07145	0.07965	0.07287	0.08635	0.29176	0.01640				
Netherlands	0.01471	0.00825	0.04675	0.09729	0.01667	0.03665	0.03914	0.03665	0.04504	0.22443	0.00731	0.00857			
Norway	0.01041	0.05320	0.04453	0.02156	0.04045	0.01682	0.03588	**0.00186***	0.03070	0.22085	0.05392	0.06714	0.03002		
Portugal	0.08738	0.13848	0.10381	0.15479	0.10818	0.06956	0.08652	0.08839	0.07703	0.33779	0.13607	0.17365	0.12163	0.10844	
Spain	0.04470	0.06057	0.01468	0.10272	0.01674	0.04604	0.05586	0.04956	0.04310	0.20621	0.04404	0.08716	0.05082	0.05445	0.05808

bold* $p>0.05$ can accept null hypothesis of no difference between populations, *i.e.* presence of gene flow between two countries. Note: E & W is England and Wales.

4.2.2.2 Temporal structuring

For PorA there were some significant changes in prevalence of some types between 2000 and 2002 ($p<0.0001$ after Bonferroni correction) (Table 4.6). P1.5-1,10-8 increased from 4.3% to 7.8% (51 to 100 isolates) ($p<0.0001$). This change is most likely to have been due to increases in this type in France and the Netherlands which together accounted for 39% of this type. It increased from 4% to 14% (7 to 29 isolates) in the Netherlands and 4% to 11% (4 to 14 isolates) in France over the three years. The odds of an isolate having this type in 2002 were 1.88 times that in 2000. The P1.5-10,2 type decreased from 1.4% to 0% although numbers were small (16 isolates to 0). There were some changes between 2000 and 2002 at the $p<0.05$ level: P1.5,2 and P1.7-2,16 decreased, and P1.22,14, P1.7-1,1 and P1.5,2-1 increased.

For FetA there was a significant increase in F3-6 between 2000 and 2002 going from 9.3% to 14.2% (93 to 144 isolates) (Table 4.7). This was most likely to have been driven by the increases in this type in the Netherlands and France where it went from 2.4% to 9% (9 to 34 isolates) and 2.4% to 5.3% (9 to 20 isolates) in each country respectively. The odds of an isolate having this type in 2002 compared to 2000 were 1.6 times. There was a significant decrease in F5-8 from 7.2% to 3.7% (72 to 37 isolates) over the three years. This is most likely due to the decrease in this type in Spain where the majority of these types originated from (40%). The odds of an isolate having this type in 2002 *versus* 2000 were about half. It decreased here from 32.6% (45 isolates) to 7.7% (10 isolates). There were some changes at the $p<0.05$ level interestingly among members of the F5 family: F5-5, F5-9 and F5-7 increased while F5-2 decreased.

Table 4.6: Change in PorA type over time (2000–2002).

PorA	year (frequency and percentage)				χ^2 p-value		
	2000	2001	2002	total	overall change	2000/2002 change	2002/2000 OR and 95% CIs
P1.5,2	219 (18.6%)	235 (19.6%)	199 (15.6%)	653 (17.9%)	0.022	0.049	0.809 (0.655–0.999)
P1.7-2,4	167 (14.2%)	153 (12.8%)	167 (13.1%)	487 (13.3%)	0.594	0.433	0.912 (0.724–1.149)
P1.7,16	77 (6.5%)	84 (7.0%)	80 (6.3%)	241 (6.6%)	0.734	0.788	0.957 (0.694–1.324)
P1.5-1,10-8	51 (4.3%)	85 (7.1%)	100 (7.8%)	236 (6.5%)	0.001	**0.0003***	**1.876 (1.332–2.673)**
P1.19,15	63 (5.3%)	45 (3.8%)	69 (5.4%)	177 (4.8%)	0.107	0.947	1.012 (0.712–1.441)
P1.18-1,3	42 (3.6%)	58 (4.8%)	44 (3.4%)	144 (3.9%)	0.140	0.875	0.966 (0.627–1.491)
P1.22,14	37 (3.1%)	35 (2.9%)	63 (4.9%)	135 (3.7%)	0.015	0.025	1.599 (1.062–2.441)
P1.5-1,10-4	40 (3.4%)	28 (2.3%)	34 (2.7%)	102 (2.8%)	0.285	0.290	0.780 (0.487–1.241)
P1.22,9	19 (1.6%)	28 (2.3%)	35 (2.7%)	82 (2.2%)	0.162	0.057	1.714 (0.984–3.08)
P1.19-1,15-11	27 (2.3%)	20 (1.7%)	33 (2.6%)	80 (2.2%)	0.294	0.637	1.131 (0.675–1.909)
P1.7-2,13-2	20 (1.7%)	23 (1.9%)	26 (2.0%)	69 (1.9%)	0.821	0.535	1.202 (0.668–2.198)
P1.21,16	24 (2.0%)	23 (1.9%)	18 (1.4%)	65 (1.8%)	0.452	0.232	0.690 (0.366–1.277)
P1.5-1,2-2	24 (2.0%)	18 (1.5%)	21 (1.6%)	63 (1.7%)	0.592	0.470	0.806 (0.441–1.460)
P1.7-2,16	24 (2.0%)	18 (1.5%)	12 (0.9%)	54 (1.5%)	0.080	0.024	0.460 (0.220–0.910)
P1.5-2,10	23 (1.9%)	15 (1.3%)	15 (1.2%)	53 (1.5%)	0.218	0.120	0.600 (0.304–1.149)
P1.22-1,14	16 (1.4%)	13 (1.1%)	22 (1.7%)	51 (1.4%)	0.404	0.463	1.271 (0.665–2.483)
P1.7-1,1	8 (0.7%)	19 (1.6%)	24 (1.9%)	51 (1.4%)	0.032	0.009	2.766 (1.284–6.660)
P1.12-1,13-1	9 (0.8%)	17 (1.4%)	12 (0.9%)	38 (1.0%)	0.257	0.635	1.228 (0.513–3.053)
P1.5-1,10-1	10 (0.8%)	17 (1.4%)	9 (0.7%)	36 (1.0%)	0.164	0.685	0.831 (0.325–2.094)
P1.5,2-1	5 (0.4%)	14 (1.2%)	16 (1.3%)	35 (1.0%)	0.071	0.026	2.915 (1.127–9.114)
P1.22,14-6	8 (0.7%)	12 (1.0%)	10 (0.8%)	30 (0.8%)	0.665	0.761	1.151 (0.448–3.068)
P1.7-2,13-1	10 (0.8%)	10 (0.8%)	10 (0.8%)	30 (0.8%)	0.982	0.858	0.923 (0.373–2.282)
P1.5-2,10-2	15 (1.3%)	5 (0.4%)	8 (0.6%)	28 (0.8%)	0.046	0.0970	0.494 (0.196–1.151)
P1.18,25	12 (1.0%)	7 (0.6%)	8 (0.6%)	27 (0.7%)	0.398	0.281	0.617 (0.238–1.513)
P1.19,15-1	8 (0.7%)	8 (0.7%)	10 (0.8%)	26 (0.7%)	0.933	0.761	1.151 (0.447–3.067)
P1.18-1,34	8 (0.7%)	8 (0.7%)	8 (0.6%)	24 (0.7%)	0.985	0.873	0.923 (0.333–2.556)
P1.5-2,10-1	6 (0.5%)	7 (0.6%)	7 (0.5%)	20 (0.5%)	0.968	0.893	1.074 (0.348–3.423)
P1.7-2,13	6 (0.5%)	8 (0.7%)	6 (0.5%)	20 (0.5%)	0.777	0.890	0.923 (0.281–3.033)
P1.18-7,9	8 (0.7%)	6 (0.5%)	4 (0.3%)	18 (0.5%)	0.434	0.195	0.469 (0.121–1.526)
P1.7,16-6	3 (0.3%)	8 (0.7%)	7 (0.5%)	18 (0.5%)	0.331	0.347	2.099 (0.566–10.304)
P1.5-10,2	16 (1.4%)	0	0	16 (0.4%)	**4.89E-08***	**2.96E-05***	n/a
P1.19-2,13-1	7 (0.6%)	5 (0.4%)	3 (0.2%)	15 (0.4%)	0.374	0.210	0.406 (0.083–1.505)
P1.22,26	5 (0.4%)	7 (0.6%)	3 (0.2%)	15 (0.4%)	0.396	0.492	0.564 (0.109–2.310)
P1.7-4,14-6	6 (0.5%)	2 (0.2%)	7 (0.5%)	15 (0.4%)	0.284	0.893	1.074 (0.348–3.423)
< 15	157	153	188	498			
total	1180	1194	1278	3652			

*** significant** after the Bonferroni correction p=0.001

Table 4.7: Change in FetA type over time (2000–2002).

FetA	year (frequency and percentage)				χ^2 p-value		
	2000	2001	2002	total	overall change	2000/2002 change	2002/2000 OR and 95% CIs
F1-5	231 (23.1%)	197 (21%)	223 (22.1%)	651 (22.1%)	0.524	0.568	0.941 (0.763–1.160)
F3-6	93 (9.3%)	137 (14.6%)	144 (14.2%)	374 (12.7%)	**0.0004***	**0.0006***	**1.617 (1.228–2.138)**
F5-1	122 (12.2%)	89 (9.5%)	108 (10.7%)	319 (10.8%)	0.151	0.281	0.860 (0.652–1.132)
F3-3	111 (11.1%)	88 (9.4%)	117 (11.6%)	316 (10.7%)	0.258	0.744	1.047 (0.794–1.381)
F5-8	72 (7.2%)	50 (5.3%)	37 (3.7%)	159 (5.4%)	**0.002***	**0.0004***	**0.490 (0.323–0.732)**
F4-1	42 (4.2%)	63 (6.7%)	45 (4.5%)	150 (5.1%)	0.023	0.786	1.061 (0.689–1.637)
F1-1	50 (5%)	58 (6.2%)	38 (3.8%)	146 (5%)	0.048	0.172	0.742 (0.479–1.141)
F1-7	43 (4.3%)	50 (5.3%)	40 (4%)	133 (4.5%)	0.322	0.695	0.916 (0.588–1.425)
F3-9	40 (4%)	43 (4.6%)	33 (3.3%)	116 (3.9%)	0.325	0.375	0.810 (0.503–1.295)
F5-5	27 (2.7%)	41 (4.4%)	47 (4.6%)	115 (3.9%)	0.053	0.021	**1.750 (1.088–2.872)**
F5-2	33 (3.3%)	12 (1.3%)	19 (1.9%)	64 (2.2%)	0.007	0.044	**0.563 (0.311–0.989)**
F5-9	7 (0.7%)	8 (0.9%)	19 (1.9%)	34 (1.2%)	0.027	0.019	**2.675 (1.161–6.958)**
F1-30	8 (0.8%)	10 (1.1%)	7 (0.7%)	25 (0.8%)	0.656	0.778	0.866 (0.297–2.464)
F5-12	6 (0.6%)	5 (0.5%)	13 (1.3%)	24 (0.8%)	0.118	0.112	2.125 (0.827–6.174)
F1-18	3 (0.3%)	5 (0.5%)	13 (1.3%)	21 (0.7%)	0.023	0.013	**4.156 (1.317–18.993)**
F1-14	9 (0.9%)	2 (0.2%)	8 (0.8%)	19 (0.6%)	0.129	0.789	0.879 (0.324–2.342)
F3-7	9 (0.9%)	1 (0.1%)	8 (0.8%)	18 (0.6%)	0.053	0.789	0.879 (0.324–2.342)
F5-7	3 (0.3%)	2 (0.2%)	11 (1.1%)	16 (0.5%)	0.014	0.034	**3.518 (1.077–16.340)**
F1-15	5 (0.5%)	5 (0.5%)	5 (0.5%)	15 (0.5%)	0.992	1.000	0.988 (0.265–3.683)
F3-1	5 (0.5%)	5 (0.5%)	3 (0.3%)	13 (0.4%)	0.691	0.505	0.604 (0.116–2.569)
F1-20	7 (0.7%)	1 (0.1%)	4 (0.4%)	12 (0.4%)	0.115	0.384	0.572 (0.144–1.950)
F3-5	2 (0.2%)	7 (0.7%)	3 (0.3%)	12 (0.4%)	0.177	1.000	1.449 (0.220–12.497)
F1-50	5 (0.5%)	5 (0.5%)	1 (0.1%)	11 (0.4%)	0.193	0.122	0.219 (0.008–1.426)
F1-55	0	3 (0.3%)	8 (0.8%)	11 (0.4%)	0.007	0.008	n/a
F4-3	1 (0.1%)	2 (0.2%)	7 (0.7%)	10 (0.3%)	0.079	0.070	6.205 (1.072–158.139)
F1-23	2 (0.2%)	4 (0.4%)	2 (0.2%)	8 (0.3%)	0.611	1.000	0.988 (0.103–9.506)
F5-13	4 (0.4%)	3 (0.3%)	1 (0.1%)	8 (0.3%)	0.415	0.216	0.272 (0.010–1.966)
F1-2	3 (0.3%)	2 (0.2%)	1 (0.1%)	6 (0.2%)	0.618	0.372	0.358 (0.012–3.091)
<=5	56	41	46	143	n/a	n/a	n/a
total	999	939	1011	2949			

*** significant** after the Bonferroni correction p=0.002

AMOVA provided some evidence of temporal structuring in Europe for the years of the study 2000, 2001 and 2002 in terms of antigens. There were statistically significant, although quite low, F_{ST} values for PorA and FetA types (both $p<0.001$) among the three years. Again, the observed F_{ST} values were similar in magnitude to those found for the MLST AMOVA tests and showed that almost 0.1% and 0.2% of structuring of PorA and FetA types respectively could be attributed to differences among the three years.

F_{ST} values of pairwise comparisons for each of the three years using PorA and FetA allele types showed some evidence of structuring. There was a significant difference between the populations of 2000 and 2001, 2000 and 2002 and 2001 and 2002 according to the PorA type (Table 4.8). Looking at the individual VRs, there was a significant difference between the years 2000 and 2002 for VR1 and VR2 and also between 2001 and 2002 for VR2. For pairwise comparisons for FetA types, there were highly significant differences between the years 2000 and 2001 and 2000 and 2002 (Table 4.9). So for all the antigens (PorA, FetA, VR1 and VR2), there was evidence of differences between the first and last years of the study, *i.e.* 2000 and 2002, indicating some change over the period in antigen distribution.

Table 4.8: Temporal structuring in Europe. F_{ST} values of pairwise comparisons of each year using PorA, VR1 and VR2 types in 18 European countries.

type		F_{ST} values	
		2000	**2001**
PorA	**2001**	0.00058	
	2002	0.00123	0.00095
VR1	**2001**	0.00023*	
	2002	0.00089	0.00047*
VR2	**2001**	0.00050*	
	2002	0.00199	0.00147

* $p>0.05$ cannot accept null hypothesis of no difference between populations, *i.e.* presence of gene flow.

Table 4.9: Temporal structuring in Europe. F_{ST} values of pairwise comparisons of each year using FetA in 16 European countries.

year	F_{ST} values	
	2000	**2001**
2001	0.00262	
2002	0.00205	0.00071*

* p>0.05 cannot accept null hypothesis of no difference between populations, *i.e.* presence of gene flow.

F_{ST} values of pairwise comparisons of the three years using the antigen allele data, indicated structuring in some countries (Tables 4.10 and 4.11). For PorA there were significant differences between the years 2000 and 2001 for Austria, France, Ireland, the Netherlands and Scotland. There were significant differences between 2001 and 2002 for Belgium and France. There were significant differences between 2000 and 2002 for Finland, France, the Netherlands, Scotland and Spain. For FetA there were significant differences between 2000 and 2001 for Belgium, Ireland, the Netherlands and Spain; between 2001 and 2002 for Belgium; for 2000 and 2002 for France, the Netherlands, Norway and Spain.

Simpson's index of diversity (with 95% CIs), was employed to estimate the diversity of each country in terms of antigen and the change in this diversity over the three years (Figures 4.7 and 4.8). For PorA the only significant difference in diversity was between the years 2000 and 2001 in Scotland where 2001 was more diverse. For FetA, significant differences were found between the years 2000 and 2002 for Italy and Spain where there were downward and upward trends in diversity index respectively. In the Netherlands there was a significant difference between the years 2000 and 2001 where there was an increase in diversity. Countries which implemented routine vaccination during this time period were the UK (1999), the Republic of Ireland (autumn 2000), the Netherlands (mid 2002), Spain (2000) and Belgium (2002). These interventions are likely to have had an impact on the diversity of antigen types found.

Table 4.10: Temporal structuring in Europe. F_{ST} values of pairwise comparisons of each year using PorA types in 16 European countries.

country		F_{ST} values	
		2000	**2001**
Austria	**2001**	0.01471	
	2002	0.01035*	-0.00122*
Belgium		-0.00309*	
		0.01032*	0.01549
Czech Rep.		0.04784*	
		0.00742*	0.05456*
Denmark		0.01153*	
		0.00035*	0.00193*
England &Wales		-0.00114*	
		0.00362*	0.00322*
Finland		0.00264*	
		0.01987	0.00619*
France		0.02176	
		0.01869	0.00881
Germany		-0.00146*	
		0.00370*	-0.00022*
Greece		-0.02937*	
		0.00612*	-0.00984*
Ireland		0.03269	
		0.02083*	-0.00318*
Italy		0.01326*	
		-0.00236*	-0.00627*
Netherlands		0.00833	
		0.00938	-0.00244*
Norway		0.00185*	
		-0.00538*	-0.00355*
Scotland		0.03279	
		0.03712	0.00369*
Spain		0.00516*	
		0.02234	0.00444*
Sweden		n/a	
		n/a	0.00656*

* p>0.05 cannot accept null hypothesis of no difference between populations, *i.e.* presence of gene flow
n/a = data not available for Sweden year 2000. Note: Data not available for 2000 and 2002 for Iceland and Portugal

Table 4.11: Temporal structuring in Europe. F_{ST} values of pairwise comparisons of each year using FetA types in 14 European countries.

country		F_{ST} values	
		2000	**2001**
Austria	**2001**	0.00608*	
	2002	-0.00582*	-0.00117*
Belgium		0.02798	
		-0.00343*	0.02887
Czech Rep.		0.01354*	
		0.00576*	0.01679*
Denmark		-0.01095*	
		0.00128*	-0.01070*
England &Wales		-0.00505*	
		-0.00062*	0.00481*
Finland		0.00359*	
		0.00384*	0.00004*
France		0.00775*	
		0.01141	0.00125*
Germany		0.00074*	
		0.00223*	-0.00468*
Greece		-0.04839*	
		-0.01279*	0.00090*
Ireland		0.03397	
		0.02986*	-0.02841*
Italy		0.01491*	
		-0.00028*	-0.00021*
Netherlands		0.01335	
		0.01729	0.00317*
Norway		0.00306*	
		0.01666	0.00579*
Spain		0.02704	
		0.03658	0.00351*

* $p>0.05$ cannot accept null hypothesis of no difference between populations, *i.e.* presence of gene flow.
Note: data not available for Scotland and Sweden and for Iceland and Portugal years 2000 and 2002.

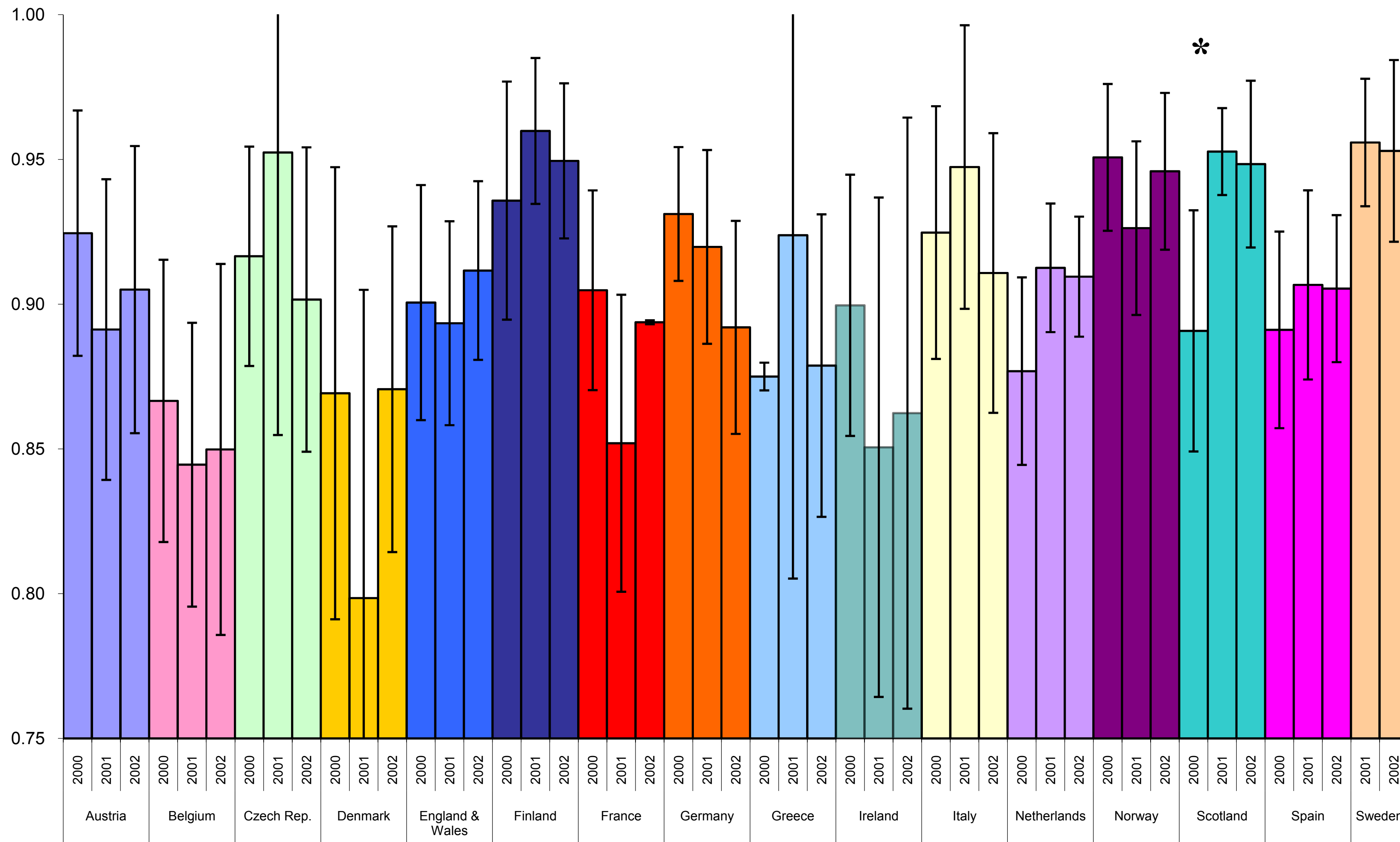


Figure 4.7: Change in diversity index (*D*) in terms of PorA type over three years in each European country. *A significant change in *D* is indicated by non-overlapping CIs. Note: Since there were isolates only for a single year for Iceland and Portugal these data were omitted.

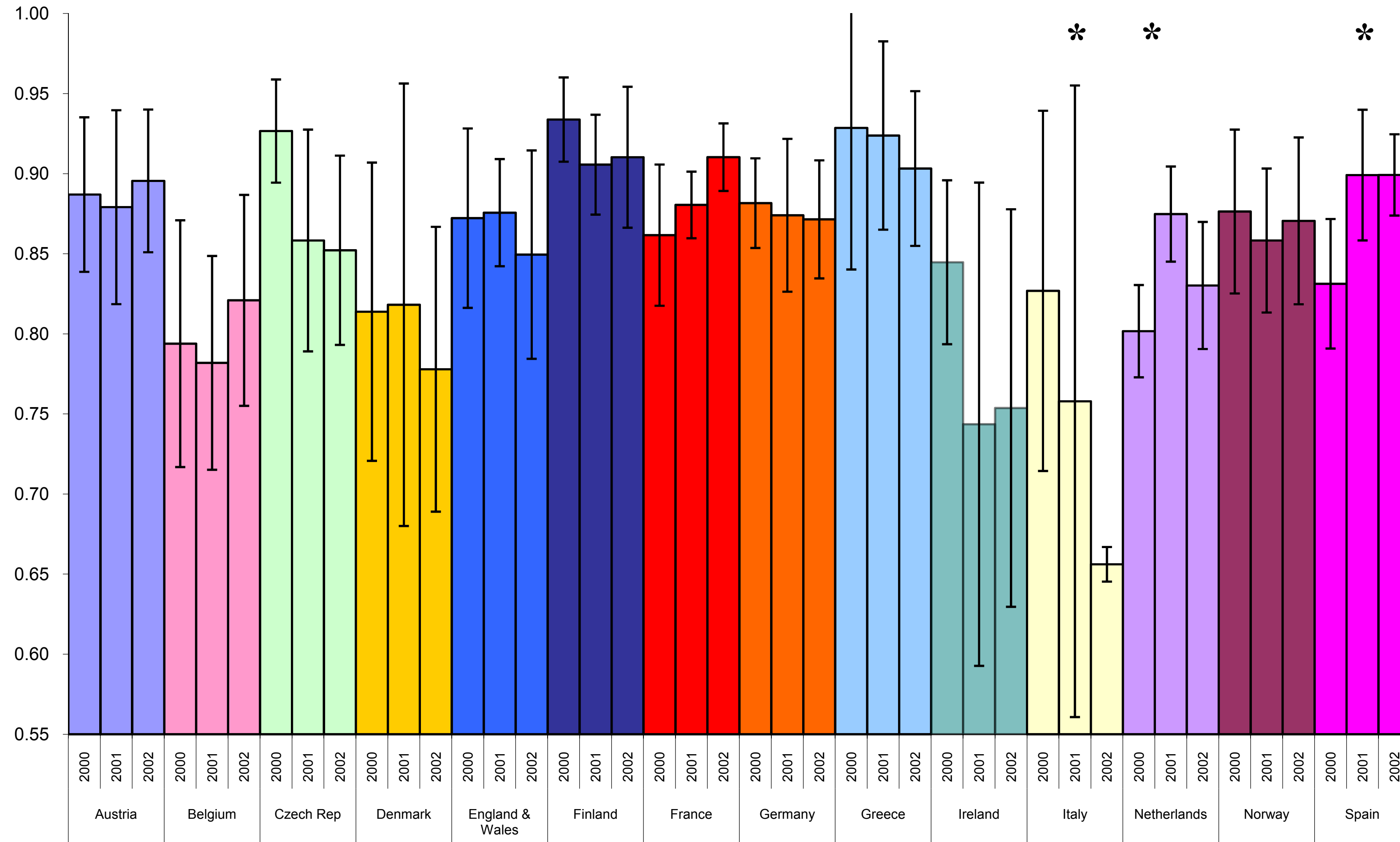
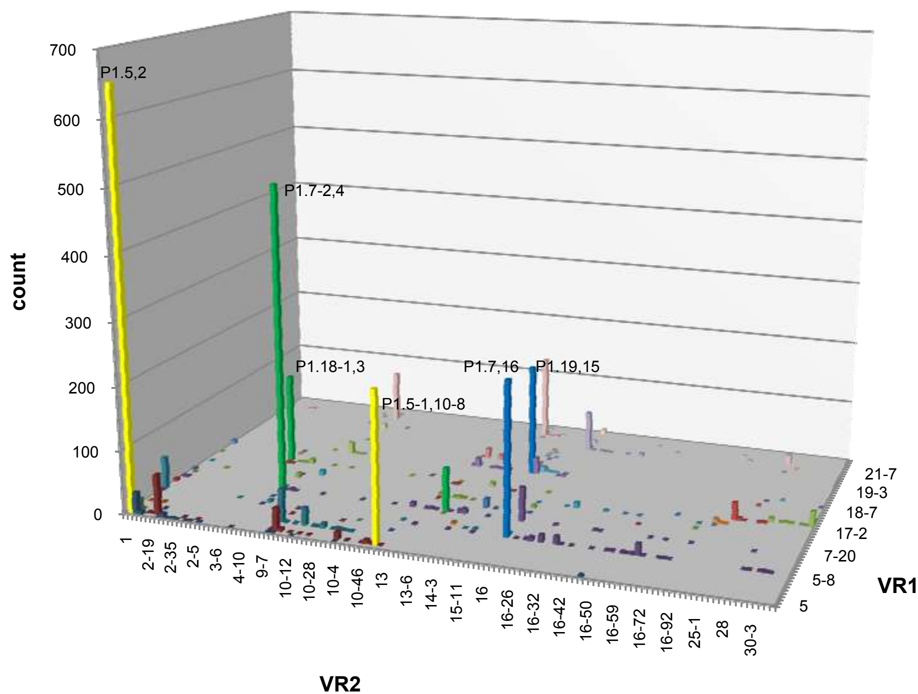


Figure 4.8: Change in diversity index (*D*) in terms of FetA type over three years in each European country. *A significant change in *D* is indicated by non-overlapping CIs. Note: Since there were isolates only for a single year for Iceland and Portugal these data were omitted.

4.2.2.3 Non-overlapping antigenic strain structure

As already demonstrated, PorA and FetA are diverse antigens but there are a relatively small number of types that account for most disease isolates. These antigens are structured in a non-overlapping manner in their association with each other. For example, PorA VR combinations occurred predominantly in non-overlapping combinations with a distribution that was significantly different from the expected distribution given the data (χ^2 test p<0.005) (Figure 4.9). The top seven PorA VR combinations did not share any alleles amongst them: P1.5,2; P1.7-2,4; P1.7,16; P1.5-1,10-8; P1.19,15; P1.18-1,3; P1.22,14. This is a product of host immune selection meaning that VR epitopes are not shared among the main disease-associated PorA types as this would be disadvantageous for the organism. The majority of VR combinations were found only once or a few times with a relatively small number of combinations accounting for most disease isolates.

(a) Observed



(b) Expected

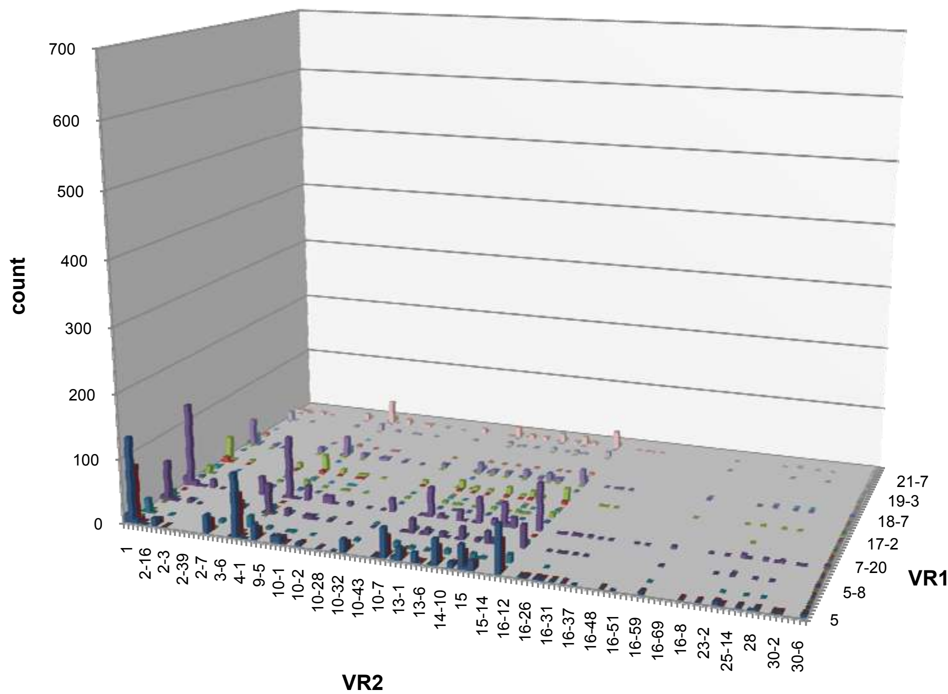


Figure 4.9: PorA VR combinations. Observed distribution (a) with the six most common combinations indicated *versus* (b) expected distribution under the null hypothesis of random reassortment χ^2 test p<0.005.

Combined PorA/FetA type information was available for 2813 isolates. There were 604 PorA/FetA combinations, 363 of which were unique (13%) (Figure 4.10; Table 4.12). The P1.7-2,4: F1-5 combination was markedly the most prevalent of the combinations accounting for 11.3% of all isolates, while the next most prevalent was P1.5,2 F3-6 accounting for 5.3% of isolates. There was a less marked overlapping structure for the FetA/PorA combinations (Figure 4.10) than for the PorA VRs (Figure 4.9). PorA type P1.5,2 in particular was associated frequently with several different FetA types, *e.g.* P1.5,2 F3-6; P1.5,2 F1-1; P1.5,2 F5-8; P1.5,2 F1-7 (Table 4.13).

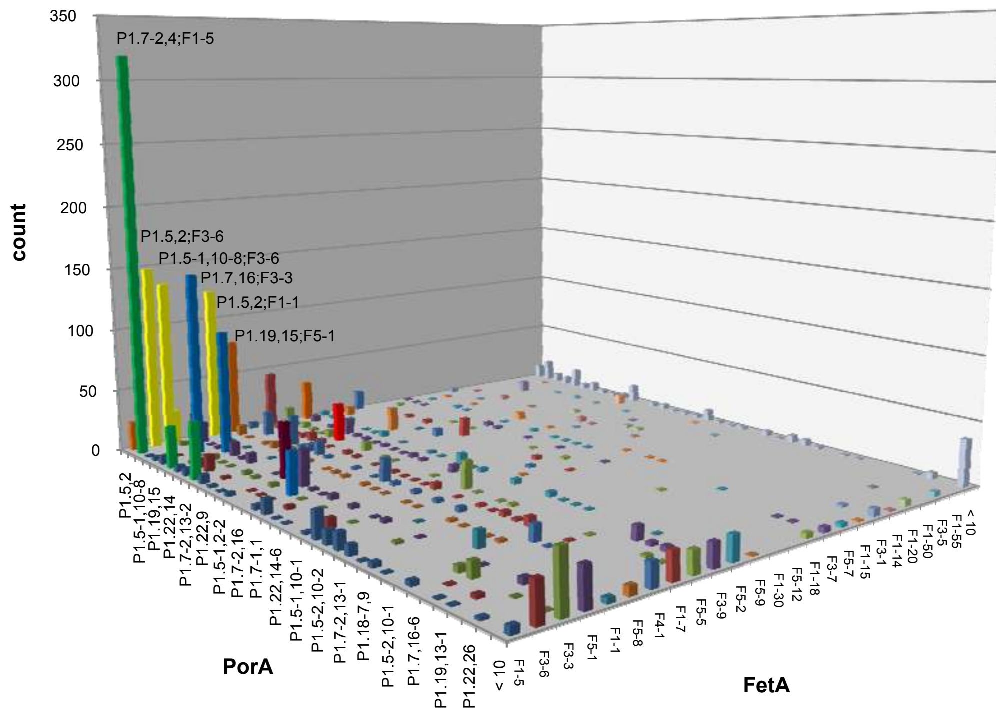


Figure 4.10: PorA/FetA combinations, with colour coding according to clonal complex for the top six combinations, *green* ST-41/44 complex, *blue* ST-32 complex, *yellow* ST-11 complex, *orange* ST-8 complex.

Table 4.12: PorA/FetA antigen combinations

PorA/FetA	frequency	percentage
P1.7-2,4 F1-5	319	11.3
P1.5,2 F3-6	148	5.3
P1.7,16 F3-3	146	5.2
P1.5-1,10-8 F3-6	139	4.9
P1.5,2 F1-1	124	4.4
P1.19,15 F5-1	99	3.5
P1.5,2 F5-8	78	2.8
P1.5,2 F1-7	45	1.6
P1.7-2,13-2 F1-5	45	1.6
P1.19-1,15-11 F5-1	44	1.6
P1.7-2,16 F3-3	34	1.2
P1.18-1,3 F1-5	33	1.2
P1.5,2 F3-9	32	1.1
P1.22,14 F5-5	31	1.1
P1.5-2,10 F5-1	29	1
P1.5,2 F3-3	27	1
P1.18-1,3 F4-1	24	0.9
P1.5,2 F1-5	23	0.8
P1.5,2-1 F5-5	22	0.8
P1.18,25 F1-5	21	0.7
P1.5,2 F5-1	21	0.7
P1.5-1,2-2 F5-8	20	0.7
P1.22,14 F5-9	19	0.7
P1.5-1,10-8 F4-1	19	0.7
P1.22-1,14 F4-1	18	0.6
P1.5,2 F1-30	16	0.6
P1.7-2,4 F5-1	16	0.6
P1.22,9 F5-12	15	0.5
P1.19,15-1 F1-5	13	0.5
P1.5-1,10-4 F3-6	13	0.5
P1.5-10,2 F1-1	13	0.5
P1.5-2,10-1 F4-1	13	0.5
P1.18-1,3 F3-9	12	0.4
P1.5,2 F5-5	12	0.4
P1.7,16-6 F3-3	12	0.4
P1.7-1,1 F1-5	12	0.4
P1.18-1,3 F5-1	11	0.4
P1.5-2,10-2 F1-5	11	0.4
P1.7-2,13-2 F4-1	11	0.4
P1.7-4,14-6 F3-9	11	0.4
P1.12-1,13-1 F1-5	10	0.4
P1.18-1,34 F1-5	10	0.4
P1.22,14-6 F1-5	10	0.4
P1.22,9 F5-1	10	0.4
P1.5-1,10-4 F1-18	10	0.4
P1.7,16-32 F3-3	10	0.4
<10 counts	1002	35.3
	2813	

4.2.3 Antigen/clonal complex association

Associations between antigens and clonal complexes have previously been described (Urwin *et al.* , 2004) and were evident here also. For example, PorA type P1.5,2 was significantly associated with ST-11, ST-8 and ST-60 complexes (Table 4.13). These three complexes accounted for 94% of the isolates of this type. P1.7-2,4 was associated with ST-41/44 complex; 95% of isolates of this type were of this complex. P1.5-1, 10-8 was associated with the ST-11 complex; P1.7,16 and P1.19,15 were associated with ST-32 complex; P1.18-1,3 with ST-22 complex and P1.22,14 was associated with the ST-213 and ST-162 complexes. The serogroup Y-associated ST-23 complex was linked with PorA types, P1.5-1,2-2; P1.5-2,10-2 and P1.5-2,10-1. These antigen associations have been noted in studies of ST-23/serogroup Y disease in the United States (Harrison *et al.*, 2006). For FetA, type F1-5 was associated with ST-41/44 complex (Table 4.14); 95% of isolates of this type were from this complex. F3-6 and F1-1 were associated with ST-11 complex; F3-3 and F5-1 were associated with ST-32 complex; F5-1 was also associated with the ST-269 complex. There were significant under-representations of antigens in clonal complexes also. For example, P1.5,2 was not associated with the ST-41/44, ST-32 and ST-269 complexes. The F1-5 type was under-represented in the ST-11, ST-32, ST-8 and ST-269 complexes. Again the associations of the serogroup Y-linked ST-23 complex and the FetA antigens F5-8 and F4-1 correspond to previous work (Harrison *et al.*, 2006).

Table 4.13: Breakdown of PorA types by frequency and association with clonal complex.

PorA	ST-41/44	ST-11	ST-32	ST-269	ST-8	ST-213	ST-22	ST-23	ST-60	ST-461	ST-35	ST-18	ST-174	ST-162	ST-167	ST-334	ST-364	ST-865	ST-254	ST-750	ST-231	ST-103	ST-1157	ST-53	ST-5	ST-226	Unass.	Others	total
P1.5,2	8	399	17	3	177		1		30				1														11		647
P1.7-2,4	457		7									1		3													12		480
P1.7,16	1	1	226	1	3					1										1		1					3		238
P1.5-1,10-8	4	220							4					1	3			1	1								2		236
P1.19,15	11		125	9	1						3								2								3	1	155
P1.18-1,3	46	1	8	3	3		47			2		4						1				3				3	13		134
P1.22,14	7	4	1	3	2	65						13		23					1								13		132
P1.5-1,10-4	4	36	2	11			1	1					12		11	1			3		4							2	88
P1.19-1,15-11		1		76			1				1																		79
P1.22,9	2	3		57	1					1			2				1			2									69
P1.7-2,13-2	66			1				1																					68
P1.21,16	14	11		2	2			1	15				5							7							2	1	60
P1.5-1,2-2	6	2	4	7	7			20	1	1			1						1								2	1	53
P1.7-2,16		1	43	5		1	1		1																		1		53
P1.5-2,10		1	38	1	8							1															1		50
P1.7-1,1	19	2	5	1	9							1						1				1					3		42
P1.22-1,14	7		1								29		1																38
P1.12-1,13-1	15		5	2					1			2					8										2		35
P1.5,2-1		33			1																						1		35
P1.7-2,13-1	19	1	3							1	1							2									2	1	30
P1.22,14-6	18	2	3		2		2					2																	29
P1.5-1,10-1	3	11	1	2				2				1			1						1	1				1	3		27
P1.5-2,10-2	15	1						10																			1		27
P1.19,15-1	18	1			1																						4		24
P1.18,25	16		1	3																									20
P1.5-2,10-1	1		1					17					1																20
<20	142	61	120	38	16	4	11	9	3	33	3	7	7	1	4	15	3	6	3		4	2	7	7	5		29	5	545
total	899	792	611	225	233	70	64	61	55	39	37	32	30	28	19	16	12	11	11	10	9	8	7	7	5	4	108	11	3414

p<0.00002 after Bonferroni correction: significant association significant dissociation

Table 4.14: Breakdown of FetA types by frequency and association with clonal complex.

FetA VR	ST-41/44	ST-11	ST-32	ST-8	ST-269	ST-23	ST-22	ST-213	ST-60	ST-35	ST-174	ST-461	ST-18	ST-162	ST-334	ST-167	ST-231	ST-865	ST-364	ST-103	ST-750	ST-254	ST-226	ST-5	Unass	Other	total
F1-5	564	3	4	27	1		2	1				1	2		1			2							5	2	615
F3-6	3	341		7	1		1						4			2		1				1			11		372
F3-3	2	31	272						2			1					1						1		6		316
F5-1	22	2	172	9	93		1			1						2									3		305
F5-8	9	6		92	2	18	1	1	2	2	1	1				1	7	2							2		147
F1-1		136	1	1	1											2									4		145
F4-1	16	25		2	2	27	25	2	3	20						2				1					18		143
F1-7	38	2	3	25	6		1	5	26			1	1	1			3		1	1					4		118
F3-9	11	3	1	37	8		5		2			18	5		12	2				5					6		115
F5-5	8	43	1	5	1		5	30	1			3	1								4				6		108
F5-2	11		5		2		1		1	4			1		1	1		1	1			3			9		41
F5-9		5		1										19	1										8		34
F1-30		24							1																		25
F1-18	6										14																20
F5-12					13								4						3								20
F3-7	4	1		2							7	1															15
F1-50			10									1															11
F5-7	5		3								1												1			1	11
F1-55	1	1			8																						10
< 10	71	18	21	1	11	4	2	1	1	2	6	1	9					4	4	1	2	1	2	4	13	7	186
total	771	641	493	209	149	49	44	40	39	29	29	28	27	20	15	12	11	10	9	8	6	5	4	4	95	10	2757

p<0.0009 after Bonferroni correction: significant association significant dissociation

4.3 Discussion

Antigenic typing of organisms can serve many purposes. Reference labs use it to index strain types that are associated with disease and can detect any emerging strains that may be related to increases in disease incidence, vaccine escape or antibiotic resistance. In epidemiological situations it can be used to detect linkage of disease cases and thus any potential outbreak. Also, it can inform the choice of intervention used, *e.g.* prophylaxis such as immunisation and administration of antibiotics. MLST can identify the clonal grouping that an isolate belongs to and whether it is one of the main virulent lineages. However, because it indexes polymorphism in relatively slowly evolving genes, on its own it may not offer enough discrimination for short term epidemiology. Isolates that share the same ST may appear epidemiologically linked but may have different antigenic markers. As antigen genes evolve much more rapidly and are often very diverse they offer a level of discrimination at the short-term level that MLST typing alone may not. Therefore, antigen typing either serologically or genotypically may be a more suitable first step in this case. Finetyping (serogroup:PorA:FetA) has been used successfully in spatio-temporal analyses of invasive disease in Germany to retrospectively identify clusters of disease using space-time scan statistics (Elias *et al.*, 2006). This has been further extended to real-time Geographical Information System (GIS) surveillance at the national level in Germany with EPIScanGIS which can potentially detect emerging clusters of disease using finetype data (Reinhardt *et al.* , 2008).

Appropriately sampled isolate datasets as part of epidemiological surveys or studies of population biology can offer insights into meningococcal population dynamics and transmission systems. Practical applications of typing data include the formulation of vaccine recipes based on the main disease-associated types and the measurement of the success of vaccine interventions. Vaccines in development include multivalent PorA recipes that are based on the main disease-associated types in Europe (van den Dobbelsteen *et al.*, 2007). This study looks at the distribution of two vaccine candidates in a European disease isolate collection. Examination of the diversity of these antigens in Europe can help inform the choice of variants to include in a vaccine that will offer the maximum coverage possible. Chapter 5 discusses the principles behind a vaccine formulation based on these antigens.

Meningococcal antigens are highly heterogeneous due to positive selection imposed on them by their exposure to the host immune system and high rates of recombination. Antigenic diversity is one of the mechanisms by which the organism can evade host immunity. Previous studies have shown that in spite of this diversity, a relatively small number of types account for most disease and the number of antigen combinations is relatively small. In this dataset there was an uneven distribution of types found and many fewer combinations than would be expected. Given the number of VR1 and VR2 types found, the number of combinations of these, the PorA type, was a theoretical 7020 (54 VR1 x 130 VR2). Instead in this dataset there were 273 PorA types found. There was also an uneven distribution of PorA/FetA combination with 604 found out of a theoretical 27027 (273 x 99). Both antigens displayed a similar pattern of distribution with a small number of types accounting for most isolates. The higher diversity of the VR2 antigen compared to VR1 and FetA may be a consequence of higher immunogenicity of this VR (Martin *et al.*, 2000b; Vermont *et al.*, 2003). There were a large number of unique types found for each antigen. Of the PorA types found, 53% (146 types) were found only once in the whole dataset. For FetA, 36% (36 types) were found just once. These types are likely to appear only rarely and disappear as they are probably less successful.

In keeping with other surveys of OMP distribution in representative disease samples, there was evidence of non-random structuring of antigen distribution, *i.e.* non-overlapping combinations (Russell *et al.*, 2008; Urwin *et al.*, 2004). This phenomenon cannot be explained by purely neutral processes and instead is a signature of host immune selection acting on immunogenic epitopes creating a structured population (Gupta *et a l.*, 1996). This model of strain structuring imposed by host immunity postulates that a dominant immune response against a polymorphic determinant (PorA, FetA) which has moderate to strong cross-protection among genotypes sharing the same alleles at strain loci will create a non-overlapping population dominated by subset of strains. These strains also can persist over long periods of time in spite of much recombination. This has been shown particularly in the case of PorA VR associations which have shown persistence over wide time spans and global spread (Russell *et al.*, 2008; Urwin *et al.*, 2004). This persistence is predicted to be observed more in the case of pathogens such as the meningococcus which have relatively short infectious periods. Recombination will continue to create strains that share alleles with the dominant

strains but they will be at a disadvantage due to host immune response against them and may exist in relatively small numbers.

In this study, PorA VR combinations in particular showed very strong non-overlapping structuring. The top seven PorA VR combinations did not share any alleles amongst them and together accounted for 57% of isolates. There were much smaller numbers of types that shared one allele with any of these major types as they are likely to be much less successful due to host selection acting against them. F_{ST} results indicated a significant level of structuring in Europe so it could be that these subordinate types have emerged or are found in regions/transmission systems where the dominant type is less prevalent and therefore there is less of a disadvantage to share an allele. For example, the P1.5-1,10-4 type was significantly associated with Scotland while the dominant P1.5-1,10-8 type was not. There was less of an overlapping distribution of PorA/FetA combinations. For example, in the top five PorA/FetA strain types, P1.5,2 had two FetA associations (F1-1 and F3-6) and F3-6 had two PorA associations (P1.5,2 and P1.5-1,10-8). This may be due to less immunogenicity against the FetA antigen leading to less intense host immune selection acting on the meningococcal population. Differences in the immunogenicity of PorA types and VRs within PorA types have been noted (Luijkx *et al.*, 2003; Martin *et al.*, 2000b; Vermont *et al.*, 2003) which may in turn impact the host-immunity-directed population structure (Russell *et al.*, 2008; Urwin *et al.*, 2004).

Structuring in the dataset was also extended to association with clonal complex. The major PorA and FetA types were mainly only associated with one lineage although this relationship was not absolute. Clonal complex is not predictive of antigen type and *vice versa*. For example, P1.5,2 was mostly associated with ST-11 complex but was also associated with the related genotype ST-8 complex, and in much smaller numbers, a number of other complexes. The ST-32 complex was associated with two major PorA types: P1.19,15 and P1.7,16. Similarly, ST-32 complex was mainly associated with FetA types F3-3 and F5-1. Many of the antigen-complex associations found in this work have been seen elsewhere, further underlining the persistence during spread of these strain types in diverse geographical locations and over long time periods. For example, a study of global disease isolates found combinations of antigens and clonal complexes across a large time span and in diverse locations (Urwin *et al.*, 2004). These associations with lineage can shift over time however and has been demonstrated in models to be the result of competition among lineages due to immune selection (Buckee

et al., 2008). A study of disease in England and Wales from 1975–1995 showed the cycling of genotypes and antigenic types even in the absence of any major interventions such as the MCC campaign (Russell *et al.*, 2008). Recombination is a tool used by the meningococcus to evade host immunity and shifts in antigenic repertoires can be associated with increased disease. In the United States in the mid 1990s an increase in serogroup Y ST-23 complex related disease was associated with antigenic shift by horizontal gene transfer at three OMP loci – PorA, PorB and FetA (Harrison *et al.* , 2006).

Previous studies of European meningococcal disease and carriage collections have shown significant geographic structuring (Claus *et al.* , 2005; Yazdankhah *et al.* , 2004). Similarly, in the present work, while there was a wide spread of the major PorA and FetA types across Europe, there was a degree of geographic structuring that played a part in overall antigen genetic diversity; 3.3% and 4.8% of PorA and FetA diversity could be attributed to geographic differences. Significant F_{ST} values indicate a degree of restricted gene flow among countries pointing to more transmission among than between countries. Similar values were obtained for F_{ST} calculations using MLST data underlining the congruence of antigen and clonal lineages. There was less evidence of temporal structuring, but as the span of this study was just three years, any major shifts in antigen prevalence may not be picked up. The continuation of surveillance is therefore essential to detect any emergent types, especially in the wake of OMP vaccine interventions.

In summary, the two vaccine candidates PorA and FetA have been found to be very diverse in a relatively recent sample of disease-associated European meningococcal isolates. This heterogeneity was structured however. A small number of types of both antigens accounted for most disease. Non-overlapping combinations of antigens are consistent with selection by host immunity and restrict the main circulating strains to a small number of dominant types found across Europe with some locally restricted gene flow. These dominant antigens were also associated with clonal lineages which are associated with disease. These strain types have been shown to persist over wide geographic and temporal spread. Together with their immunogenicity, these features make these antigens attractive vaccine candidates as a relatively small number of types could have the potential to give broad coverage that may have long range potential in terms of time and distance. This will be discussed further in the following chapter.

CHAPTER 5: Using population biology to inform vaccine design

5.1 Introduction

Single clone epidemics in New Zealand, Norway and Cuba have successfully been controlled by the use of vaccines containing their respective epidemic PorA type antigen (Bjune *et al.* , 1991; O'Hallahan *et al.* , 2005; Rodriguez *et al.* , 1999). In the endemic setting this would be inappropriate, because of the large diversity of this particular antigen, there would be protection against only a relatively small portion of the meningococcal population. Consequently, the two potential routes for vaccine development are to (i) use more conserved antigens or (ii) use a cocktail of diverse antigens. To date, conserved antigens such as NadA and NspA have been shown to be protective, but vary in expression or may not be expressed in all isolates tested (Comanducci *et al.* , 2002; Moe *et al.* , 1999). The diversity of the antigens PorA and FetA point to strong immune selection and indeed both are known to elicit human immune response and are expressed in the majority of meningococci. Therefore, despite their diversity they are attractive as antigen candidates.

Also particularly relevant in vaccine development is the finding that PorA and FetA antigenic diversity is structured in a number of ways. Firstly, a relatively small number of types account for most disease with a large number of types that may appear only once or a few times and only transiently. Secondly, this uneven distribution also manifests in non-overlapping antigen combinations which is consistent with models of structure generated by host immune selection. In spite of much recombination these antigen associations persist and are stable over time and distance (Gupta *et al.* , 1996; Urwin *et al.* , 2004). Thirdly, there are associations with respect to clonal complexes which also persist over time and place but which can shift in a way that is consistent with competition among lineages (Buckee *et al.* , 2008; Urwin *et al.* , 2004). These factors combine to potentially simplify vaccine formulation, meaning that a small number of antigen types in a recipe could potentially provide coverage against the

majority of disease-associated lineages with the possibility of long-range vaccine protection due to persistence of antigen types in the population.

The aim of this chapter was to outline the use of molecular epidemiology as a rationale for the design of a PorA/FetA based-vaccine. PorA and FetA data from a representative European disease dataset, described in Chapter 4, were employed to formulate a vaccine recipe containing a cocktail of antigen variants that would provide the best coverage with regard to disease in Europe and by extension potentially other global regions. Epidemiological information collected at the same time was merged with this antigen data so further detailed analysis of the potential coverage of the vaccine could be carried out.

5.2 Merging EU-MenNet data with EMEC epidemiological database

The EU-MenNet disease isolate collection sequence data (MLST, PorA, FetA) were merged with epidemiological information (serogroup, age, sex) collected separately for the isolates by the EMEC/EU-IBIS (Trotter *et al.*, 2007) where possible. This allowed for further analysis of the EU-MenNet data and investigation of the potential coverage of a vaccine recipe among epidemiological categories such as age and serogroup.

5.2.1 Breakdown of combined dataset

A total of 3705 of the samples were linked to EMEC/EU-IBIS data. Serogroup information was available for 3700 (88.5%) isolates. The most common serogroup in the EU-MenNet dataset was serogroup B which accounted for 65.1% (2412) of disease isolates (Figure 5.1). Serogroup C accounted for 27.7% (1025) of isolates, so together these two serogroups accounted for almost 93% of all disease samples. After B and C, W-135 accounted for 3.9% (141), Y 2.1% (76), non-groupable 0.5% (17), 29E 0.2% (17), A 0.2% (6), X 0.1% (5) and serogroup Z 0.03% (1).

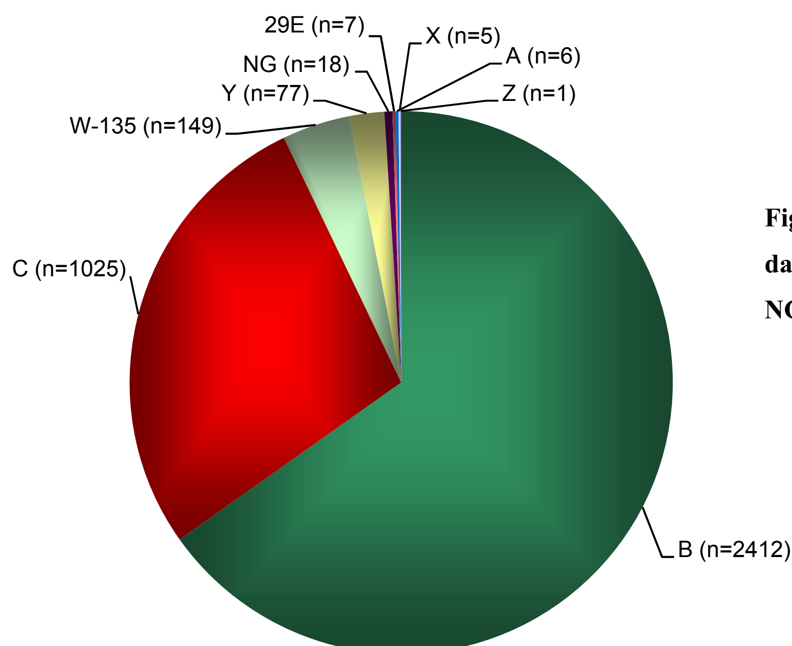


Figure 5.1: Breakdown of dataset by serogroup. NG=non-groupable.

Along with complete genotypic characterisation for MLST, PorA and FetA loci, information was available for 2456 (58.7%) isolates. A total of 1130 unique strain types (serogroup, PorA, FetA and ST) (Jolley *et al.* , 2007) were present among these 2456 isolates, giving a diversity index of 0.989. Of the isolates for which a complete strain type was available, a total of 15 (1.3%) strain types were observed 20 times or more, accounting for 794 (32.3%) of these isolates. Of the remaining strain types, 896 (36.5%) were observed only once. Together with isolates belonging to the same clonal complex, the 13 most prevalent strain types accounted for 43% of isolates. There were 727 unique serogroup, PorA, FetA and clonal complex strain types (Figure 5.2). A total of 17 types had a frequency of more than 20 and accounted for 47% of isolates and 504 unique types accounted for 21% of isolates. The most prevalent strain type was B:P1.7-2,4:F1-5:(cc41/44) which accounted for 12% (288) of isolates. The next most prevalent type was B:P1.7,16:F3-3:(cc32), which accounted for 5% of isolates.

Diversity indices of strain types in European countries ranged from 0.596 in Iceland to 0.983 in Germany although isolates were only available for the year 2001 for Iceland and Portugal. F_{ST} values of pairwise comparisons of strain types by country indicated a significant level of structuring of strain types in Europe ($p<0.05$, Table 8, Appendix). There were significant associations of particular strain types and certain countries (Table 9, Appendix). For example, B:P1.7-2,4:F1-5:ST-41(cc41/44) and C:P1.5,2:F3-6:ST-11(cc11) were associated with Belgium. Strain type B:P1.7,16:F3-3:ST-32(cc32) was associated with Denmark; W-135:P1.5,2:F1-1:ST-11(cc11) with France; C:P1.5-1,10-8:F3-6:ST-11(cc11) with Iceland; C:P1.5,2:F3-3:ST-11(cc11) with Germany; B:P1.7-2,4:F1-5:ST-42(cc41/44) with the Netherlands and B:P1.22,14:F5-5:ST-213(cc213) with England and Wales (all $p<0.00007$).

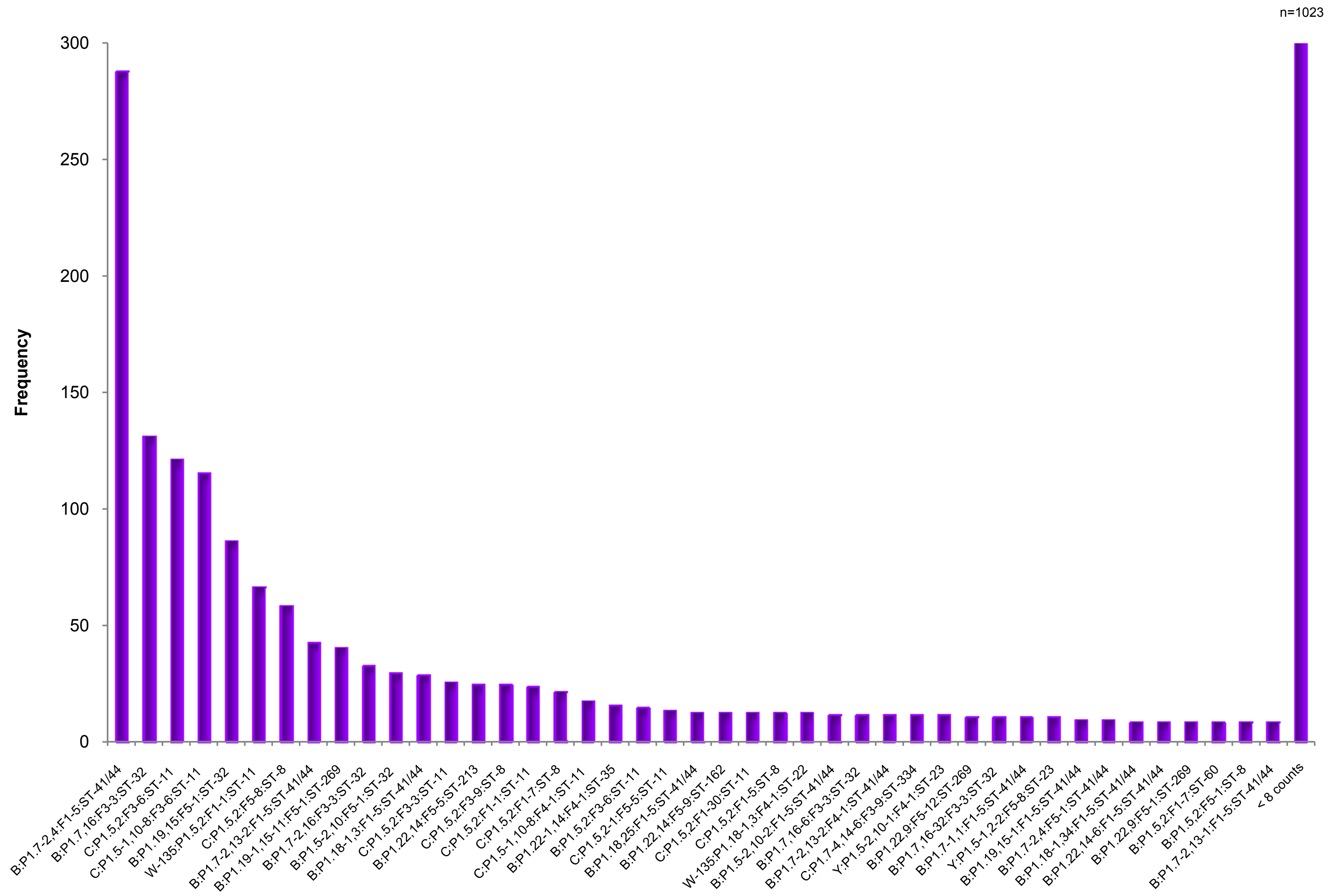


Figure 5.2: Prevalence of strain types (serogroup:PorA:FetA:cc).

5.2.1.1 Age group clonal complex association

Age data were available for 3226 isolates. The ST-11 complex was more associated with those over four years, while ST-41/44 complex was more associated with those nine years and under (Figure 5.3). Overall χ^2 tests were performed for each complex represented by 50 or more isolates. Significant age effects were seen for the ST-41/44, ST-11, ST-32, ST-8, ST-23 complexes, and those unassigned to any complex. Individual χ^2 and logistic regression tests were then performed for each age range for these complexes (Tables: 5.1 and 5.2).

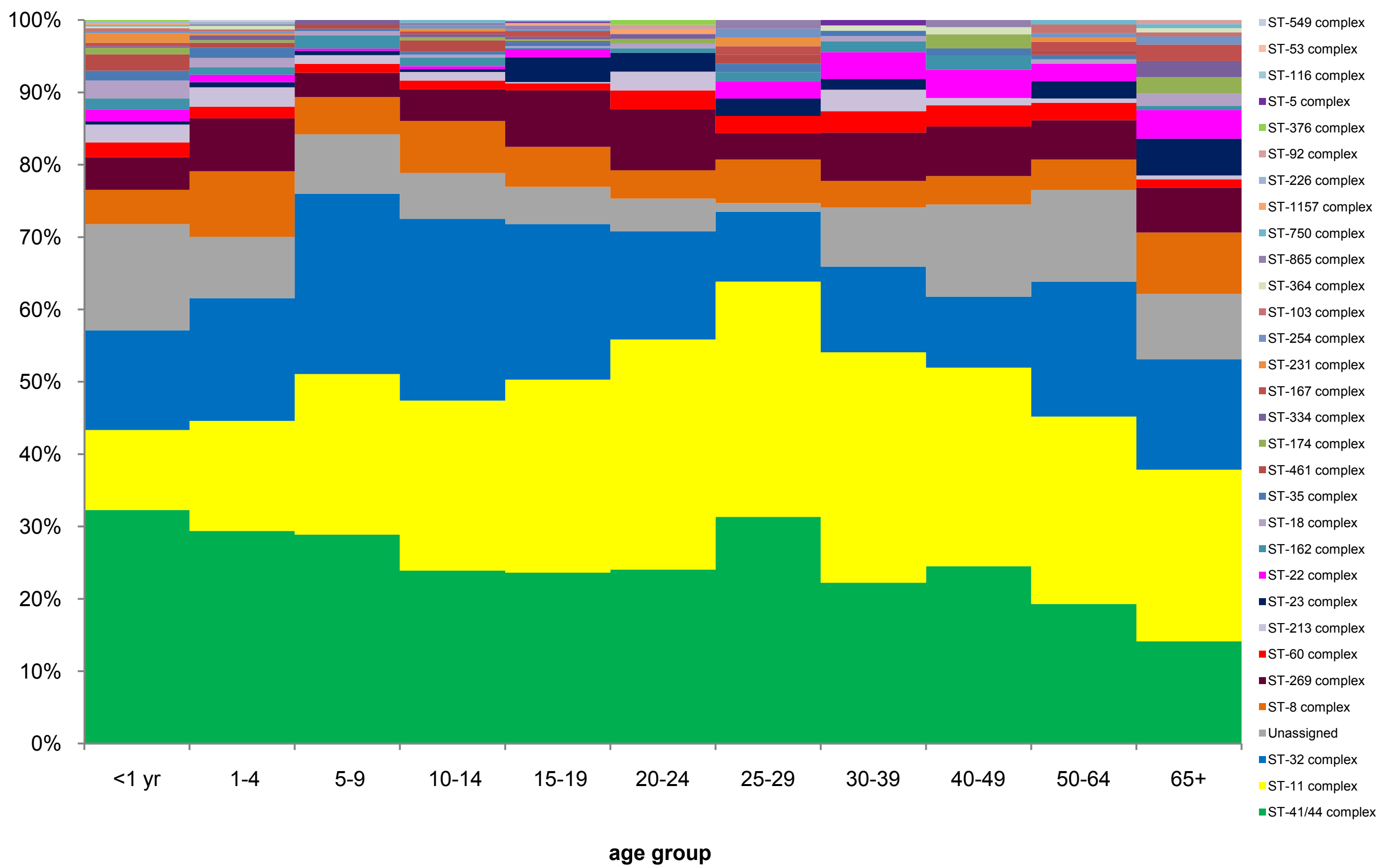


Figure 5.3: Distribution of clonal complexes by age group.

Table 5.1: Breakdown of clonal complex by frequency and association with age group.

clonal complex	age range									
	<1 yr	1–4	5–9	10–14	15–19	20–24	25–44	45–64	65+	total
ST-41/44	143	253	95	60	124	37	68	45	25	850
ST-11	49	131	73	59	140	49	79	62	42	684
ST-32	61	146	82	63	113	23	33	32	27	580
ST-8	21	78	17	18	29	6	10	11	15	205
ST-269	20	63	11	11	41	13	16	12	11	198
ST-60	9	14	4	3	5	4	8	5	2	54
ST-213	11	23	4	3	1	4	5	1	1	53
ST-23	2	6	2	1	18	4	4	4	9	50
Unassigned	65	73	27	16	27	7	18	28	16	277
Others	62	74	14	17	27	7	22	23	29	275
total	443	861	329	251	525	154	263	223	177	3226

p<0.0056 after Bonferroni correction: significant association/dissociation *versus* expected

Table 5.2: Breakdown of clonal complex and association with age group by odds ratio with and without controlling for serogroup (in brackets).

clonal complex	age range								
	<1 yr	1–4	5–9	10–14	15–19	20–24	25–44	45–64	65+
ST-41/44	1.58	1.48(1.20)							0.49
ST-11	0.42(0.61)	0.61(0.66)			1.34	1.69(1.71)	1.61(1.67)	1.43(1.78)	
ST-32	(0.55)		1.85(1.92)	1.68(1.75)	1.38(1.59)		0.68		
ST-8		1.47(2.00)			(0.61)	(0.39)	(0.41)		
ST-269	(0.58)				(1.59)				
ST-60							(2.57)		
ST-213		1.75			0.09(0.11)				
ST-23		0.32			2.51				3.55
Unassigned	2.01(1.82)				0.52(0.57)				(1.66)

significant association/dissociation; all p<0.05

The ST-41/44 complex was over-represented in the <1 year and under-represented in the 65+ year age ranges (Table 5.1), while the ST-11 complex was under-represented in the under five year olds but over-represented in the 15–44 year age range. Increased ST-32 complex disease was associated with the 5–14 year age group and there was increased ST-8 complex-related disease in the 1–4 year age range. The ST-23 complex was significantly associated with the 15–19 years old age group and the over-65s. STs unassigned to any complex were over-represented in children <1 year old and under-represented in the 15–19 year age range.

A sense of magnitude of associations and independence of the association from serogroup was obtained by performing logistic regression (Table 5.2). The odds of ST-41/44 complex being associated with the under ones was 1.48 (95% CIs 1.27–1.96) and 1.58 for the 1–4 years age group (95% CIs 1.26–1.74). Controlling for serogroup, the greater odds of disease associated with this serogroup for the 1–4 years age group remained but was close to unity (OR 1.20; 95% CIs 1–1.43). Those more than 65 years of age were about half as likely to have disease associated with this complex as other age groups (OR 0.49; 95% CIs 0.32–0.76). Between the ages of 15 and 64 the odds of disease associated with ST-11 *versus* other complexes was between 1.34 and 1.69 times. The association remained after controlling for serogroup for ages between 20 and 64 (1.67–1.78). For the under ones, the odds of disease associated with ST-11 complex was 0.42 (95% CIs 0.31–0.57) and 0.61 times (95% CIs 0.50–0.74) for the 1–4 years age group. This negative association remained for both groups after controlling for serogroup (OR 0.61; 95% CIs 0.4–0.91 and OR 0.66; 95% CIs 0.51–0.86). ST-32 complex was between 1.38 and 1.85 times more likely to be associated with the 5–19 age group than other age groups. This association was independent of serogroup with the odds increasing from between 1.59 and 1.92.

Other associations of note included the odds of ST-8-associated disease in the 1–4 age group were twice that of other age groups independent of serogroup (95% CIs 1.48–2.71). The ST-23 complex was 2.51 (95% CIs 1.45–4.33) and 3.55 (95% CIs 1.73–7.28) times more likely to be associated with the 15–19 years and 65+ age groups respectively. STs unassigned to any clonal complex were, independent of serogroup, 1.82 (95% CIs 1.30–2.53) and 1.66 (95% CIs 1.03–2.67) times more likely to be associated with the <1 years and 45–64 years age groups respectively and about half as likely to be associated with the 15–19 years age group (OR 0.57; 95% CIs 0.33–0.82).

5.2.1.2 Age group antigen association

Serogroup B was proportionally more associated with the under four years age group (76%) *versus* older age groups while serogroup C was more associated with the over fours (Figure 5.4). The non-BC serogroups were particularly associated with the 65+ age group, where 18% of the disease isolates from this group were non-BC.

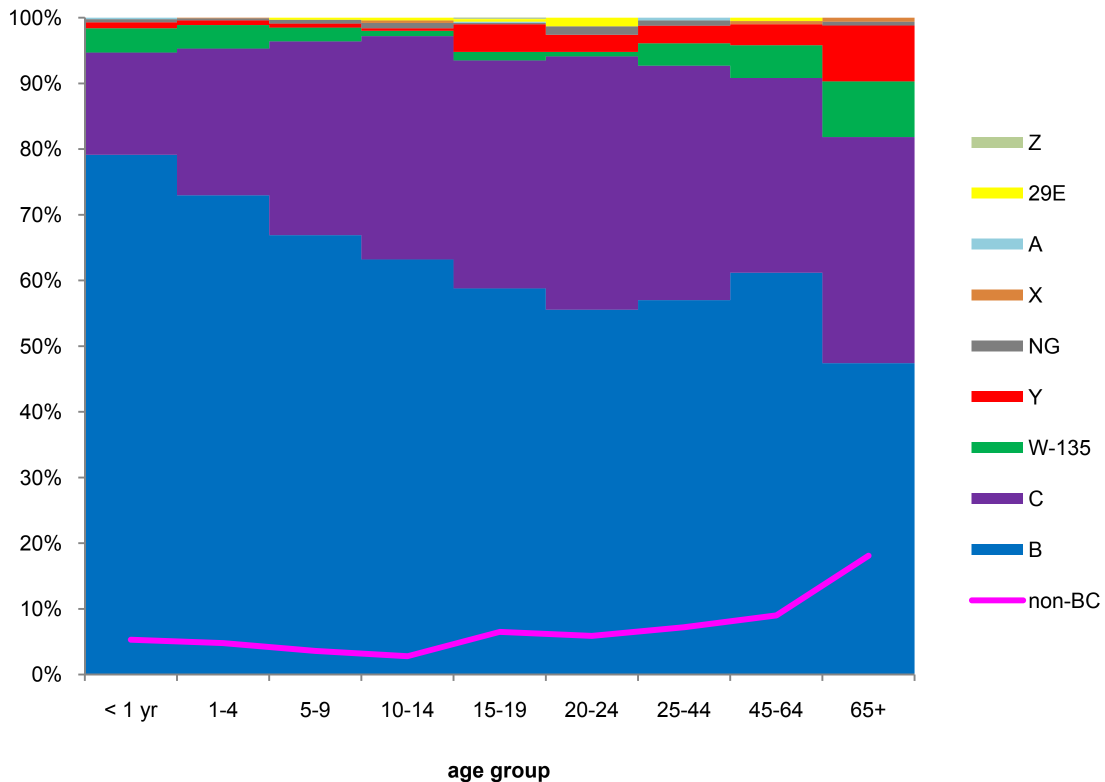


Figure 5.4: Breakdown of serogroup by age group.

χ^2 tests were carried out to test for association of antigen and age group. There were no significant associations between PorA type age group (Figure 5.5) by the χ^2 test. The F3-6 FetA type was significantly under-represented in under five year olds and significantly over-represented in the 15–19 year age group (χ^2 test, both $p<0.005$) (Figure 5.6). Logistic regression was carried out to determine magnitude of association and the independence of association from serogroup and clonal complex. The odds of F3-6 being associated with the under ones was 0.31 (95% CIs (0.18–0.52)). This negative association was independent of serogroup and clonal complex (OR 0.37; 95%

CIs 0.19–0.72). The odds of F3-6 being associated with the 1–4 years age group was 0.46 (95% CIs 0.33–0.64). This negative association was independent of serogroup and clonal complex although almost crossed unity (OR 0.63; 95% CIs 0.40–0.996). The odds of F3-6 being associated with the 15–19 years age group was 1.97 (95% CIs 1.47–2.64). This positive association was independent of serogroup and clonal complex (OR 1.90; 95% CIs 1.19–3.02).

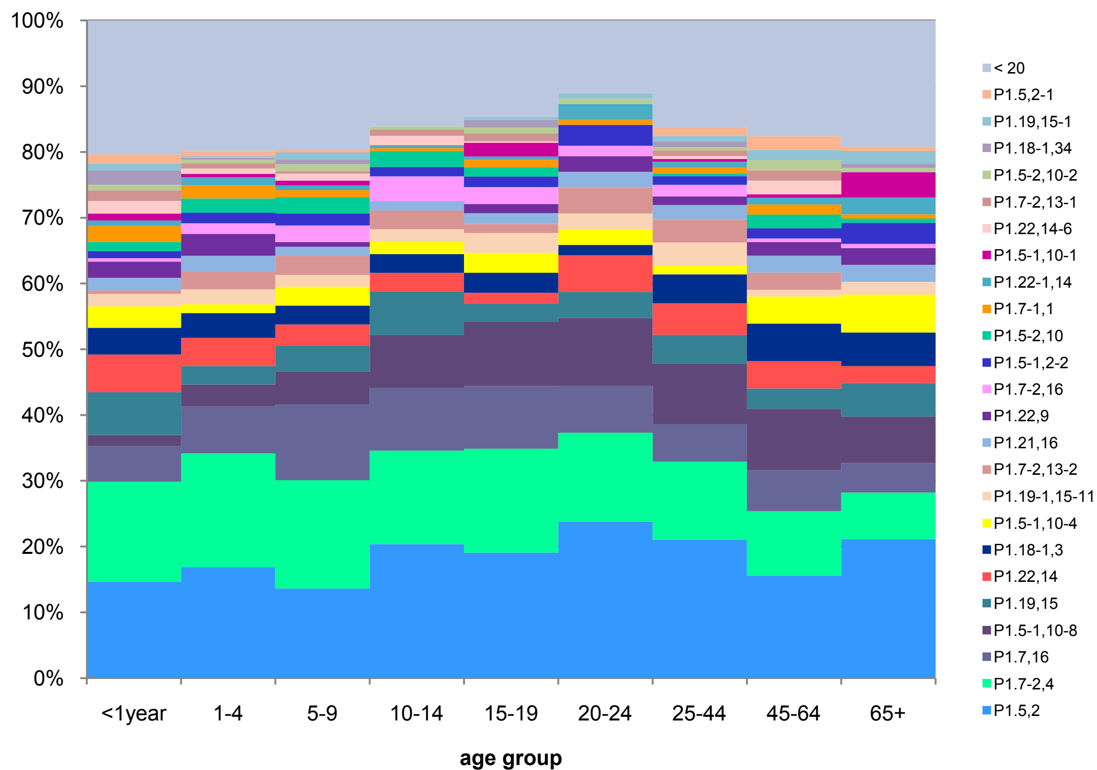


Figure 5.5: PorA distribution by age group.

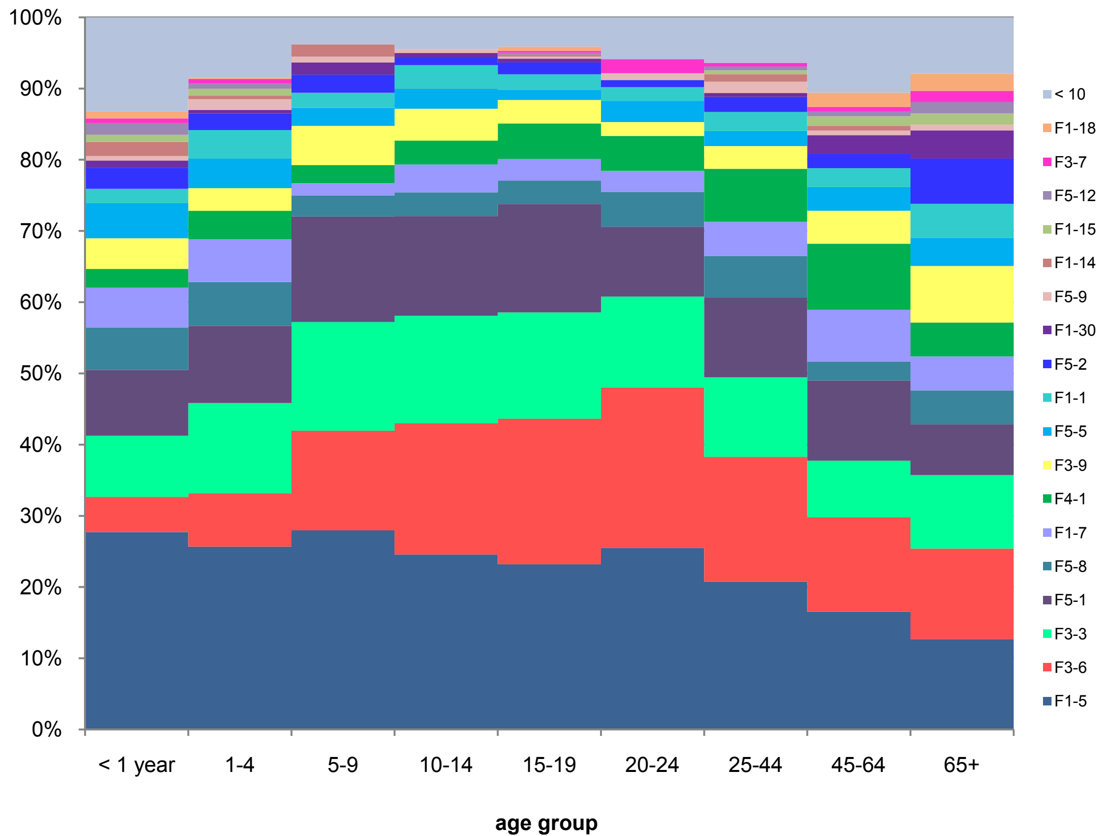


Figure 5.6: FetA distribution by age group.

5.3 Rationale for PorA/FetA vaccine

PorA and FetA are considered suitable candidates to include in an OMV vaccine formulation due to their natural presence in OMVs, their expression in most meningococci, their immunogenicity, including their ability to elicit bactericidal antibodies, and also their structured antigenic diversity (Jodar *et al.*, 2002; Urwin *et al.*, 2004; Wedege *et al.*, 1998). This structuring manifests in a relatively small number of PorA and FetA types accounting for most disease with a large number of transient types. The association of antigenic epitopes appears in a non-random non-overlapping distribution which is consistent with host immune selection (Gupta *et al.*, 1996). In spite of much recombination these antigen associations persist and are stable over time and distance (Gupta *et al.*, 1996; Urwin *et al.*, 2004). In the EU-MenNet disease dataset, this antigenic structuring was apparent with a non-random distribution of epitopes significantly different from the expected null hypothesis of random association

of types. The major hyperinvasive lineages were present across Europe although with some variation in frequency distributions (Chapter 3). These main lineages were also shown to be associated with PorA and FetA antigen types (Chapter 4) and were made up of a relatively small number of types. These associations have been shown to also be relatively stable over time and distance but can shift, which may be explained by competition among lineages (Buckee *et a l.*, 2008; Urwin *et al.* , 2004). These factors together help to simplify vaccine formulation, meaning that a small number of antigen types in a recipe can provide coverage against the majority of disease-associated lineages with potential for long-range vaccine protection due to persistence of antigen types in the population.

Using the PorA and FetA sequence typing data generated from the EU-MenNet study, it was possible to design a multivalent vaccine formulation that would correspond to the most common antigens representing the main hyperinvasive lineages associated with disease in Europe (ST-11, ST-32, ST-41/44, ST-8 and ST-269). These antigens are also found in stable associations with each other so the most common PorA/FetA strain types in the recipe are non-overlapping. The vaccine recipe contains four each of PorA and FetA antigens (Table 5.3): PorA types P1.5,2; P17-2,4; P17,16; P1.19,15; FetA types F1-5, F3-6, F5-1 and F3-3. Together, these antigens make up 78.8% of the disease isolates in the EU-MenNet dataset which is also the theoretical coverage attained by a cocktail vaccine containing them. A specially-developed tool on the EU-MenNet database allowed the calculation of potential coverage of a given recipe containing a certain number of antigen components (one component consists of one PorA and one FetA antigen) and the number of epitopes covered, one, two or three of VR1, VR2 or FetA,. Potential coverage is the percentage of isolates that share at least one of the antigens, VR1, VR2 or FetA, with the recipe.

Table 5.3 shows the putative coverage of the proposed vaccine recipe. It is important to note that coverage may be underestimated, as it does not take into account any potential cross-protection there may be among antigen type families, *e.g.* members of the P1.5 family include P1.5, P1.5-2, P1.5-1 and so on. Therefore, 78.8% may be a baseline of the potential coverage of this recipe. A vaccine containing even just one PorA and FetA antigen gives coverage of almost 44%, whereas if just one of each of the PorA and FetA (P1.5,2 and F1-5) were included coverage would be just 20% and 22% respectively.

Table 5.3: Potential coverage of PorA/FetA recipe in EU-MenNet dataset with number of antigens covered.

no. of components	additive components			potential coverage	no. of antigens covered		
	VR1	VR2	FetA		1	2	3
1	5	2	F1-5	1226/2813 (43.6%)	670	533	23
2	7-2	4	F3-6	1605/2813 (57.1%)	569	544	492
3	7	16	F5-1	2138/2813 (76.0%)	860	740	538
4	19	15	F3-3	2216/2813 (78.8%)	697	664	855

Figure 5.7 shows that as the number of components in the recipe increases, the amount of triple antigen coverage also increases. This is likely to be beneficial since the density of deposited antibody on the meningococcal surface is liable to be greater, augmenting the vaccine's protective effect. The percentage of double and triple antigen coverage in the dataset for the four component vaccine was 30% and 38.6% respectively.

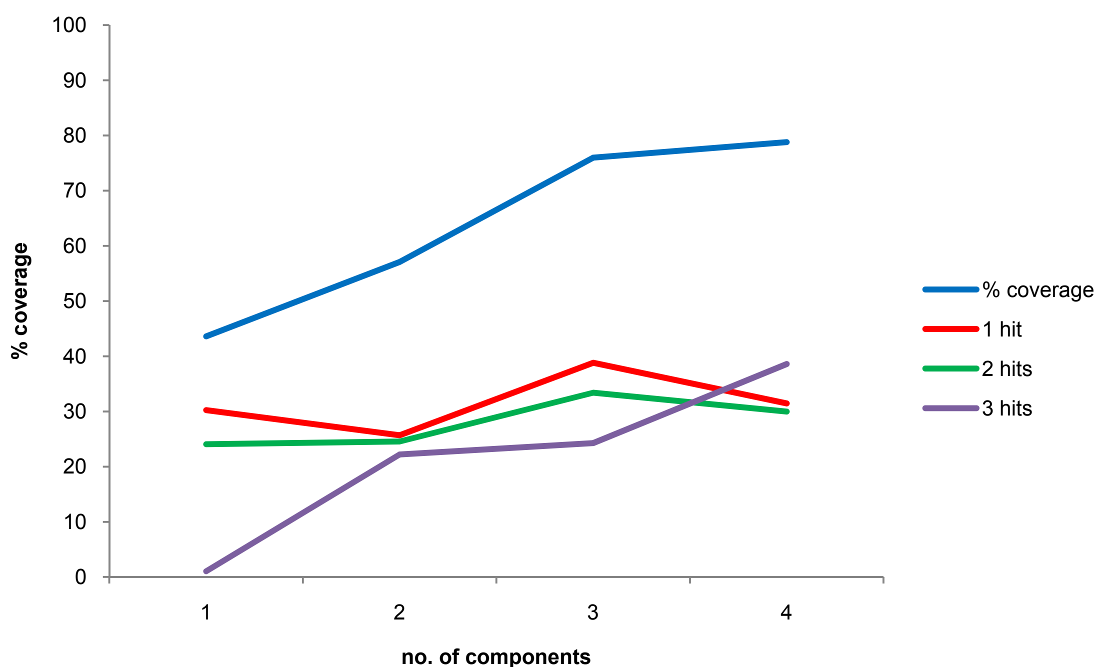


Figure 5.7: Potential coverage of PorA/FetA recipe in EU-MenNet dataset with number of antigens covered (hits).

5.3.1 Stability of vaccine antigen associations

As previously mentioned, there are associations between PorA and FetA antigen types and these also show associations with clonal complex and serogroup. The variants to be included in the recipe are no exceptions to this, showing both strong associations with each other and with clonal complex and serogroup. This is beneficial for the long-term effectiveness of the vaccine and increases the proportion of multiple antigen coverage with the formula. Table 5.4 shows the vaccine antigen associations and their make up with regard to complex and serogroup. The association with complex/serogroup was not absolute with a number of antigen combinations being found with different serogroups/complexes. These alternative associations were rare, however, and there were still dominant associations with perhaps the exception of P1.5,2: F3-6 having 88% serogroup C but 10% having serogroup B.

Examination of the publicly accessible PubMLST website and repository (http://pubmlst.org/) underlined the stability of the relationships over time and geographical distance. 96% of P1.7-2,4: F1-5 were serogroup B ST-41/44 complex with the earliest one found of this type from 1985. Isolates of this strain type were found in Europe, New Zealand and Chile. 96% of P1.19,15: F5-1 were serogroup B and ST-32 complex. Isolates of this strain type were found from 1983 to 2009 and from Europe, South Africa and Cuba. 90% of P1.7,16:F3-3 types were serogroup B and ST-32 complex, with isolates of this type being found from 1976 to 2009 and in various countries including USA, Brazil, Cuba, Cuba and in Europe. Again the P1.5,2: F3-6 pairing was slightly more diverse in that 89% were serogroup C ST-11 complex with 9% being serogroup B. The serogroup B strain types were found in the Netherlands in 1961 and 1962, the UK in 1995 and the Czech Republic in 2000.

Table 5.4: Breakdown of associated PorA/FetA vaccine recipe antigens with serogroup and clonal complex.

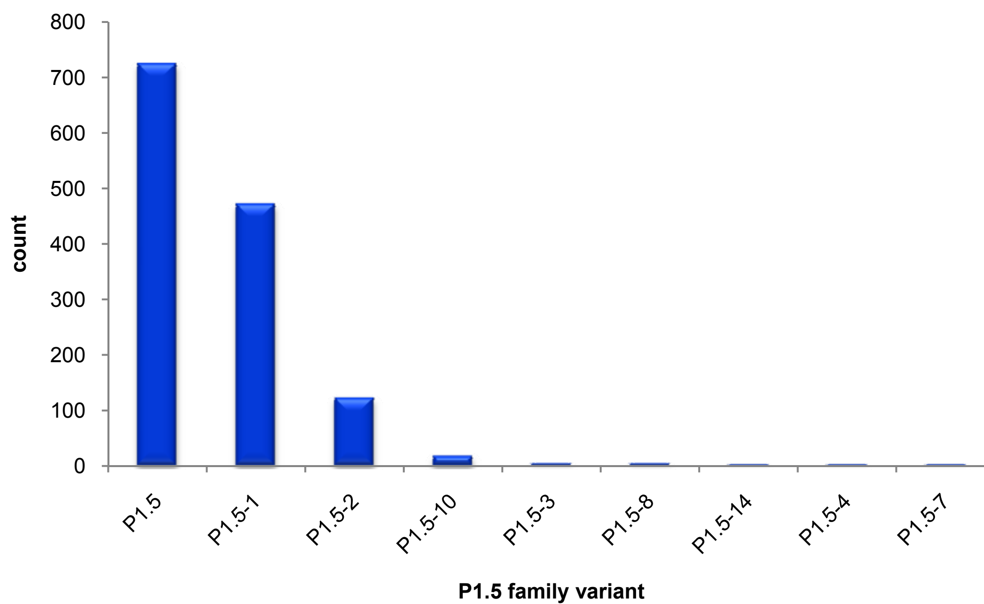
PorA	FetA	serogroup	clonal complex	count	%
P1.7-2,4	F1-5	B	ST-41/44 complex	288	97.6
		B	Unassigned	3	1
		C	ST-41/44 complex	3	1
		NG	ST-41/44 complex	1	0.3
				295	
P1.7,16	F3-3	B	ST-32 complex	131	97
		C	ST-32 complex	3	2.2
		NG	ST-32 complex	1	0.7
				135	
P1.5,2	F3-6	C	ST-11 complex	121	87.7
		B	ST-11 complex	14	10.1
		C	ST-8 complex	3	2.2
				138	
P1.19,15	F5-1	B	ST-32 complex	86	91.5
		B	Unassigned	4	4.3
		C	ST-32 complex	2	2.1
		W-135	ST-32 complex	1	1.1
		Y	ST-32 complex	1	1.1
				94	

5.3.2 Breakdown of vaccine antigen and associated variant family members

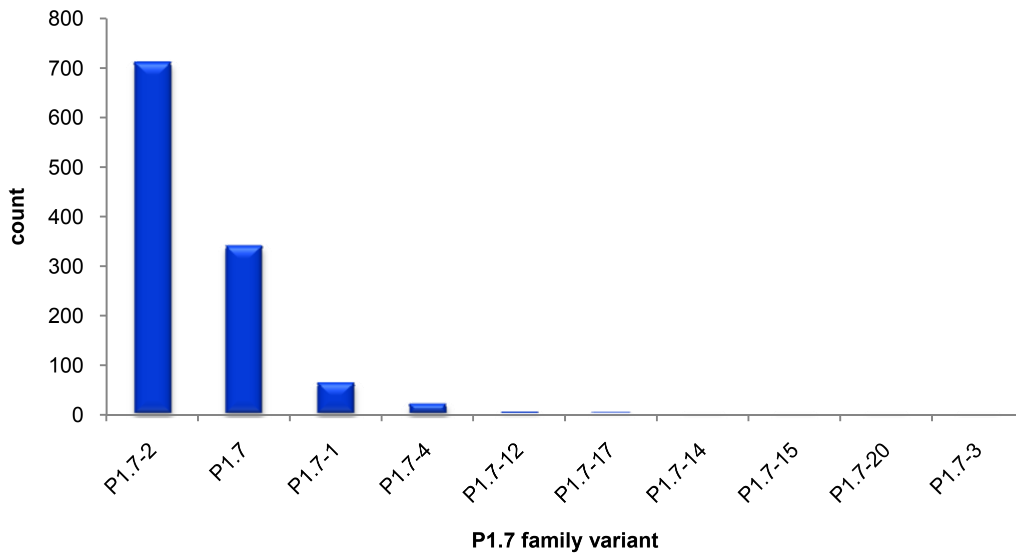
Cross-protection amongst members of antigen variant families is still to a large extent unknown and, as the vaccine recipe theoretical coverage does not assume cross-immunity, potential coverage may be some way above 79%. Each family is however dominated by one or two types with a larger number of types found very infrequently (Figures 5.8 (a) – (j)). Therefore, cross-protection may add only a relatively small amount of extra coverage. The PorA P1.7 and FetA F3 families each have two variant members included in the formula: P1.7 and P1.7-2 and F3-3 and F3-6 respectively. Antigen families particularly dominated by a single type include P1.2, P1.4, P1.16 and F1.

Figure 5.8 (a) – (j): Breakdown of each vaccine antigen-associated variant family.

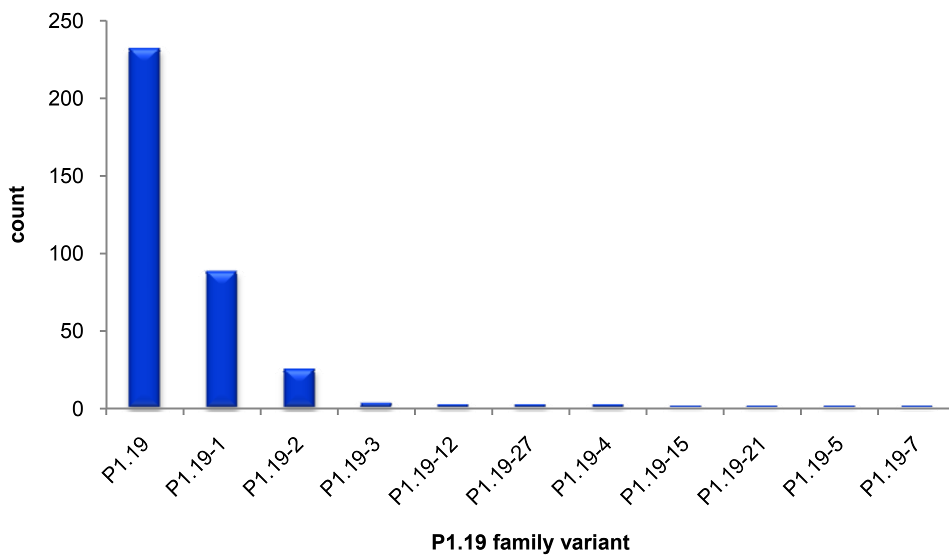
(a)



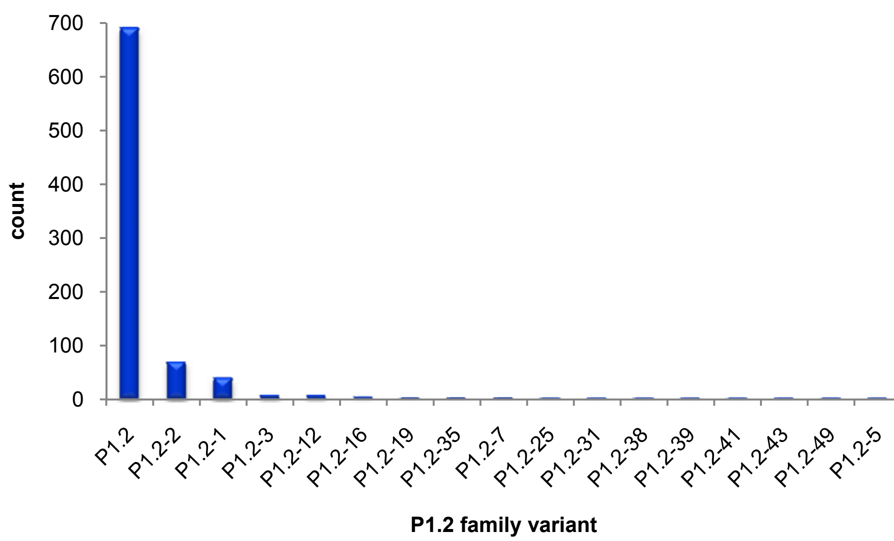
(b)



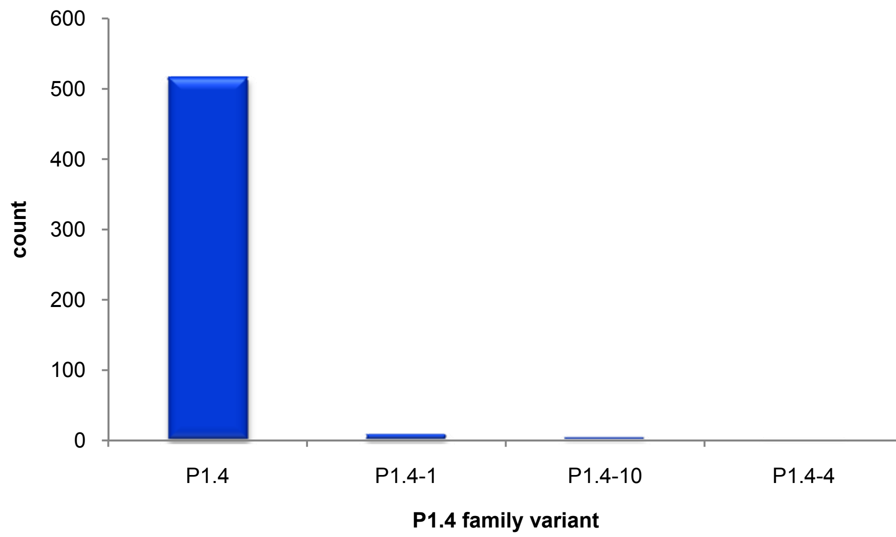
(c)



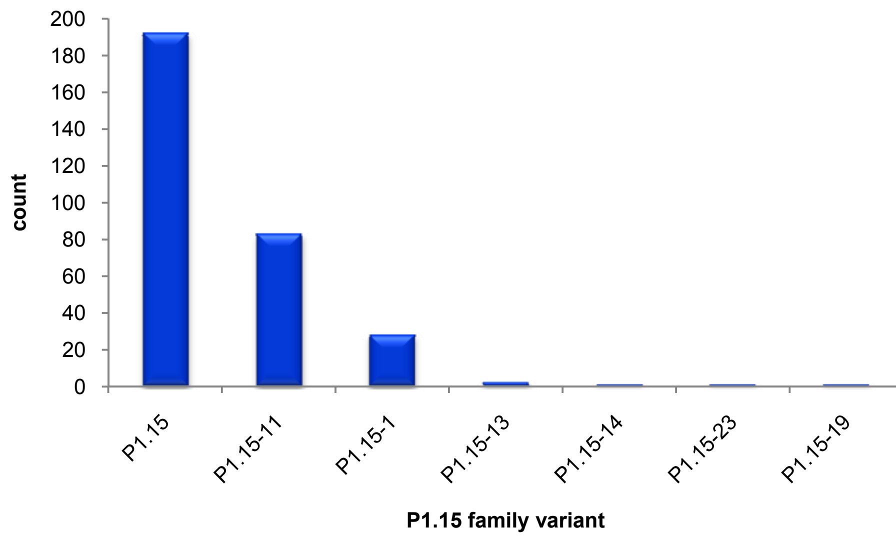
(d)



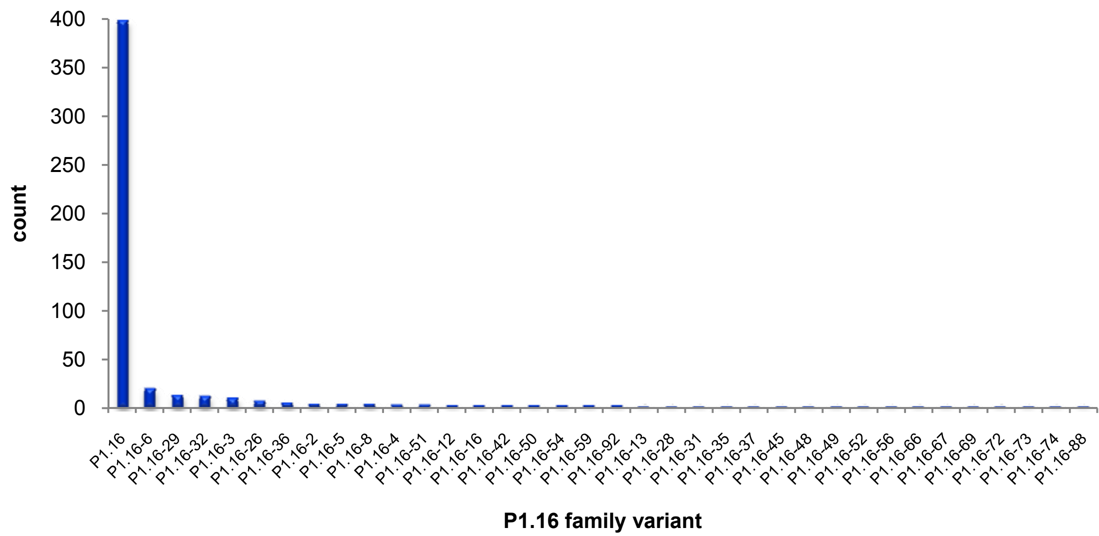
(e)



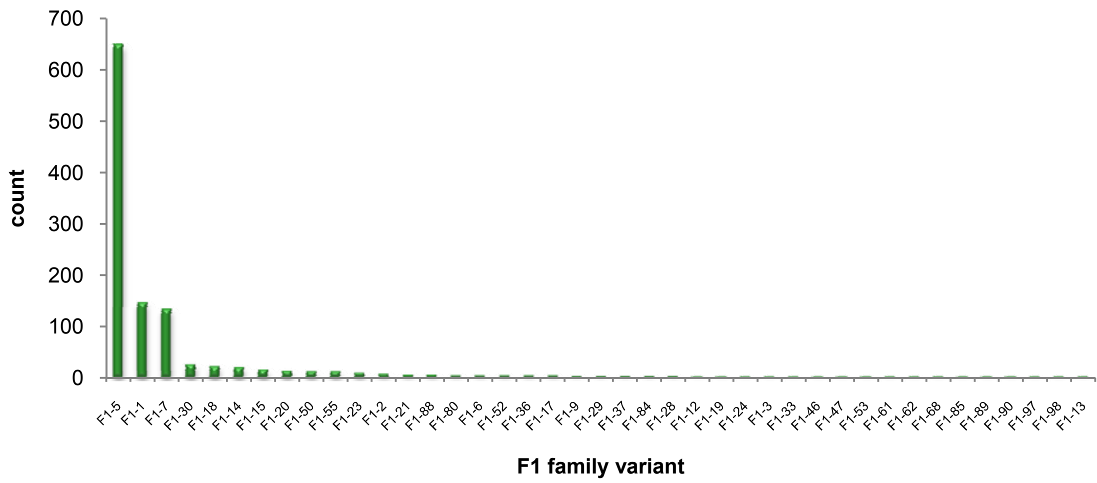
(f)



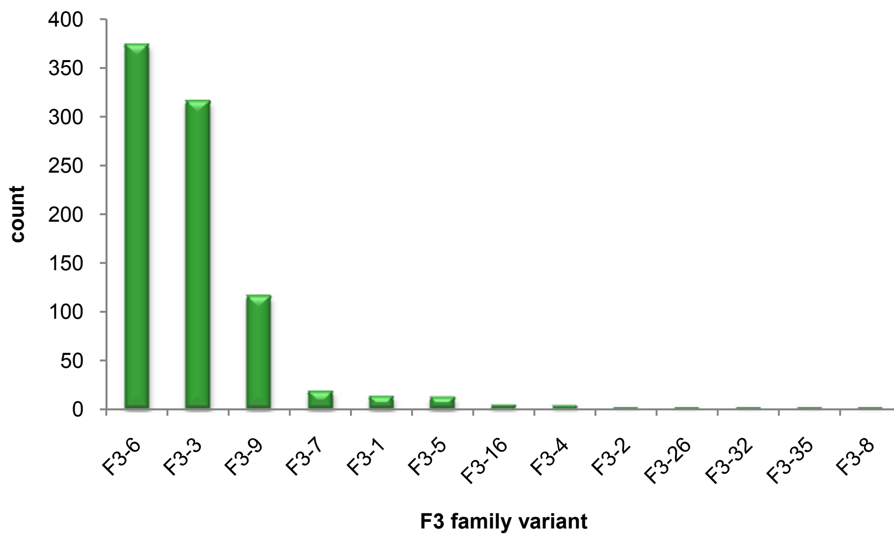
(g)



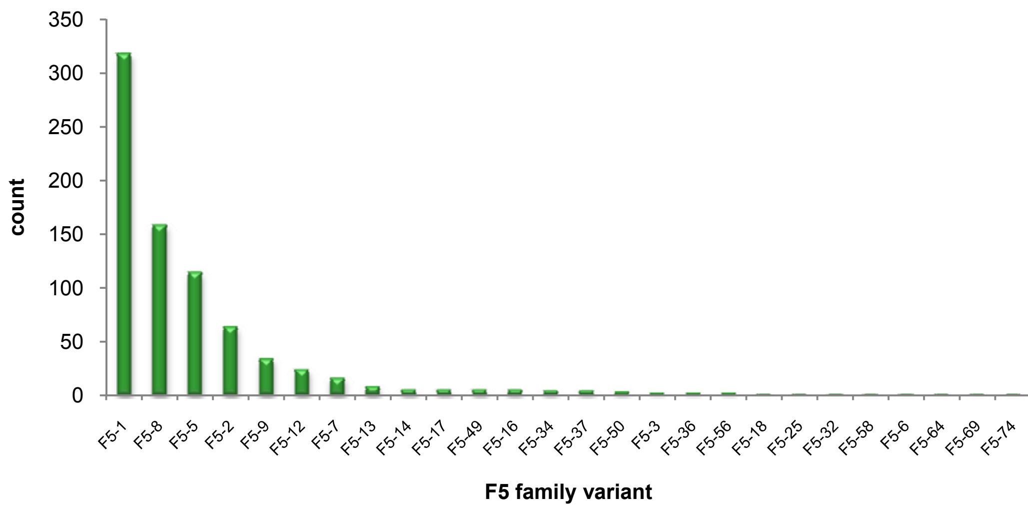
(h)



(i)

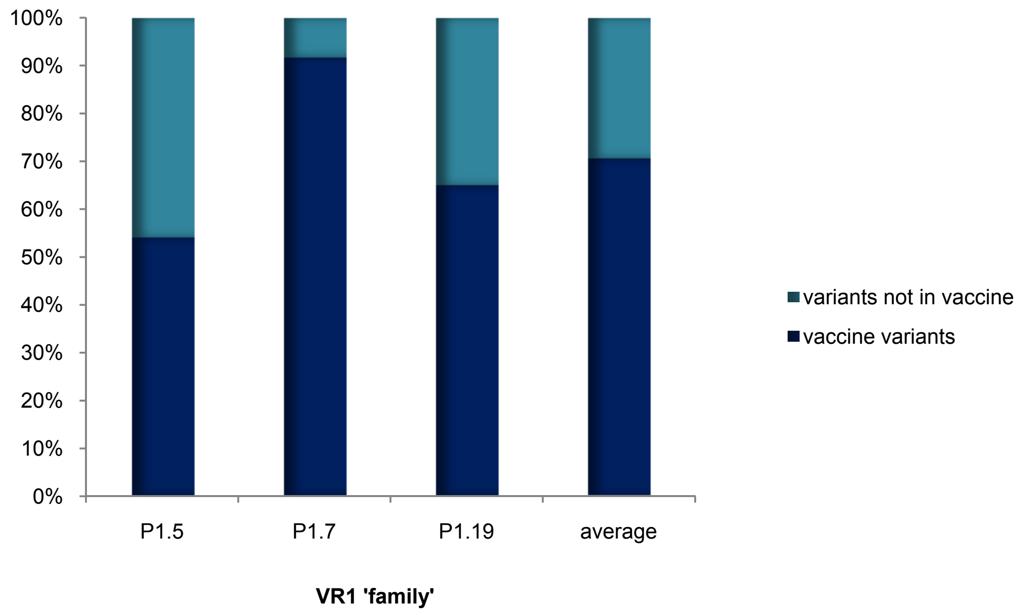


(j)

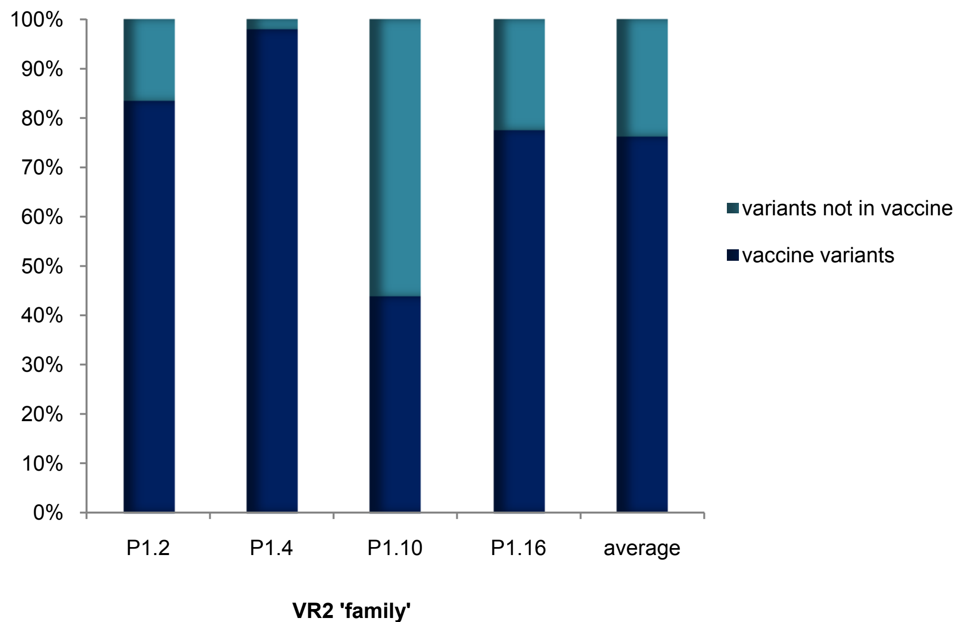


The vaccine antigens accounted for between 41% and 98% of their families (F5 and P1.4 respectively) (Figure 5.9 (a) – (c)). The VR2 vaccine antigens accounted for the most of their respective families (on average 80.2%). VR1 vaccine antigens on average accounted for 70% of their families and the FetA vaccine antigens accounted for 59.8% on average.

(a)



(b)



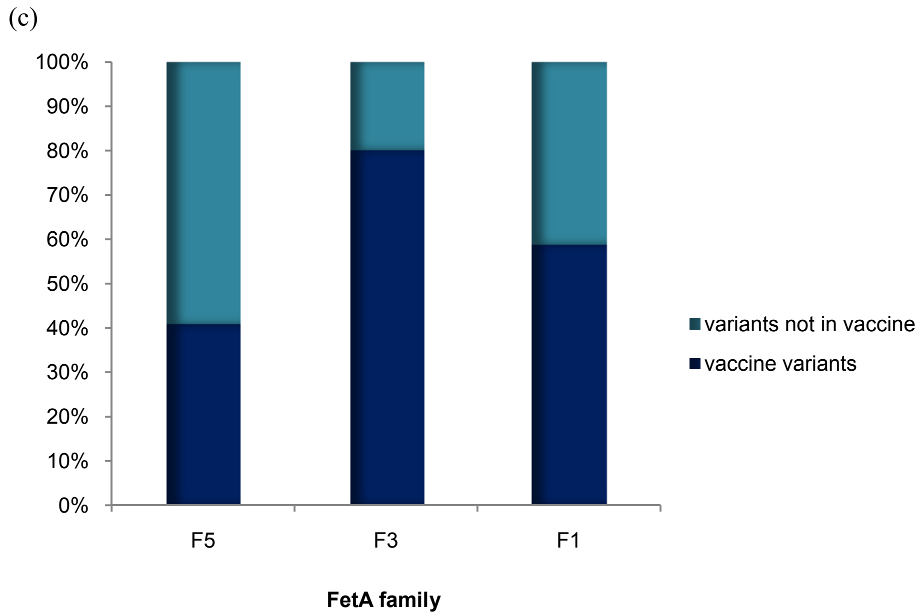


Figure 5.9: Coverage of (a) VR1 (b) VR2 and (c) FetA variant families in vaccine recipe.

5.3.3 PorA/FetA Recipe coverage

5.3.3.1 By year

There were no noticeable differences or changes in potential coverage and number of antigens covered of the vaccine recipe for each year over the three year period of the study (Table 5.5).

Table 5.5: Breakdown of coverage of PorA/FetA vaccine by year including number of antigens covered.

year	potential coverage	no. of antigens covered		
		1	2	3
2000	78.2% (769/983)	229	224	316
2001	79.6% (681/855)	215	224	242
2002	78.6% (766/975)	253	216	297
total	2216/2813 (78.8%)	697	664	855

5.3.3.2 By country

Overall there was between 58% and 94% potential coverage across Europe (Table 5.6). The country with the highest potential coverage in the EU dataset was Iceland with 94.1% coverage. It must be noted however that there was just one year's data from this country (2001). The country with the highest coverage that had all three years data was Italy with 89.1% coverage. The country with the lowest coverage was Greece with 57.8%. This is most likely due to the fact that the most common clonal complex in Greece is ST-162 (19%) which is rarely found elsewhere and is associated with the PorA type P1.22,14 and FetA type F5-9. Eight out of the 16 countries had at least 81% coverage. The country with the highest percentage of triple antigen coverage was Belgium with 59.6%.

Table 5.6: Breakdown of coverage of PorA/FetA vaccine by country in EY-MenNet database including number of antigens covered.

country	potential coverage	no. of antigens covered			
		1	2	3	total
Austria	76.1% (102/134)	26	41	35	102
Belgium	84.7% (183/216)	41	33	109	183
Czech Rep.	63.6% (63/99)	18	13	32	63
Denmark	85.3% (116/136)	24	32	60	116
England & Wales	82% (123/150)	48	26	49	123
Finland	64% (80/125)	32	28	20	80
France	76% (266/350)	88	114	64	266
Germany	75.6% (236/312)	59	58	119	236
Greece	57.8% (26/45)	8	7	11	26
Iceland	94.1% (16/17)	13	0	3	16
Ireland	84.1% (90/107)	40	10	40	90
Italy	89.1% (57/64)	18	21	18	57
Netherlands	86.9% (438/504)	157	140	141	438
Norway	80.9% (144/178)	49	40	55	144
Portugal	64.7% (22/34)	1	20	1	22
Spain	74.3% (254/342)	75	81	98	254
total	78.8% (2216/2813)	697	664	855	2216

5.3.3.3 By serogroup

The vaccine recipe gave potential coverage of 80% and 85.5% for the B and C serogroups respectively (Table 5.7). There was no serogroup A coverage but it occurs very rarely in Europe; there were just six isolates with this capsule in this dataset. The next lowest coverage was for the also relatively rare serogroup Y with 22.8%. Serogroup B had particularly high percentage of triple antigen coverage with 47% of isolates being covered at all three epitopes.

Table 5.7: Breakdown of coverage of PorA/FetA vaccine by serogroup in EU-MenNet database including number of antigens covered.

serogroup	potential coverage	no. of antigens covered			
		1	2	3	total
29E	75% (3/4)	0	3	0	3
A	0% (0/4)	0	0	0	0
B	80% (1308/1634)	390	300	618	1308
C	85.5% (590/690)	208	204	178	590
NG	77.8% (7/9)	3	1	3	7
W-135	68.4% (80/117)	5	73	2	80
X	100% (3/3)	2	1	0	3
Y	22.8% (13/57)	11	1	1	13
total	78.8% (2216/2813)	697	664	855	2216

5.3.3.4 Restricted to serogroup B

Although this vaccine recipe is designed to be non-serogroup specific, it is informative to look at the distribution of serogroup B-related disease types in Europe and potential vaccine coverage, since it is the predominant disease-associated serogroup in Europe and other developed regions and is the only major disease-associated serogroup without a comprehensive vaccine. The principal clonal complexes in Europe associated with serogroup B were ST-41/44 (38.6%), ST-32 (26.4%) and ST-269 (9.2%) complexes (Table 5.8). Together these three complexes accounted for 74.2% of isolates. The

major hyperinvasive serogroup B-associated lineages are found across Europe, if at varying frequencies (Figure 5.10).

Table 5.8: Breakdown of serogroup B by clonal complex.

clonal complex	frequency	percentage
ST-41/44 complex	904	38.62
ST-32 complex	618	26.40
ST-269 complex	215	9.18
Unassigned	203	8.67
ST-213 complex	55	2.35
ST-11 complex	52	2.22
ST-60 complex	44	1.88
ST-8 complex	42	1.79
ST-461 complex	38	1.62
ST-35 complex	37	1.58
ST-162 complex	34	1.45
ST-18 complex	31	1.32
ST-364 complex	12	0.51
ST-22 complex	10	0.43
ST-231 complex	9	0.38
ST-865 complex	8	0.34
ST-254 complex	6	0.26
ST-103 complex	5	0.21
ST-226 complex	3	0.13
ST-174 complex	2	0.09
ST-23 complex	2	0.09
ST-37 complex	2	0.09
ST-376 complex	2	0.09
ST-1157 complex	1	0.04
ST-167 complex	1	0.04
ST-175 complex	1	0.04
ST-334 complex	1	0.04
ST-53 complex	1	0.04
ST-549 complex	1	0.04
ST-750 complex	1	0.04
total	2341	

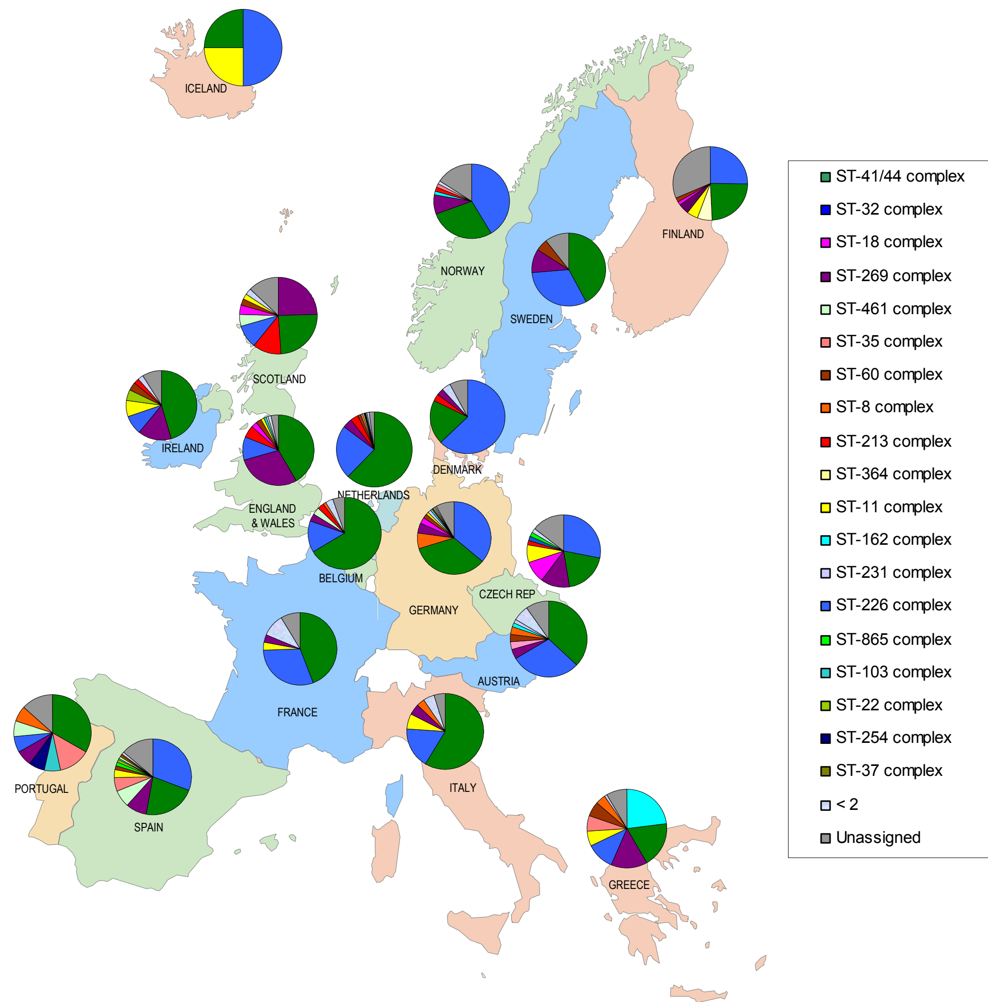


Figure 5.10: Distribution of clonal complexes in Europe restricted to serogroup B.

Restricting the dataset to serogroup B only gave slightly higher overall coverage of the recipe at 80% (Table 5.9). Again discounting Iceland and Portugal because they only have data from one year, the highest coverage was found in the Netherlands with 88.5% and the country with the lowest coverage was the Czech Republic with 59.4%. Nine out of the 16 countries had at least 83% coverage. The country with the highest percentage of triple antigen coverage was Belgium with 62.7%.

Table 5.9: Breakdown of coverage of PorA/FetA vaccine by country in EU-MenNet database restricted to serogroup B including number of antigens covered.

country	potential coverage	no. of antigens covered		
		1	2	3
Austria	62/86 (72.1%)	26	19	17
Belgium	102/117 (87.2%)	22	16	64
Czech Rep.	38/64 (59.4%)	14	6	18
Denmark	88/100 (88%)	19	18	51
England &Wales	90/104 (86.5%)	34	16	40
Finland	41/56 (73.2%)	12	18	11
France	45/54 (83.3%)	6	21	18
Germany	154/198 (77.8%)	42	24	88
Greece	24/37 (64.9%)	8	6	10
Iceland	4/4 (100%)	2	0	2
Ireland	65/77 (84.4%)	32	6	27
Italy	44/50 (88%)	12	17	15
Netherlands	285/322 (88.5%)	91	74	120
Norway	99/118 (83.9%)	35	21	43
Portugal	4/13 (30.8%)	1	2	1
Spain	163/234 (69.7%)	30	40	93
total	1308/1634 (80%)	386	304	618

5.3.3.5 By age group

Looking at potential coverage of the vaccine recipe by age group, there was at least 71% coverage in all age groups. The joint highest coverage was in the 5–9 and 10–14 years age group (87.2%). Coverage in the age groups with the highest risk of disease (<1 year, 1–4 years and later adolescents) was lower at 71%, 81% and 86% respectively but still high (Figure 5.11). The highest percentage of triple antigen coverage was in the 20–24 years age group (45.4%).

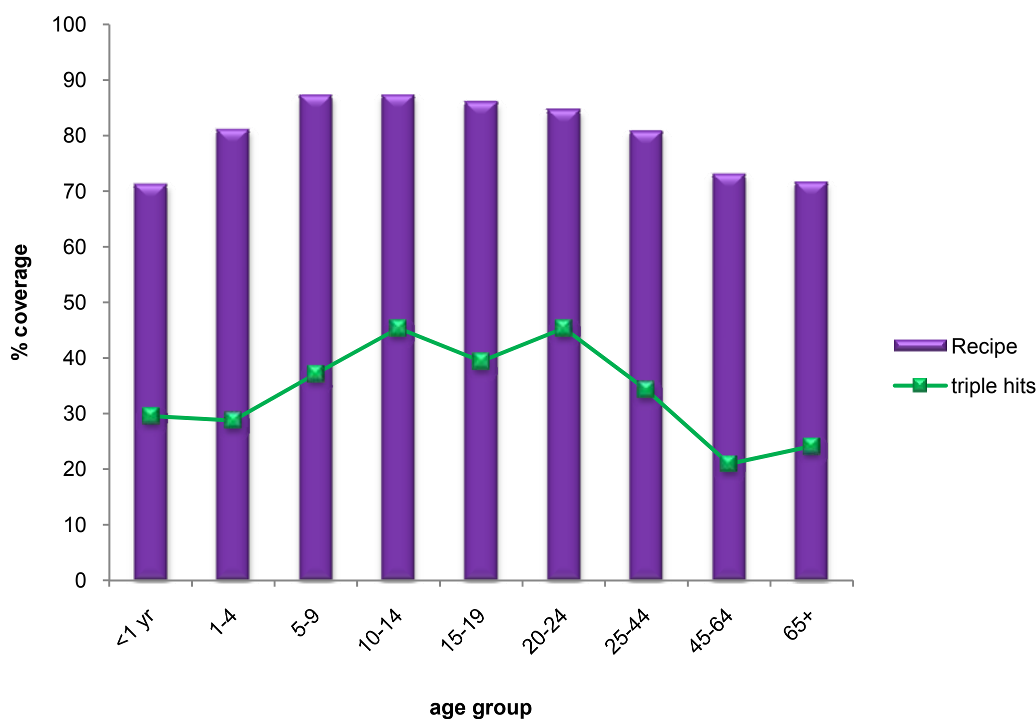


Figure 5.11: PorA/FetA recipe coverage by age group.

5.3.4 Comparison with other vaccines

The potential coverage offered by the vaccine recipe was compared to that of other currently available or in-development vaccines (Figure 5.12). These vaccines are: NonaMen, ACYW, Meningococcal C conjugate (MCC), New Zealand (P1.7-2,4), Norwegian (P1.7,16) and Cuban (P1.19,15). The putative PorA/FetA vaccine recipe compared well against all of these vaccines and presented the best coverage in the EU-MenNet dataset. The next most comprehensive coverage was provided by the NonaMen vaccine which contains nine PorA type antigens namely: P1.7,16; P1.5-1,2-2; P1.19,15-1; P1.5-2,10; P1.12-1,13; P1.7-2,4, P1.22,14; P1.7-1,1; P1.18-1,3,6. It gave a theoretical coverage of 72.2%. The ACYW vaccine gave a theoretical coverage of 34% which is unsurprising given that A, Y and W-135-related disease is still relatively uncommon in Europe at present. The MCC vaccine targets just serogroup C and therefore gave a potential coverage of 27.7%. As the New Zealand, Norwegian and Cuban vaccines were all developed in response to single clone epidemics and therefore target just that PorA type, their potential coverage is limited: 19.8%, 13.2% and 6.7%

respectively. Again however, it must be noted that cross-protection is not assumed and coverage may as a result be underestimated.

Comparing the PorA/FetA and NonaMen recipes, there was higher coverage overall for the PorA/FetA recipe but theoretical serogroup B coverage was higher for NonaMen at 86% *versus* 80% (Figure 5.12 inset). Potential coverage for serogroup C was lower for NonaMen than PorA/FetA (45% *versus* 85%).

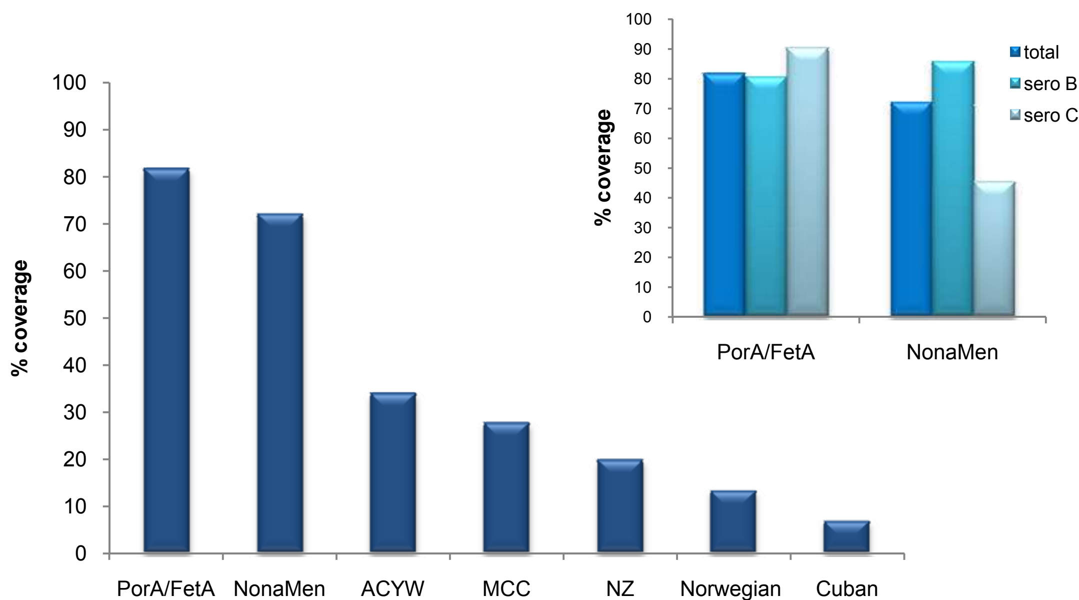


Figure 5.12: Potential coverage of PorA/FetA recipe compared to currently available vaccines NonaMen, ACYW, MCC, New Zealand, Norwegian and Cuban. Inset: comparison of coverage of major disease-associated serogroups B and C by PorA/FetA and NonaMen recipes.

The recipes were compared for potential coverage in countries in the EU-MenNet dataset with and without routine MCC vaccine programmes at the time of the study period: *with MCC*, England and Wales, Republic of Ireland and Spain; *without MCC*, Austria, Czech Republic, Denmark, Finland, France, Germany, Greece, Italy, Norway, Portugal. The Netherlands, Belgium and Iceland introduced MCC routine vaccination in 2002, and in Greece from 2001 most paediatricians have administered it as a single dose to patients aged 12 months or more. Therefore, these countries were excluded from the comparison. The PorA/FetA recipe gave higher potential coverage

than NonaMen for both countries that have implemented routine MCC vaccination and those that have not (Figure 5.13). Higher coverage was found for both in countries with MCC vaccination (80% and 74% for PorA/FetA and NonaMen respectively). Theoretical coverage in non-MCC countries was 75% and 72% for PorA/FetA and NonaMen recipes respectively.

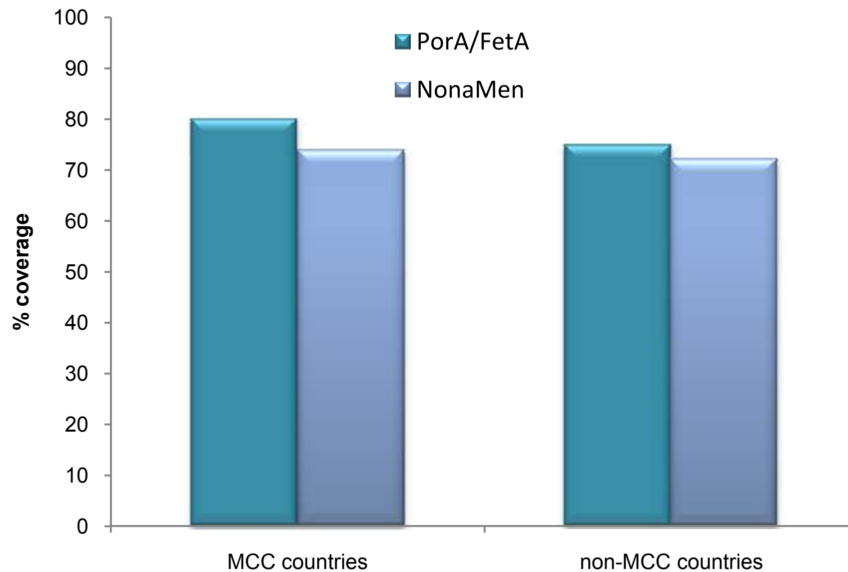


Figure 5.13: Potential coverage of PorA/FetA recipe compared to NonaMen recipe in countries which have implemented MCC *versus* those that have not.

5.4 Discussion

Molecular epidemiology has become an integral part of population biology and disease surveillance of many infectious diseases in recent times. A more recent application and extension of this utility has been to measure the impact of vaccine interventions (Maiden *et al.*, 2008) and the design of vaccines based on the most prevalent disease-associated types (van den Dobbelsteen *et al.* , 2004). It is necessary to know the distribution of antigens in the population before a vaccine is formulated to in order to arrive at the best possible coverage. Continued surveillance is also key in monitoring

disease-associated strains circulating in the population, assessing vaccine effectiveness, detecting any emergent escape variants or novel variants arising due to the selection pressure imposed by the vaccine and any subsequent shift in the pathogen population structure which may necessitate re-formulation of the vaccine. This comprehensive and relatively contemporary survey of meningococcal disease in Europe was the basis from which to formulate a vaccine based on the two immunogenic vaccine candidates PorA and FetA. The rationale for use of these particular antigens in a vaccine is their immunogenicity, persistent structured diversity in terms of non-overlapping strain structuring and association with hyperinvasive lineages. Previous work on disease datasets from England and Wales and a global collection demonstrated that a relatively simple formula containing a small number of the most prevalent PorA and FetA antigens (Russell *et al.*, 2008; Urwin *et al.*, 2004) could protect against the majority of disease strains assuming an adequate immune response.

In the European disease dataset, the antigen distribution was found to be structured with non-overlapping antigen combinations and associations with clonal complex. Antigenic diversity implies a degree of host immune response against them which is a prerequisite for a vaccine candidate. Indeed, non-overlapping strain structuring is a feature induced by an immune response directed against a dominant antigenic determinant where there is moderate to strong cross-protection among types that share the same alleles (Gupta *et al.*, 1996). This leads to a meningococcal population which consists of a stable set of independently transmitted strains that do not share polymorphic determinants. Identification of this signature of antigen immunogenicity has been put forward as a strategy for identifying dominant epitopes (Gupta *et al.*, 1996). Antigen relationships with clonal lineage are striking, with the majority of clonal complexes being made up of just one antigen type, *e.g.* 94% of P1.7-2,4 were ST-41/44 complex. The relationships are not absolute, however, and as this study only covers a relatively short time period, any major shift in association may not be detected. Antigen clonal complex associations can change over time even in the absence of any interventions (Buckee *et al.*, 2008; Russell *et al.*, 2008). Models of the effect of immune selection on the meningococcal population structure show this to be consistent with dominant lineages being out-competed and replaced by others over time (Buckee *et al.*, 2008). So otherwise unrelated strains can have the same PorA/FetA combinations.

PorA and FetA are two well characterised, immunogenic and protective antigens (Pettersson *et al.*, 1990; Wedege *et al.*, 1998). They show similar diversity and structuring in the meningococcal population and were thus proposed as antigens in a new formulation. Choosing two different types of antigens produced by genes that are distant from each other on the genome reduces the chances of escape variants as the organism would have to change at two loci to completely escape immune attack. PorA also has two epitopes in separate loop regions so this makes a potential number of three epitope hits per organism. The antigen components chosen for the recipe were the most prevalent in the European disease population and encompass the major hyperinvasive lineages. Many of these antigens have been found associated with disease over wide time periods and distances (Russell *et al.*, 2008). This supports the prospective longevity of vaccine coverage. A previous study of a temporally and geographically diverse set of 78 disease isolates found that a combination of six PorA and five FetA variants gave homologous theoretical protection for all isolates (Urwin *et al.*, 2004). Another study of 323 disease isolates collected over a period of 20 years in England and Wales found that combinations of the four most prevalent PorA and FetA variants could potentially offer 80% coverage (Russell *et al.*, 2008). Coverage in the England and Wales collection remained around the 80% mark across the 20 year span of the study illustrating the stability of antigen associations and their persistence. In this work, a vaccine cocktail containing four each of PorA and FetA antigens gave a potential coverage of almost 79%. In fact the addition of just a further four PorA and FetA antigens would give a theoretical coverage of 96.1%. The potential coverage of the four component recipe compares well with any of the currently, or soon to be, available vaccines such as NonaMen. Since the full extent of cross-protection is still relatively unknown, coverage may be higher still. So far, by the standard correlate of protection the SBA, cross-protection amongst variant members of PorA families has been shown to be quite variable (Findlow *et al.*, 2005; Martin *et al.*, 2000b; Vermont *et al.*, 2003). If protection amongst members of the same antigen family is assumed, then potential coverage could be as high as 98%. It is therefore important to establish the amount of cross-protection for recipe variants in order to know the full potential of any cocktail vaccine.

Many meningococcal vaccines being developed at the moment focus solely on targeting serogroup B disease. One of the attributes of a vaccine based on sub-capsular antigens is that it can prospectively provide protection across all serogroups. The

antigens in this formulation are associated with the major disease-associated serogroups in Europe (B and C) but also serogroups more associated with disease in other regions such as Y and W-135. There is no reason to believe that these serogroups may not increase in prevalence over time. Serogroup Y for example, while still relatively uncommon in Europe, is one of the major disease-associated serogroups in the United States. It was uncommon prior to the mid-1990s and now accounts for about a third of all disease cases (Rosenstein *et al.*, 1999). Serogroup W-135 was introduced into many countries after the Hajj pilgrimage in 2000 and 2001 although the number of cases was relatively low (Taha *et al.* , 2000). Serogroup B is currently still the predominant capsule associated with meningococcal disease in Europe and in this collection accounted for 65% of disease isolates. While this recipe is non-serogroup specific, potential serogroup B coverage was 80%. The next most dominant serogroup, serogroup C gave a potential coverage of 85%.

Conjugate vaccination strategies have been shown to affect prevalence of disease, carriage and also the population biology of the meningococcus dramatically. The MCC campaign in the UK introduced in 1999 significantly reduced disease attributable to serogroup C by 86.7% in the targeted age groups by 2001 (Balmer *et al.*, 2002). Direct protection from the vaccine itself was shown by high effectiveness in all vaccinated age groups of >87% more than one year after immunisation (Miller *et al.* , 2001). Effectiveness decreased over time in infants given the routine 2, 3 and 4 month schedule (Trotter *et al.* , 2004). Indirect protection in the form of herd immunity was assumed to be an important contributor to reduction of disease as (i) there was a significant reduction of 67% in the attack rate in the unimmunised population before *versus* after the introduction of the vaccine (Ramsay *et al.* , 2003) and (ii) despite the reduction in direct protection in some of the immunised cohort, there was no evidence of increased serogroup C disease. Changes in the carriage of the meningococcus also point to the effect of herd immunity. There were significant reductions in carriage of meningococci expressing serogroup C and also having the serogroup C-specific allele ($siaD_C$) (Maiden *et al.*, 2008). MCC vaccination campaigns have been implemented in other European countries: Republic of Ireland, Spain, the Netherlands, Iceland, Belgium, Portugal, Germany, Switzerland, and Greece (Trotter & Ramsay, 2007) (www.euvac.net). There has been no evidence to date of capsule replacement or capsule switching by vaccine escape variants filling the niche left by serogroup C organisms as a result of the intervention (Alonso *et al.* , 2007; Trotter *et al.* , 2006b).

This is not completely out of the question however and examples of this phenomenon have occurred in other organisms. For example, following the introduction of a seven valent pneumococcal vaccine in the United States, disease associated with non-vaccine antigens have increased in incidence in the population (Brueggemann *et al.*, 2007).

In the case of meningococcal OMPs, there are several mechanisms by which the organism may evade the host immune response that may have significance for non-capsular vaccine interventions long term. These are: mutation, recombination, gene deletion, variable expression and insertional inactivation. A consequence of antigen changes is that a population may be immunologically susceptible to these new antigen types resulting in a rise in disease. Recombination and mutation are means by which the meningococcus creates variation to evade host immunity by antigenic shift and drift respectively. Antigenic shift has been implicated in increased disease in the United States in the mid-1990s where an increase in serogroup Y ST-23 complex-related disease was associated with antigenic shift by horizontal gene transfer at three OMP loci – PorA, PorB and FetA (Harrison *et al.* , 2006). At the same time an increase in serogroup C disease was associated with a change in the FetA VR. As observed by Harrison *et al*, the shift in the antigenic repertoire for the serogroup Y-related strains is consistent with strong immune selection resulting in a new repertoire with non-overlapping alleles with the original (Gupta *et al.* , 1996). Point mutation, creating single amino acid changes in PorA VR2, have been associated with a geographically-restricted outbreak in France and also an increase in endemic disease in England and Wales (McGuinness *et al.*, 1991; Taha *et al.*, 2001). The finding of disease-associated isolates with deletions of the vaccine candidate genes (Claus *et al.*, 2007; Marsh *et al.*, 2007; van der Ende *et al.*, 2003) may be of concern although numbers found so far are quite small and so far none have been found with both *fetA* and *porA* genes simultaneously deleted. Meningococci deficient in the *porA* gene were found in several strains associated with an outbreak in the Netherlands in 2001 (van der Ende *et al.* , 2003). Deletions of the *fetA* gene have also been found in clinical isolates though only rarely (Claus *et al.*, 2007; Marsh *et al.*, 2007). There are varied mechanisms by which antigens may vary expression. In the case of PorA, variation in the length of the homopolymeric tract of guanidine and/or thymidine residues in the promoter region, along with the homopolymeric tract of adenine residues in the coding region of the *porA* gene can result in variable expression (van der Ende *et al.*, 1995; van der Ende *et al.*, 2000). Additionally, lack of PorA expression can be caused by point mutation or

insertional inactivation (Jelfs *et al.*, 2000b; Newcombe *et al.*, 1998; van der Ende *et al.*, 2000). Vaccine pressure may induce escape variants with altered protein repertoires or with a lack of protein expression through various means including gene deletion. The proposed vaccine formula with a cocktail of proteins from two different antigens may be the best means of avoiding these potential issues.

Reduction in carriage and the potential additional benefits of herd immunity have not been demonstrated as yet for protein vaccines (Dellicour & Greenwood, 2007). A study in Norway showed no difference in carriage in those immunised with an OMV vaccine and control groups and also no extra indirect protection for those unvaccinated in schools which received the intervention *versus* students in schools which received a placebo (Bjune, 1992). Although the mechanisms by which herd immunity operate are still not well understood, particularly with the new generation of serogroup B substitute vaccines, monitoring its effects are important in the design and implementation of vaccine interventions. In the case of MCC, it was shown that immunisation of teenagers, who are the most prevalent carriers and thus where most transmission occurs, was an important element of the herd immunity effect in the UK intervention. The long-term effects of protein vaccines on meningococcal population biology are relatively unknown and it will be necessary to monitor this closely and separate the impact of vaccine interventions from natural fluctuations in the microbial population. There is a concern that vaccine interventions may eliminate from carriage, strains that are non-invasive but provide natural immunity. This is thought not to have been important in the case of MCC as carriage of serogroup C organisms was relatively low already. However, as cocktail vaccines such as the PorA/FetA recipe with multiple antigens span the serogroups, the effect on the population of both disease and carriage-related strains may be quite different.

Meningococcal serogroups and clonal complexes are differentially distributed amongst the age groups in terms of disease association. Serogroup B was proportionally more associated with the younger, more at-risk, age groups, the under fours, while serogroup C, which is associated with a higher case-fatality ratio (EU-IBIS, 2007; Yazdankhah *et al*., 2004), was more associated with those above nine years of age. The non-BC serogroups were associated with older age groups particularly those above 65 years of age. In tandem, the serogroup B-associated ST-41/44 complex along with unassigned STs were significantly associated with the under ones, while the hypervirulent ST-11 complex was associated with the adolescent age group. These

differences are likely the result of waning maternal immunity in under ones leaving them vulnerable to less virulent and more carriage-associated clonal complexes. Adolescents, on the other hand who have the highest rates of carriage (Claus *et al.* , 2005) and are therefore less immunologically naïve may be more likely stricken by more virulent strains. There was at least 71% potential coverage of the recipe in all age groups. There were no major differences in the coverage amongst the age groups as the recipe contains antigens associated with the major hyperinvasive lineages. Targeting of particular age groups for immunisation is an important consideration in the planning of a vaccine intervention especially as the impact of OMV vaccines on the meningococcal population is still not fully understood. The effect on carriage and thus herd immunity can have important implications for overall disease dynamics. The experience of the MCC in the UK showed how the inclusion of older age groups, in that case up to age 19, in the catch-up campaign had an important effect of additional protection for the population by the reduction of meningococcal transmission including virulent types.

This work has demonstrated the use of molecular epidemiology in the design of a relatively simple formulation that could offer protection against a large section of currently circulating meningococcal strains. Continued and detailed surveillance of meningococcal strains is required to assess the effect of a vaccine on disease and carriage, to monitor the effectiveness of the vaccine, to detect changes in the microbial population whether through the emergence of new genotype/antigen combinations and/or diversification of current ones. It will be important to measure the direct effect of the vaccine and how it bears upon the meningococcal population and separate this from any natural fluctuations that may have occurred. Individual countries are moving towards more detailed molecular epidemiology and real time databases such as EMERT (http://emgm.eu/emert/), which uses the new meningococcal typing scheme to index variation, have been established to aid surveillance across Europe. Additional future work which will shed more light on the potential coverage of the vaccine include antigen expression studies and serum bactericidal assays measuring the protection afforded by such a formula.

CHAPTER 6: Variation of the factor H-binding protein

6.1 Introduction

Factor H-binding protein (fHbp), previously known as lipoprotein 2086 (LP2086) and Genome Derived Neisserial Antigen 1870 (GNA1870), (Fletcher *et al.*, 2004; Masignani *et al.*, 2003) is a surface-exposed 29 kDa globular lipoprotein that has a critical role in modulating the activity of the alternative complement pathway where it binds the regulatory protein factor H (fH) (Madico *et al.*, 2006). fH maintains homeostasis of the complement system and also, by attachment to host cells and tissue, prevents potential damage to them by inhibiting complement activation (Rodríguez de Córdoba *et al.*, 2004). fHbp is present in all meningococci although levels of expression may vary in different isolates (Fletcher *et al.*, 2004; Masignani *et al.*, 2003). In comparison to other vaccine antigens, it is relatively sparse in its epitope surface-exposure in most meningococcal strains (Welsch *et al.*, 2004). Expression of fHbp has been found to be key for survival in *ex vivo* human blood and human serum particularly in high-expressing strains (Seib *et al.*, 2009; Welsch *et al.*, 2008).

fHbp is a principal component of two recombinant protein vaccines in clinical trials at the time of writing (Anderson *et al.*, 2009; Biolchi *et al.*, 2009). It is unique as a vaccine candidate in that it is able to elicit serum antibodies that activate classical complement pathway bacteriolysis and also prevent fH binding to the meningococcal cell surface thus making it more susceptible to bactericidal activity (Madico *et al.*, 2006; Welsch *et a l.*, 2008). It has been shown to be relatively diverse antigenically and is divided into two subfamilies (subfamily A and B) or three variants (variants 1, 2 and 3) according to nomenclature system (Fletcher *et al.*, 2004; Masignani *et al.*, 2003). Like other vaccine candidates, molecular epidemiology should play an important part in vaccine design and implementation. It is essential to determine the diversity of the protein in the population to know the number of variants required to provide the broadest possible coverage.

In this Chapter, variation of fHbp in a reference set of diverse meningococcal isolates was surveyed. The association of particular variants with clonal complex and serogroup was established, and the levels of recombination and selection acting on it

were determined. Also, a novel unified nomenclature scheme was developed that was independent of subfamily/variant and a Web-accessible database established to facilitate querying of sequences and submission of new allele sequences.

6.2 Results

6.2.1 Diversity of fHbp gene and protein

The *fHbp* gene was found in all 107 isolates and among these a total of 28 unique gene sequences encoding 27 different amino acid sequences were identified (Table 6.1). Unique peptide and nucleotide sequences were arbitrarily assigned allele numbers in order of discovery and entered into a database (http://neisseria.org/nm/typing/fhbp/), providing a comprehensive repository of reported fHbp diversity.

On the basis of 798 unambiguously aligned nucleotides there were a total of 299 variable nucleotide sites (Figure 6.1). Two broad groups of sequence variant types were evident from this alignment. This corresponded to the previously identified groups: subfamily A/variant 2 and subfamily B/variant 1 (Fletcher *et al*., 2004). Within subfamily A/variant 2, there were five putative subfamily A/variant 3 sequences which differ from the rest of subfamily A/variant 2 mainly in the N-terminal first ~100 amino acids (a.a.) (Masignani *et al.*, 2003). The subfamily A/variant 2 (including variant 3) was significantly more diverse in terms of allele types than subfamily B/variant 1 (*D*=0.91 (95% CIs 0.87–0.95) *vs.* 0.80 (95% CIs 0.75–0.85) respectively). F_{ST} analysis of nucleotide sequences of the two subfamilies/variants gave a value of 0.873 ($p<0.005$), indicating a high level of genetic differentiation between them.

Table 6.1: Study isolates and association with clonal complex, sequence type (ST), year, country, disease, serogroup, peptide, nucleotide allele and subfamily/variant.

isolate name	complex	ST	subfamily/variant	year	country	disease	serogroup	peptide	allele
393	ST-1 complex	1	SubB/v1	1968	Greece	carrier	A	4	4
20	ST-1 complex	1	SubB/v1	1963	Niger	invasive (unspecified/other)	A	4	4
129E	ST-1 complex	1	SubB/v1	1964	West Germany	invasive (unspecified/other)	A	4	4
254	ST-1 complex	1	SubB/v1	1966	Djibouti	invasive (unspecified/other)	A	4	4
106	ST-1 complex	1	SubB/v1	1967	Morocco	invasive (unspecified/other)	A	4	4
6748	ST-1 complex	1	SubB/v1	1971	Canada	invasive (unspecified/other)	A	4	4
S5611	ST-1 complex	1	SubB/v1	1977	Australia	invasive (unspecified/other)	A	4	4
79126	ST-1 complex	3	SubB/v1	1979	China	invasive (unspecified/other)	A	38	37
79128	ST-1 complex	3	SubB/v1	1979	China	invasive (unspecified/other)	A	38	37
371	ST-1 complex	1	SubB/v1	1980	India	invasive (unspecified/other)	A	4	4
322/85	ST-1 complex	2	SubB/v1	1985	East Germany	invasive (unspecified/other)	A	14	14
120M	ST-1 complex	1	SubB/v1	1967	Pakistan	meningitis and septicaemia	A	4	4
139M	ST-1 complex	1	SubB/v1	1968	Philippines	unspecified	A	1	1
BZ 133	ST-1 complex	1	SubB/v1	1977	Netherlands	invasive (unspecified/other)	B	4	4
890326	ST-103 complex	28	SubA/v2	1989	Netherlands	invasive (unspecified/other)	Z	25	25
NG P20	ST-11 complex	11	SubB/v1	1969	Norway	invasive (unspecified/other)	B	5	5
D1	ST-11 complex	11	SubA/v2	1989	Mali	carrier	C	22	22
M597	ST-11 complex	11	SubA/v2	1988	Israel	invasive (unspecified/other)	C	22	22
BRAZ10	ST-11 complex	11	SubA/v2	1976	Brazil	unspecified	C	22	22
F1576	ST-11 complex	11	SubA/v2	1984	Ghana	unspecified	C	22	22
500	ST-11 complex	11	SubA/v2	1984	Italy	unspecified	C	22	22
MA-5756	ST-11 complex	11	SubA/v3	1985	Spain	unspecified	C	67	54
90/18311	ST-11 complex	11	SubA/v2	1990	Scotland	unspecified	C	22	22
L93/4286	ST-11 complex	11	SubA/v2	1993	England	invasive (unspecified/other)	C	95	92
38VI	ST-11 complex	11	SubA/v3	1964	USA	unspecified	B	98	95
860800	ST-167 complex	29	SubA/v2	1986	Netherlands	invasive (unspecified/other)	Y	23	23
EG 328	ST-18 complex	18	SubB/v1	1985	East Germany	invasive (unspecified/other)	B	37	36
EG 327	ST-18 complex	19	SubA/v2	1985	East Germany	invasive (unspecified/other)	B	33	41
1000	ST-18 complex	20	SubB/v1	1988	USSR	invasive (unspecified/other)	B	5	5
528	ST-18 complex	18	SubB/v1	1989	USSR	invasive (unspecified/other)	B	37	36
E26	ST-198 complex	39	SubA/v3	1988	Norway	carrier	X	94	91
A22	ST-22 complex	22	SubA/v2	1986	Norway	carrier	W-135	16	16
71/94	ST-23 complex	23	SubA/v2	1994	Norway	invasive (unspecified/other)	Y	25	25
297-0	ST-254 complex	49	SubB/v1	1987	Chile	carrier	B	13	13
NG F26	ST-269 complex	14	SubA/v2	1988	Norway	carrier	B	19	19
NG 6/88	ST-269 complex	13	SubA/v2	1988	Norway	invasive (unspecified/other)	B	32	40
44/76	ST-32 complex	32	SubB/v1	1976	Norway	invasive (unspecified/other)	B	1	1
NG 080	ST-32 complex	32	SubB/v1	1981	Norway	invasive (unspecified/other)	B	1	1
NG144/82	ST-32 complex	32	SubB/v1	1982	Norway	invasive (unspecified/other)	B	1	1
BZ 83	ST-32 complex	34	SubB/v1	1984	Netherlands	invasive (unspecified/other)	B	1	1
EG 329	ST-32 complex	32	SubB/v1	1985	East Germany	invasive (unspecified/other)	B	1	1
BZ 169	ST-32 complex	32	SubB/v1	1985	Netherlands	invasive (unspecified/other)	B	1	1
NG PB24	ST-32 complex	32	SubB/v1	1985	Norway	invasive (unspecified/other)	B	1	1
8680	ST-32 complex	32	SubB/v1	1987	Chile	invasive (unspecified/other)	B	13	13
204/92	ST-32 complex	33	SubB/v1	1992	Cuba	invasive (unspecified/other)	B	1	1
196/87	ST-32 complex	32	SubB/v1	1987	Norway	unspecified	C	1	1
E32	ST-334 complex	31	SubA/v2	1988	Norway	carrier	Z	19	19
SWZ107	ST-35 complex	35	SubA/v2	1986	Switzerland	invasive (unspecified/other)	B	16	16
NG E31	ST-364 complex	15	SubA/v2	1988	Norway	carrier	B	34	43

DK 353	ST-37 complex	37	SubA/v2	1962	Denmark	invasive (unspecified/other)	B	24	24
BZ 232	ST-37 complex	38	SubA/v2	1964	Netherlands	invasive (unspecified/other)	B	24	24
A4/M1027	ST-4 complex	4	SubB/v1	1937	USA	invasive (unspecified/other)	A	5	5
10	ST-4 complex	4	SubB/v1	1963	Burkina Faso	invasive (unspecified/other)	A	5	5
26	ST-4 complex	4	SubB/v1	1963	Niger	invasive (unspecified/other)	A	5	5
255	ST-4 complex	4	SubB/v1	1966	Burkina Faso	invasive (unspecified/other)	A	5	5
S3131	ST-4 complex	4	SubB/v1	1973	Ghana	invasive (unspecified/other)	A	5	5
690	ST-4 complex	4	SubB/v1	1980	India	invasive (unspecified/other)	A	5	5
C751	ST-4 complex	4	SubB/v1	1983	Gambia	invasive (unspecified/other)	A	5	5
1014	ST-4 complex	4	SubB/v1	1985	Sudan	invasive (unspecified/other)	A	5	5
2059001	ST-4 complex	4	SubB/v1	1990	Mali	invasive (unspecified/other)	A	5	5
D8	ST-4 complex	4	SubB/v1	1990	Mali	invasive (unspecified/other)	A	5	5
243	ST-4 complex	4	SubB/v1	1966	Cameroon	unspecified	A	5	5
NG H36	ST-41/44 complex	47	SubA/v2	1988	Norway	carrier	B	19	19
NG E30	ST-41/44 complex	44	SubA/v2	1988	Norway	carrier	B	32	42
BZ 147	ST-41/44 complex	48	SubB/v1	1963	Netherlands	invasive (unspecified/other)	B	4	4
BZ198	ST-41/44 complex	41	SubB/v1	1986	Netherlands	invasive (unspecified/other)	B	5	5
88/03415	ST-41/44 complex	46	SubB/v1	1988	Scotland	invasive (unspecified/other)	B	14	14
91/40	ST-41/44 complex	42	SubB/v1	1991	New Zealand	invasive (unspecified/other)	B	14	14
400	ST-41/44 complex	40	SubB/v1	1991	Austria	invasive (unspecified/other)	B	14	14
AK50	ST-41/44 complex	41	SubB/v1	1992	Greece	invasive (unspecified/other)	B	4	4
M-101/93	ST-41/44 complex	41	SubB/v1	1993	Iceland	invasive (unspecified/other)	B	4	4
931905	ST-41/44 complex	41	SubB/v1	1993	Netherlands	invasive (unspecified/other)	B	14	14
50/94	ST-41/44 complex	45	SubB/v1	1994	Norway	invasive (unspecified/other)	B	4	4
M40/94	ST-41/44 complex	41	SubB/v1	1994	Chile	invasive (unspecified/other)	B	35	32
N45/96	ST-41/44 complex	41	SubB/v1	1996	Norway	invasive (unspecified/other)	B	1	1
NG H15	ST-41/44 complex	43	SubA/v2	1988	Norway	carrier	B	19	19
80049	ST-5 complex	5	SubB/v1	1963	China	carrier	A	39	38
F4698	ST-5 complex	5	SubB/v1	1987	Saudi	carrier	A	14	14
153	ST-5 complex	5	SubA/v2	1966	China	invasive (unspecified/other)	A	22	22
154	ST-5 complex	6	SubB/v1	1966	China	invasive (unspecified/other)	A	36	35
S4355	ST-5 complex	5	SubB/v1	1974	Denmark	invasive (unspecified/other)	A	5	5
7891	ST-5 complex	5	SubB/v1	1975	Finland	invasive (unspecified/other)	A	5	5
11-004	ST-5 complex	5	SubB/v1	1984	China	invasive (unspecified/other)	A	5	5
H1964	ST-5 complex	5	SubB/v1	1987	UK	invasive (unspecified/other)	A	5	5
92001	ST-5 complex	7	SubB/v1	1992	China	invasive (unspecified/other)	A	5	5
14/1455	ST-5 complex	5	SubB/v1	1970	USSR	unspecified	A	14	14
IAL2229	ST-5 complex	5	SubB/v1	1976	Brazil	unspecified	A	5	5
F6124	ST-5 complex	5	SubA/v2	1988	Chad	invasive (unspecified/other)	A	25	25
860060	ST-750 complex	24	SubA/v3	1986	Netherlands	invasive (unspecified/other)	X	84	69
BZ 10	ST-8 complex	8	SubA/v2	1967	Netherlands	invasive (unspecified/other)	B	18	39
B6116/77	ST-8 complex	10	SubA/v2	1977	Iceland	invasive (unspecified/other)	B	16	16
BZ 163	ST-8 complex	9	SubA/v2	1979	Netherlands	invasive (unspecified/other)	B	16	16
G2136	ST-8 complex	8	SubA/v2	1986	England	invasive (unspecified/other)	B	16	16
AK22	ST-8 complex	8	SubA/v2	1992	Greece	invasive (unspecified/other)	B	16	16
SB25	ST-8 complex	8	SubA/v2	1990	South Africa	invasive (unspecified/other)	C	16	16
94/155	ST-8 complex	66	SubA/v2	1994	New Zealand	invasive (unspecified/other)	C	16	16
312 901	ST-8 complex	8	SubA/v2	1996	England	invasive (unspecified/other)	C	16	16
CN100	Unassigned	21	SubB/v1	1941	England	invasive (unspecified/other)	A	4	4
NG E28	Unassigned	26	SubB/v1	1988	Norway	carrier	B	14	14
3906	Unassigned	17	SubA/v2	1977	China	invasive (unspecified/other)	B	18	39
NG 4/88	Unassigned	30	SubB/v1	1988	Norway	invasive (unspecified/other)	B	4	4
NG 3/88	Unassigned	12	SubA/v3	1988	Norway	invasive (unspecified/other)	B	96	93
NG G40	Unassigned	25	SubA/v2	1988	Norway	carrier	B	24	24

EG 011	Unassigned	36	SubA/v2	1986	East Germany	invasive (unspecified/other)	B	24	24
NG H41	Unassigned	27	SubA/v2	1988	Norway	carrier	B	25	25
DK 24	Unassigned	16	SubA/v2	1940	Denmark	invasive (unspecified/other)	B	97	94
NG H38	Unassigned	36	SubA/v2	1988	Norway	carrier	B	24	24

There was 63% nucleotide sequence identity shared between the two main groups and greater identity within them: 85% nucleotide site identity within subfamily A/variant 2 and 87% nucleotide site identity within subfamily B/variant 1. The overall mean Kimura 2-parameter *p*-distance among all gene sequences was 0.165. The within-group mean *p*-distances were 0.046 for subfamily A/variant 2 and 0.032 for subfamily B/variant 1 with a mean average *p*-distance between the two subfamilies/variants of 0.302. There was 56% deduced amino acid sequence identity shared between the two groups. Within subfamily A/variant 2 there was 81% amino acid site identity and within subfamily B/variant 1 there was 87% amino acid site identity. The overall mean *p*-distance among the amino acid sequences was 0.17. The within-group mean *p*-distances for subfamily A/variant 2 and subfamily B/variant 1 were 0.052 and 0.038 respectively. The mean average *p*-distance between the two subfamilies/variants was 0.31. Without the subfamily A/variant 3 sequences, subfamily A/variant 2 nucleotide sequences identity was 90% and amino acid site identity was 88%.

Sequence variability was found throughout the gene and encoded protein (Figure 6.1). There was a marked difference in variability, however, between the N–terminal first ~105 a.a. and the C-terminal region of ~161 a.a. where sequences were more variable. Amino acid sequence identity of the C-terminal region between the two groups was 48%, however, there was more identity within the groups (subfamily A/variant 2, 87%; subfamily B/variant 1, 84%). For the N-terminal region, there was 67% amino acid identity between the two subfamilies/variants and 70% within subfamily A/variant 2 and 93% within subfamily B/variant 1. In the absence of the subfamily A/variant 3 sequences, for subfamily A/variant 2, there was 90% amino acid identity in the C-terminal region and 86% in the N-terminal region.

```
                                           11111111111111111111111111111111111111111111111111222222222222222222222222222222222222
           1111222233444444455566668888899900111112222222223333333344444555555566677777888899000000111122222233333444555555555556
          503490347570235679024023413789356060457912345678901234689012380145679136013571268470345680279013459134782790123456891
fHbp1     -GVAAGDAAGQTDQVRKEKKAKYGNVRIRQDQLEVSLTFTQIQDSEHSGKMVAKQRIGDIASDKEGGRTRTGAGKTTAANKSLDAADIPGRSSVLNQASSGIGKVEKTVNGIRHGLA
fHbp4     -........S............................L..V...............................S.....H.....S...K.............Q..E.A........
fHbp5     -....V...S............................L..V...............................S.....H.....S...K.............Q..E.A........
fHbp13    -..............................K......L..V....D.................K..S...........H...ET.Y..E.......D.....Q..E.A...H....
fHbp14    -........S............................L..E..P.........RK........KDVM...........H...ET.Y..EH......D.....Q..E.A...H....
fHbp35    -.....................................L..E..L.........RK........KDVM...........H....V.Y..EH......D....E.....A...H....
fHbp36    -....V...S....................N.......L..V...............................S.....H.....S...K.............Q..E.A........
fHbp37    -..........M.............I............L..V....D.R................SD....S.......H.........E.......D.....Q.............
fHbp38    -......V..............................L..E..L...R.....RK........KDVM...........H....V.Y..EH......D....E.....A...H....
fHbp39    -.....................................L..E..L...R.....RK........KDVM...........H....V.Y..EH......D....E.....A...H....
fHbp16    -........S........................IDVVLIK.NNPDKIDSLINQSLVSGLGANQ-D.KEHKS.......H.TQE..ELAESLDTRGSETHALDRIT.IGEKVHE.IG
fHbp18    -........S...............I.....R..IDVVLIK.NNPDKIDSLINQSLVSGLGANQ-D.KEHKS.......H.TQE..ELAESLDTRGSETHALDRIT.IREKVHE.IG
fHbp19    -........S........................IDVVLIK.NNPDKIDSLINQSLVSGLGANQ-S.KEHKS.......H.TQES.ELAESLDTRGGETHALDRIT.IREKVHE.IG
fHbp22    -........S........................IDVVLIK.NNPDKIDSLINQSLVSGLGANQ-S.KEHKSPNRHSTKYRTQES.ELAESLDTRGGETHALDRIT.IREKVHE.IG
fHbp23    -.................................IDVVLIK.NNPDKIDSLINQSLVSGLGANQ-S.KEHKSPNRHSTKYRTQES.ELAESLDTRGGETHALDRIT.IREKVHE.IG
fHbp24    -........S......................TAINVVLIK.NNPDKIDSLINQSLVSGLGANQ-D.KEHKSPNRHSTKYRTQES.ELAESLDTRGGETHALDRIT.IREKVHE.IG
fHbp25    -.......TS......................TAINVVLIK.NNPDKIDSLINQSLVSGLGANQ-D.KEHKSPNRHSTKYRTQES.ELAESLDTRGGETHALDRIT.IREKVHE.IG
fHbp32    -...T.............................IDVVLIK.NNPDKIDSLINQSLVSGLGANQ-S.KEHKS.......H.TQES.ELAESLDTRGGETHALDRIT.IREKVHE.IG
fHbp33    -..........M..............H........DVVLIK.NNPDKIDSLINQSLVSGLGANQ-..KEHKS.......H.TQE..ELAESLDTRGSETHALDRIT.IGEKVHE.IG
fHbp34    S...T.Y..S........................IDVVLIK.NNPDKIDSLINQSLVSGLGANQ-..KEHKSPNRHSTKYRTQES.ELAESLDTRGGETHALDRIT.IREKVHE.IG
fHbp95    -..........M....................TAINVVLIK.NNPDKIDSLINQSLVSGLGANQ-D.KEHKSPNRHSTKYRTQES.ELAESLDTRGGETHALDRIT.IREKVHE.IG
fHbp97    G...V...T.......................TAINVVLIK.NNPDKIDSLINQSLVSGLGANQ-D.KEHKSPNRHSTKYRTQES.ELAESLDTRGGETHALDRIT.IREKVHE.IG
fHbp67    G.........K.EDISQGTTSRFKAI........IDVVLIK.NNPDKIDSLINQSLVSGLGANQ-S.KEHKSPNRHSTKYRTQES.ELAESLDTRGGETHALDRIT.IREKVHE.IG
fHbp84    G...T.....K.EDIPQGTTS.FKAI.VQK..TAIDVVLIK.NNPDKIDSLINQSLVSGLGANQ-..KEHKS.......H.TQE..ELAESLDTRDSETHALDRIT.IGEKVHESIG
fHbp94    -.........K.EDIPQGTTS.FKAI.VQK..TAIDVVLIK.NNPDKIDSLINQSLVSGLGANQ-..KEHKS.......H.TQE..ELAESLDTRGSETHALDRIT.IGEKVHESIG
fHbp96    G..TT.....K.EDIPQGTTS.FKAI.VQK..TAINVVLIK.NNPDKIDSLINQSLVSGLGANQ-.DKEHKSPNRH.TNYRT..S.ELAESLDTRGSETHALDRIT.IGEKVHE.IG
fHbp98    GSI.T.....K.EDIPQGTTS.FKAI.VQK..TAIDVVLIK.NNPDKIDSLINQSLVSGLGANQ-..KEHKS.......H.TQE..ELAESLDTRGSETHALDRIT.IGEKVHE.IG
```

Figure 6.1: Aligned unique fHbp peptide sequence variable sites. fHbp1 to fHbp39 are subfamily B/variant 1; fHbp16 to fHbp97 are subfamily A/variant 2 and fHbp67 to fHbp98 are subfamily A/variant 3.

The amino acid Glu/Lys at position 154 was present in subfamily B/variant 1 isolates but not subfamily A/variant 2 isolates (Figure 6.1). There was an absence, in subfamily A/variant 2 isolates, of Arg^{204} (here at a.a. 212) considered to be key in antibody binding in the subfamily B/variant 1 antigen (Giuliani *et al*., 2005), where it was substituted with serine. There were 81% of subfamily B/variant 1 isolates that contained Arg^{204}; the rest had a histidine residue at this position. Two subfamily B/variant 1 isolates (id EG 328 and 528; peptide 37) had a substitution of a G for a T in the final stop codon and were thus extended for a further 9 bases. The N-terminal region separates the subfamily A/variant 3 isolate sequences from the other subfamilies/variants (Figure 6.1). Subfamily A/variant 3 sequences contained an insertion at 67–69 a.a. of lysine, aspartic acid and asparagine, not present in the other variants. This insertion has previously been noted as being present in a subset of subfamily A protein sequences (Fletcher *et a l.*, 2004) and is considered one of the 'signatures' for fHbp sequences of this type (Murphy *et al.* , 2009). Three of the five subfamily A/variant 3 sequences also contained a 5 a.a. glycine-rich insertion at the N-terminal end. This insertion has previously been found in sequences of both subfamilies/variants (Beernink & Granoff, 2009; Fletcher *et al.*, 2004) and is thought to be used as a means of lengthening the chain that attaches the folded protein to accommodate differences in LOS length on the outer membrane (Mascioni *et al.*, 2008).

6.2.2 Phylogenetic analysis of fHbp

Genealogical analysis using CLONALFRAME, and phylogenies constructed with neighbour-joining and maximum-likelihood methods resolved protein and nucleotide sequences into two major groups, with the subfamily A/variant 3 isolates branching off from the rest of the subfamily A/variant 2 (Figures 6.2, 6.3). CLONALFRAME, gives equal weight to genetic events that result in one nucleotide change, as single horizontal genetic exchange events that result in many nucleotide changes and did not separate the putative subfamily A/variant 3 isolate sequences from the other subfamily A/variant 2 isolates (Figure 6.2), although they were more distant from them in neighbour-joining and maximum-likelihood phylogenies. There was some clustering of complexes within trees. For example, the serogroup A-associated complexes ST-1 and ST-4/5 were mostly grouped together within subfamily B/variant 1: all 14 ST-1 complex and all 11 ST-4 complex isolates were subfamily B/variant 1 while 10/12 ST-5 complex isolates were of this group. Also, all ST-32-associated isolate sequences were found in this subfamily/variant group. This fHbp type-complex association will be further discussed in section 6.2.3. CLONALFRAME also allows for exploration of genetic events such as horizontal genetic transfer and substitutions in branches above nodes. This will be discussed in more detail in section 6.2.4.

The split decomposition method was also used to assess the relationships amongst the fHbp sequences (Figures 6.4 (a) – (c)). This method does not impose a tree-like structure on a set of sequences if none exits and is therefore especially appropriate for organisms in which recombination occurs. It illustrates all the possible evolutionary paths between linked sequences. All the splitsgraphs were well supported and gave fit values of at least 95%. Split decomposition analysis indicated extensive recombination within fHbp in each of the major groups, but also some among them. Figure 6.4 (a) shows a splitsgraph of all 27 fHbp peptide types and like the other phylogenetic methods used in the study resolves them into two major groups. Figures 6.4 (b) and (c) depict splitsgraphs of the subfamily A/variant 2/3 (17 types) and subfamily B/variant 1 (10 types) respectively and show networks and multiple pathways between sequences which is consistent with much recombination within each group. Within the subfamily A grouping there was a distinct separation of the variants 2 and 3

with them occupying different parts of the graph, although peptide 67 appeared almost midway between the two.

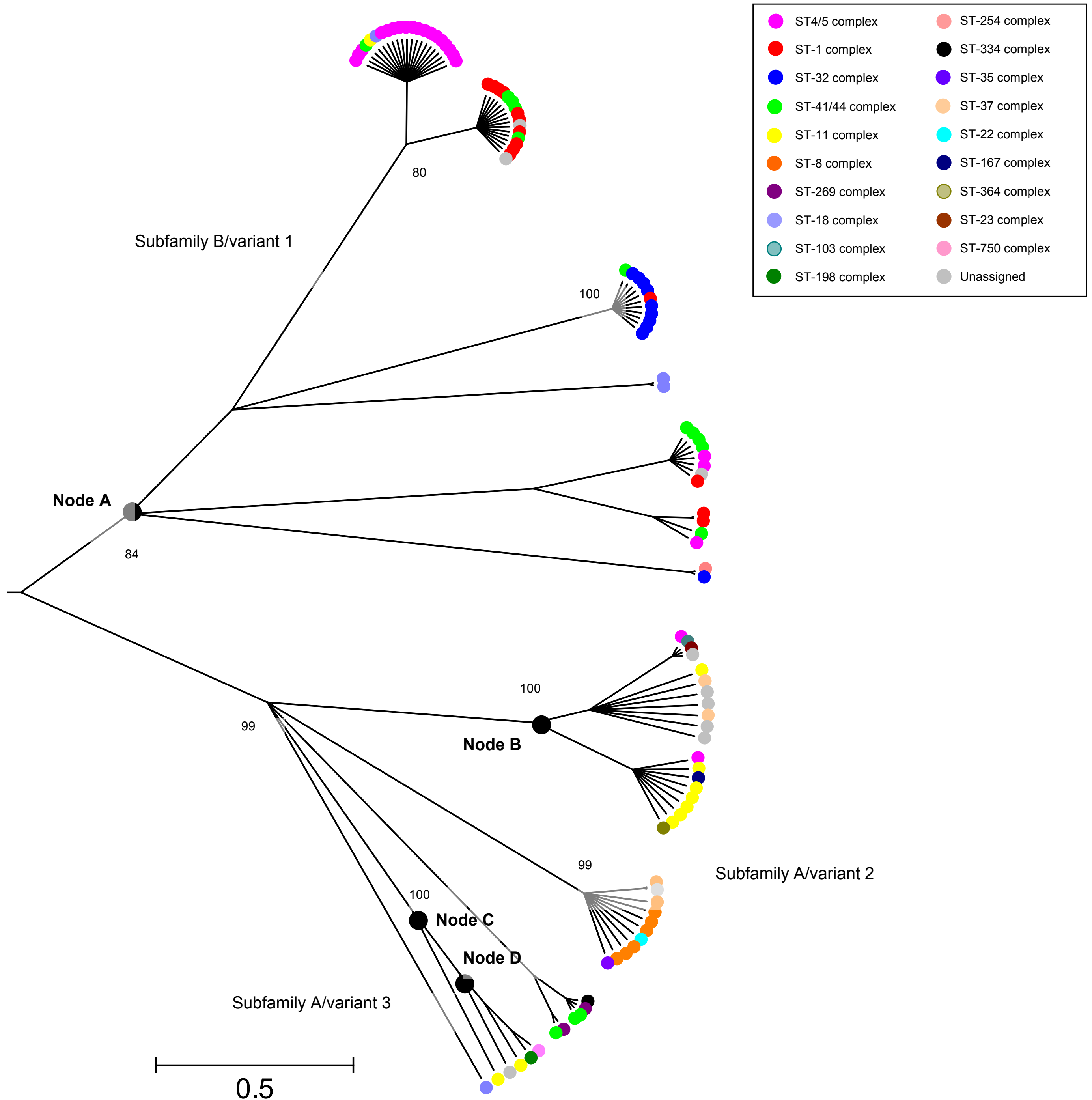
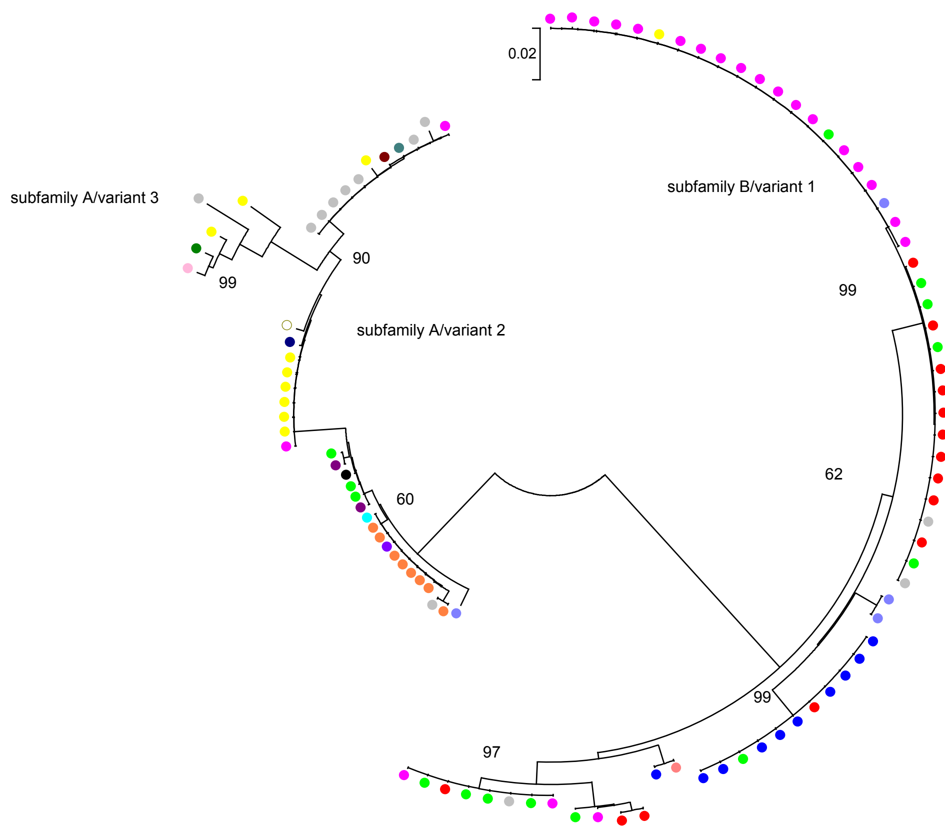


Figure 6.2: A 75% majority-rule consensus ClonalFrame radial tree of 107 aligned nucleotide sequences with colour coding according to clonal complex and confidence values in nodes. Nodes are defined as the most recent common ancestor of the isolates in the branch above it. Bootstrap values shown.

(a)



(b)

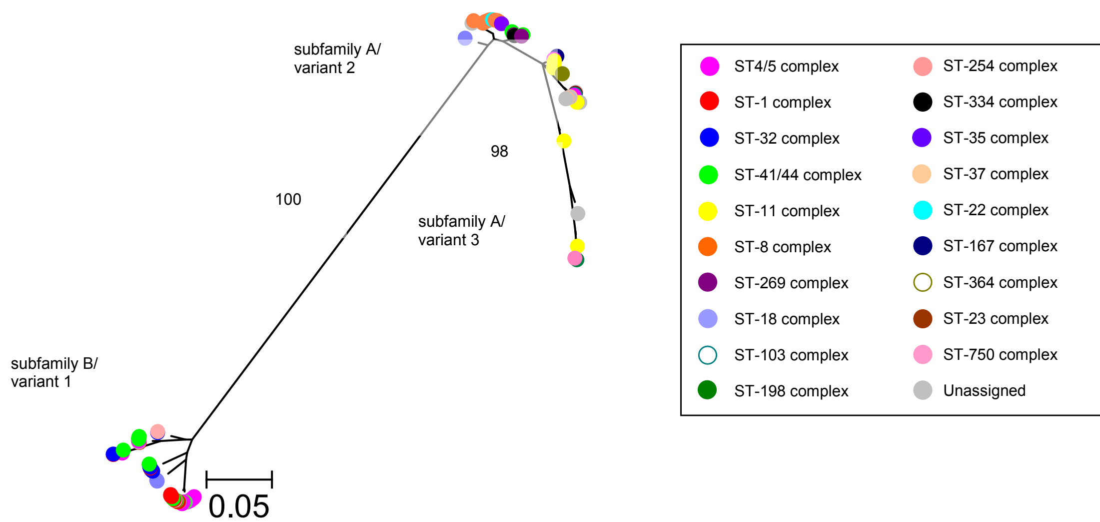


Figure 6.3: (a) Neighbour-joining tree constructed from aligned peptide sequences indicating association with clonal complex. (b) Maximum likelihood tree for the 107 *fHbp* gene sequences. Bootstrap values shown for both trees.

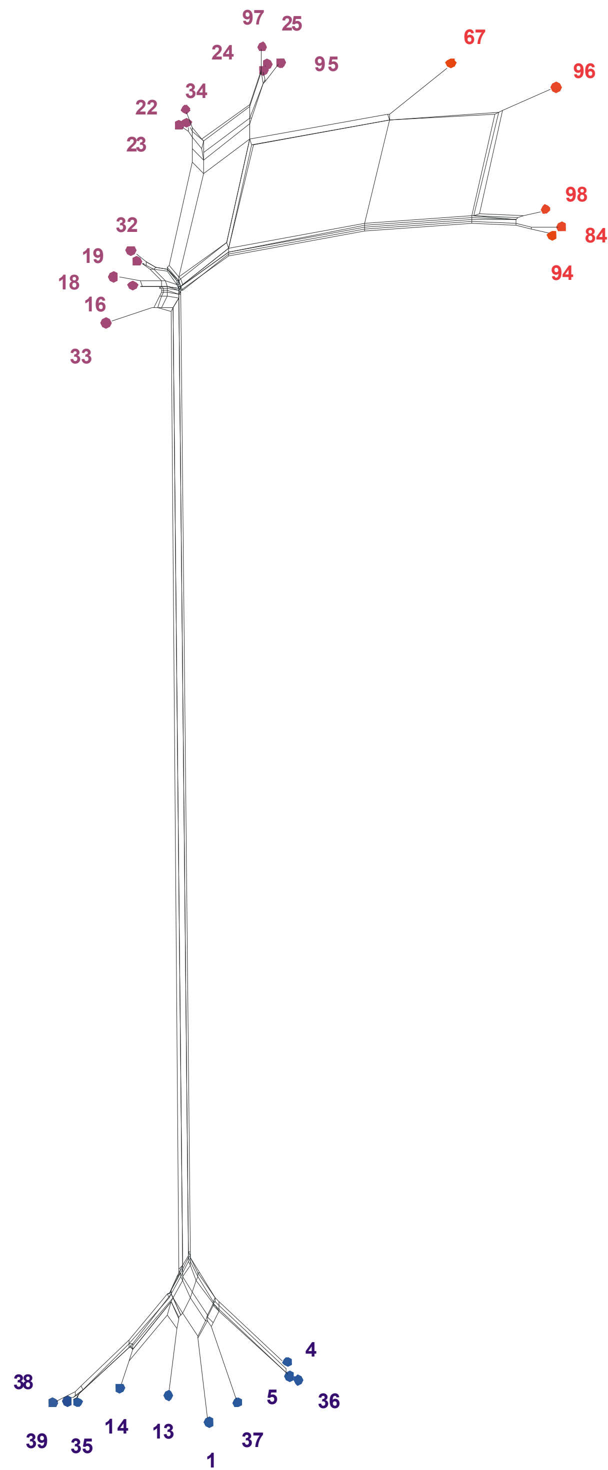
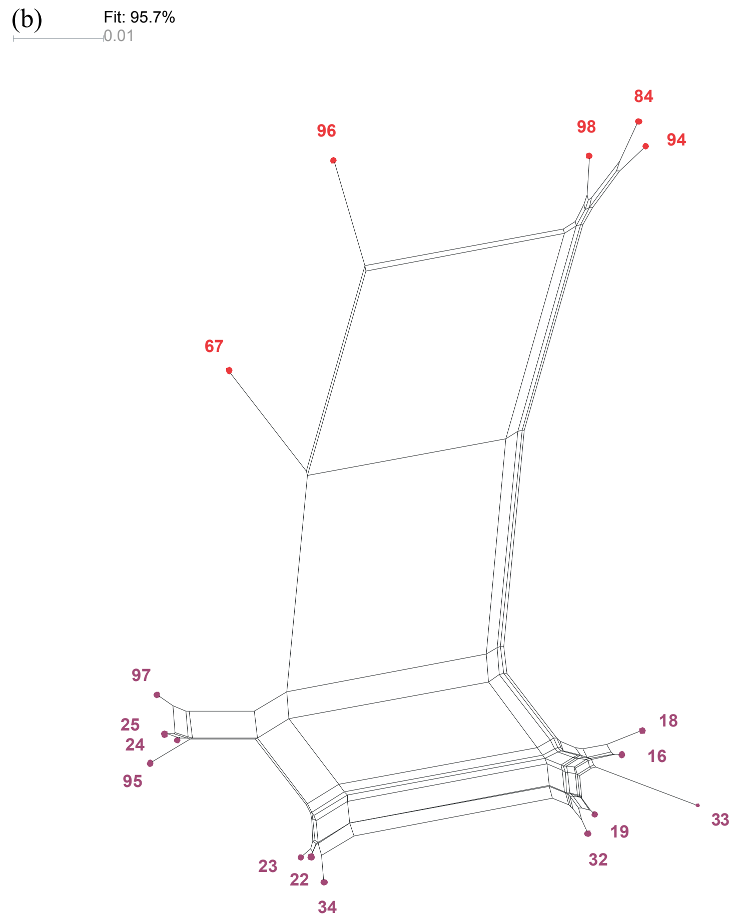


Figure 6.4 (a): SPLITSTREE analysis of all nucleotide sequences corresponding to 27 peptide fHbp types with associated fHbp number indicated. Subfamily A/variant 2 types are coloured purple, subfamily A/variant 3 coloured red and subfamily B/variant 1 coloured blue. Fit value indicated.



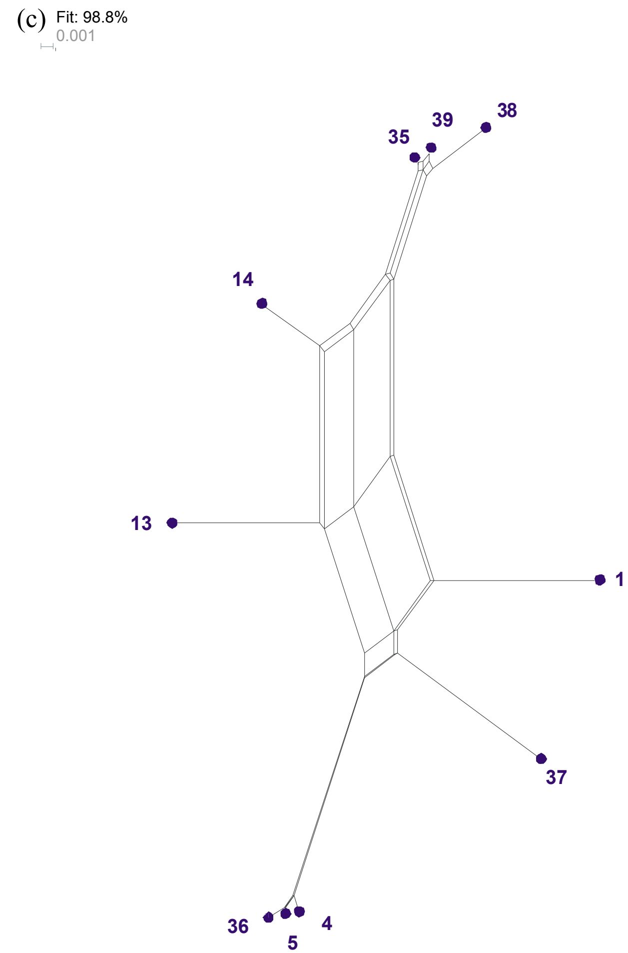


Figure 6.4 (b): SPLITSTREE analysis of subfamily A/variant 2/3 (17 types) and (c) subfamily B/variant 1 (10 types). Fit values indicated. Colour scheme as in (a).

A *Neisseria gonorrhoeae* peptide sequence was obtained from GenBank (accession EEH61327) and used to create a NJ tree along with the 107 study peptide sequences. The gonococcal sequence appeared to be a member of the subfamily A/variant 3 family indicating a shared genetic pool and probable horizontal genetic exchange between the two species or that a common ancestor had this fHbp type (Figure 6.5).

Figure 6.5: Neighbour-joining tree constructed from aligned peptide sequences including a *N. gonorrhoeae* sequence from GenBank (accession EEH61327).

6.2.3 Clonal complex/serogroup association

The distribution of subfamily/variant peptides was not random among clonal complexes, showing some clustering with particular meningococcal genotypes although the relationship was not absolute (Figure 6.2, 6.3, 6.6). The ST-11 complex was associated with subfamily A/variant 2 and in particular a cluster within this subfamily/variant (6 of 8 isolates were fHbp peptide 22; Fisher's exact test $p<0.005$). Similarly, the ST-8 complex was found to be only associated with this subfamily/variant and 7 of 8 isolates had the fHbp 16 peptide (Fisher's exact test $p<0.005$). The serogroup A-associated complexes ST-4 and ST-5 were clustered together and associated mainly with subfamily B/variant 1. The ST-4 complex was particularly

homogeneous as all isolates had the fHbp 5 peptide (Fisher's exact test p<0.005). All the ST-32 complex isolates were found associated with subfamily B/variant 1 with 9/10 isolates having the fHbp 1 peptide type (Fisher's exact test p<0.005). The ST-1 complex was significantly associated with fHbp peptide 4 (10/14 isolates, Fisher's exact test p<0.005). The other main hyperinvasive lineage, ST-41/44 complex, was more diverse with respect to the subfamilies/variants observed.

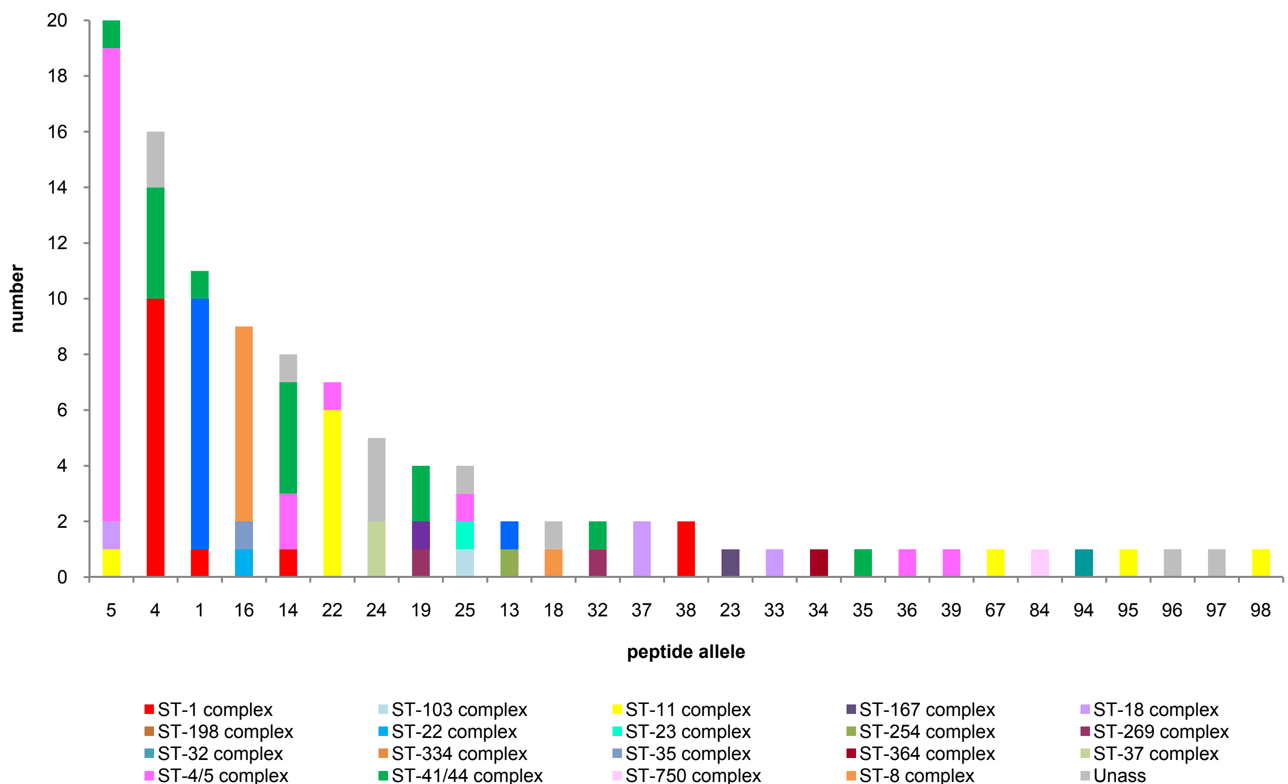


Figure 6.6: Distribution of clonal complexes within fHbp peptide types.

Similarly, there was a relationship between serogroup, particularly non-B serogroups, and variants/subfamilies (Figure 6.7), although this was at least in part due to the known association of clonal complex with serogroup (Trotter *et al.* , 2007). A total of 55% of subfamily B/variant 1 were serogroup A (Fisher's exact test p<0.005) compared to 4% for subfamily A/variant 2. Serogroup C was found in 1.6% of subfamily B/variant 1 while it accounted for 35% of subfamily A/variant 2 (Fisher's exact test p<0.005). There were no W-135, Y or Z subfamily B/variant 1 types while

each serogroup accounted for 4% of subfamily A/variant 2 isolates. Serogroup B was more evenly distributed accounting for 50% of subfamily A/variant 2 isolates and 44% of subfamily B/variant 1 isolates. However, 70% of serogroup B disease-associated isolates were subfamily B/variant 1.

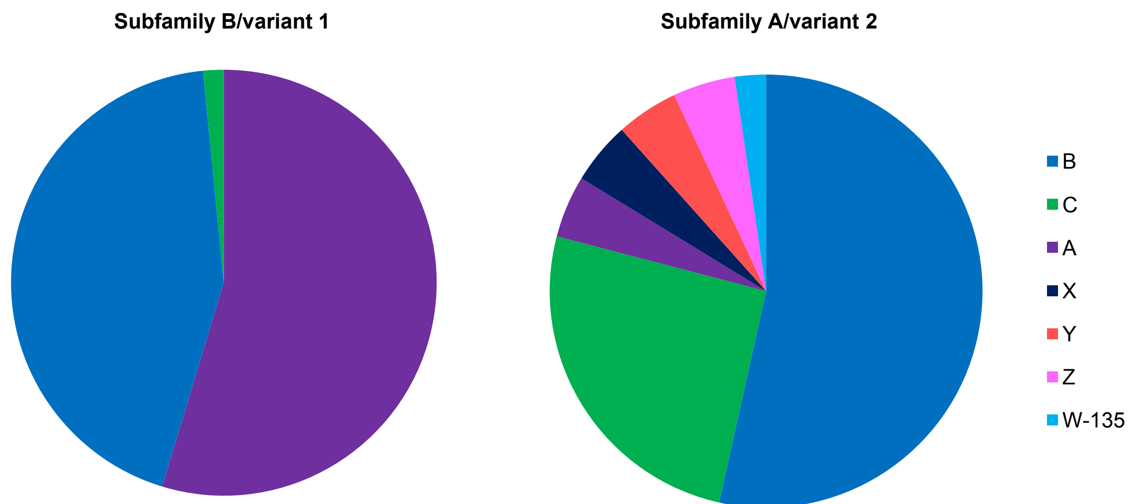


Figure 6.7: Subfamily/variant-serogroup association.

Association between subfamily/variant and whether the isolate was related to disease or carriage has been noted previously (Muzzi *et al.*, 2009) and was supported by the findings here (Table 6.2, Figure 6.8). 68.4% of disease isolates were subfamily B/variant 1 while 64.7% of carriage isolates were subfamily A/variant 2 (Figure 6.8). The odds of a subfamily B/variant 1 isolate being related to disease *versus* carriage is about five times, while a subfamily A/variant 2 isolate is about four times more likely to be related to carriage *versus* disease.

Table 6.2: Association of disease/carriage and subfamily/variant.

subfamily/variant	disease	carriage	odds ratio	95% CIs
subfamily B/variant 1	54	5	5.01	1.64–17.64
subfamily A/variant 2	23	11	0.23	0.070–0.688
subfamily A/variant 3	2	1	0.40	0.030–12.99
	79	17		

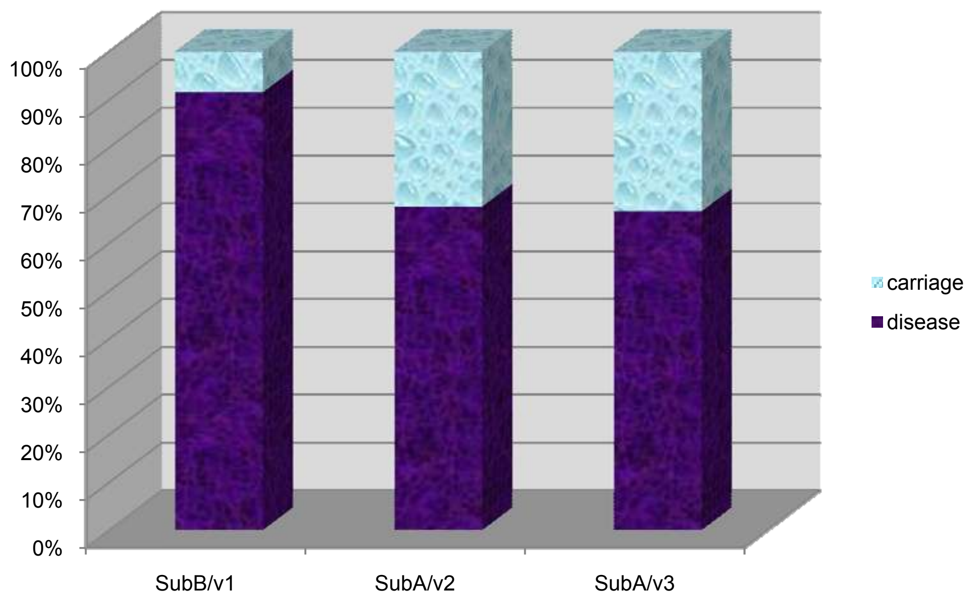


Figure 6.8: Association of disease state with subfamily/variant.

6.2.4 Evidence for recombination

The SPLITSTREE analysis gave results which were consistent with extensive recombination within the *fHbp* gene. Additional analyses were carried out to explore this further. Maximum χ^2 analysis identified putative recombination sites after nucleotide sites 281 and 326 (Table 6.3).

Table 6.3: Maximum χ^2 squared test for recombination. Putative recombination sites in fHbp locus.

sequence 1	sequence 2	putative recombination site	max χ^2	p-value
fHbp-1	fHbp-16	326	122.46	<0.001
fHbp-1	fHbp-19	326	122.46	<0.001
fHbp-1	fHbp-22	326	141.58	<0.001
fHbp-1	fHbp-23	326	141.58	<0.001
fHbp-1	fHbp-24	281	125.23	<0.001
fHbp-1	fHbp-25	281	122.41	<0.001
fHbp-1	fHbp-39	326	112.88	<0.001
fHbp-1	fHbp-40	326	117.05	<0.001
fHbp-1	fHbp-41	326	118.65	<0.001
fHbp-1	fHbp-42	326	119.74	<0.001

CLONALFRAME analysis indicated strong evidence of horizontal genetic exchange within the gene in five major segments (Figure 6.9). Firstly, in the C-terminal region from around 300 bp onwards there were two regions where lateral exchange events were likely to have occurred (node A Figure 6.2 and Figure 6.9 (a)). Also, in the N-terminal region there was strong evidence of lateral gene transfer which may have given rise to this variant within the subfamily A/variant 2 group (nodes C and D (Figure 6.9 (c) and (d)). Another point of recombination identified was above node B which contains subfamily A/variant 2 sequences (Figure 6.9 (b)).

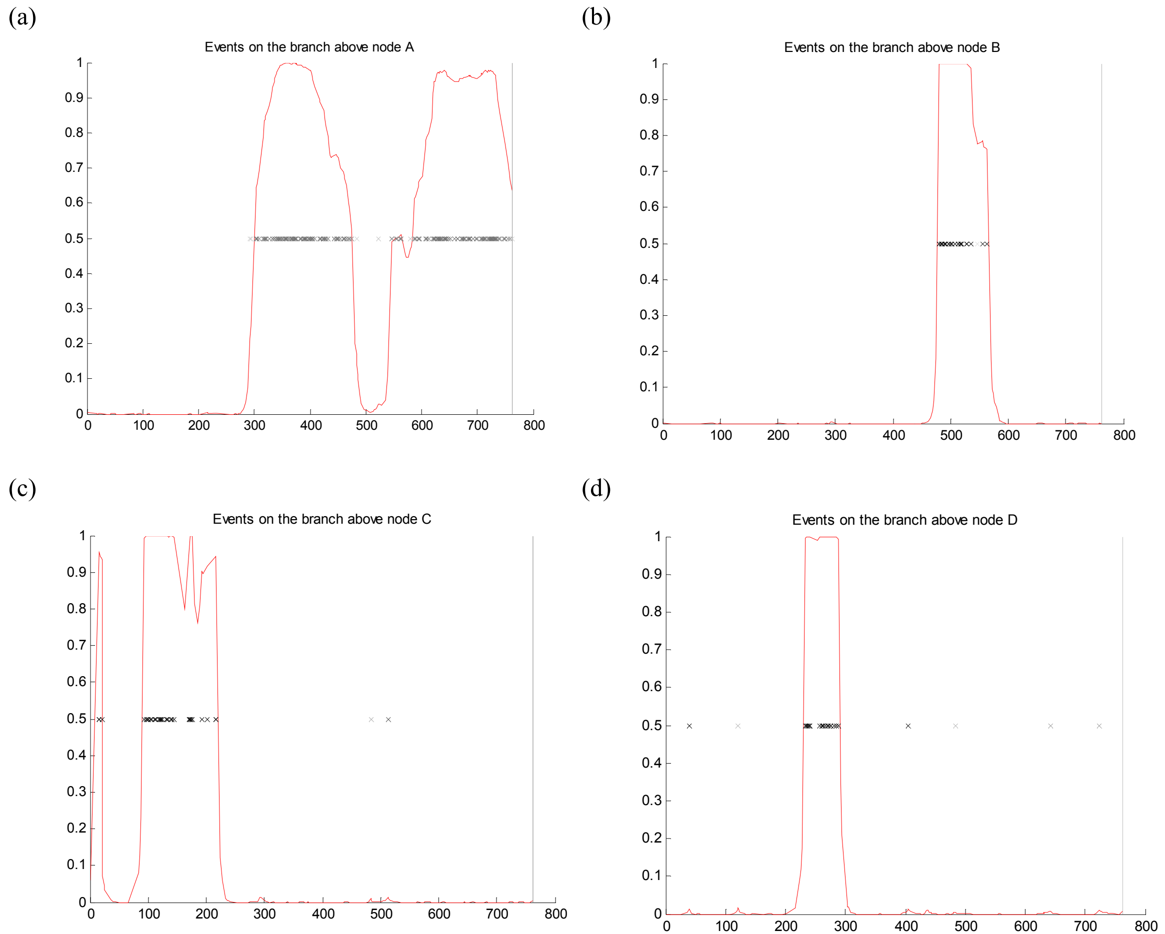


Figure 6.9: Representation of *fHbp* gene recombination events (a) – (d). The nucleotide sequence of *fHbp* gene is on the x axis with the *red* line indicating the probability for an import from zero to one (y axis). The panels depict genetic events above nodes A, B, C and D shown in the 75% majority-rule consensus CLONALFRAME tree panel (Figure 6.2). Each inferred substitution is indicated by a cross, the intensity of which indicates the posterior probability for that substitution. In panel (a) horizontal genetic exchange is depicted occurring from bases 300 to 500 and 550 to 800; in panel (b) from 450–600; panel (c) horizontal genetic exchange is depicted occurring from bases 100 to 250 and in panel (d) at about 200 and 300 bases.

6.2.5 Evidence for selection

The *fHbp* locus had an average d_N/d_S ratio of 0.35 indicating a level of purifying selection against amino acid change. Previous estimates have been 0.51 ± 0.07 (Bambini *et al.* , 2009) and is comparable to that of other antigenic genes such as *fetA* (0.314) (Thompson, 2001). The Synonymous Nonsynonymous Analysis Program (SNAP) algorithm (http://www.hiv.lanl.gov/) was used to analyse pairwise synonymous and nonsynonymous differences in the nucleotide sequences according to the method of Nei and Gojobori (Nei & Gojobori, 1986). Analysis showed that most selection occurred after about 100 codons in the C-terminal region and that it was mainly non-synonymous substitutions occurring at sites indicating positive selection (Figure 6.10).

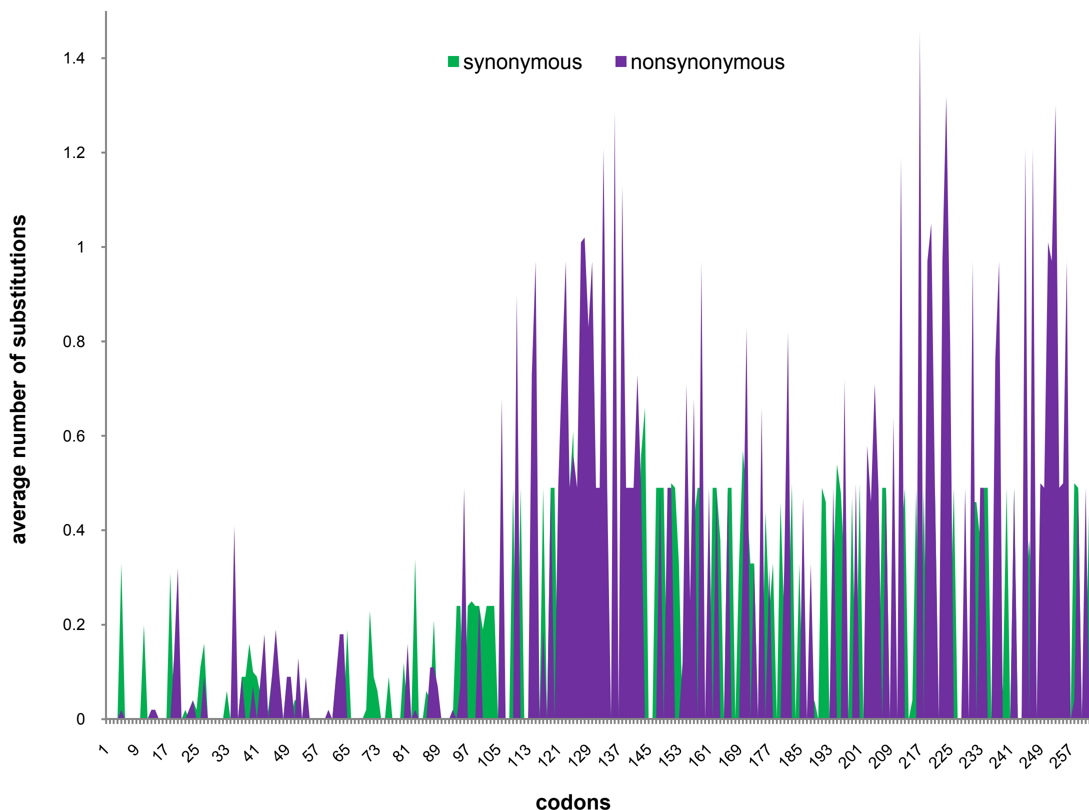


Figure 6.10: SNAP analysis of 107 aligned fHbp sequences with average number of non-synonymous and synonymous substitutions along the sequence indicated.

Codon-by-codon analysis of selection on the gene was possible using OMEGAMAP. Separate analyses for each of the subfamilies/variants, including subfamily A/variant 3, indicated that in the C-terminal region (after ~318 nucleotides encoding 106 amino acids) there was diversifying immune selection ($\omega>1$) acting on particular areas (Figure 6.11 (a) – (f)). Subfamily B/variant 1 and subfamily A/variant 2 (not including variant 3) shared one positively selected codon (147 and 151 in subfamily B/variant 1 and subfamily A/variant 2 respectively). Subfamily B/variant 1 displayed positive selection at the four codons 146–149 (ω 3.41–3.52) and also at codons 195–204 (ω 1.02–1.66). Subfamily A variants 2 and 3 shared the positively selected sites from codons 169–181. The per-site point estimate of ω inferred for each of the subfamily/variant isolates was used to colour a three-dimensional pdb file of the solution structure of a complex between a subfamily B/variant 1 GNA1870/fHbp protein and a region of the fH protein (code 2W81) (Schneider *et al.* , 2009). The temperature colouring of the protein enabled the demonstration of the regions under positive selection on the 3D model (Figure 6.12 (a) (b) & (c)). These regions did not overlap with residues involved interactions with the fH molecule (Schneider *et al.* , 2009).

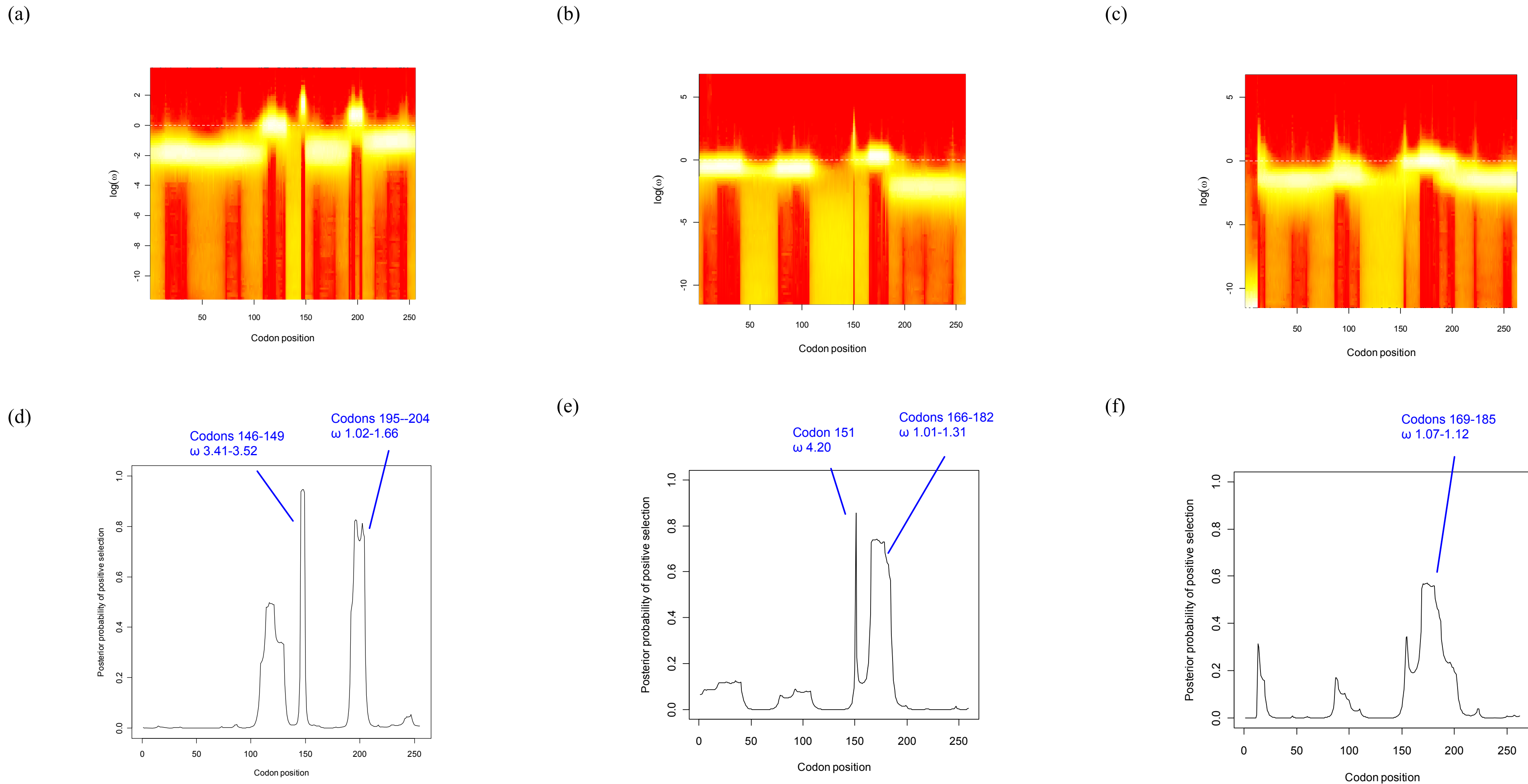
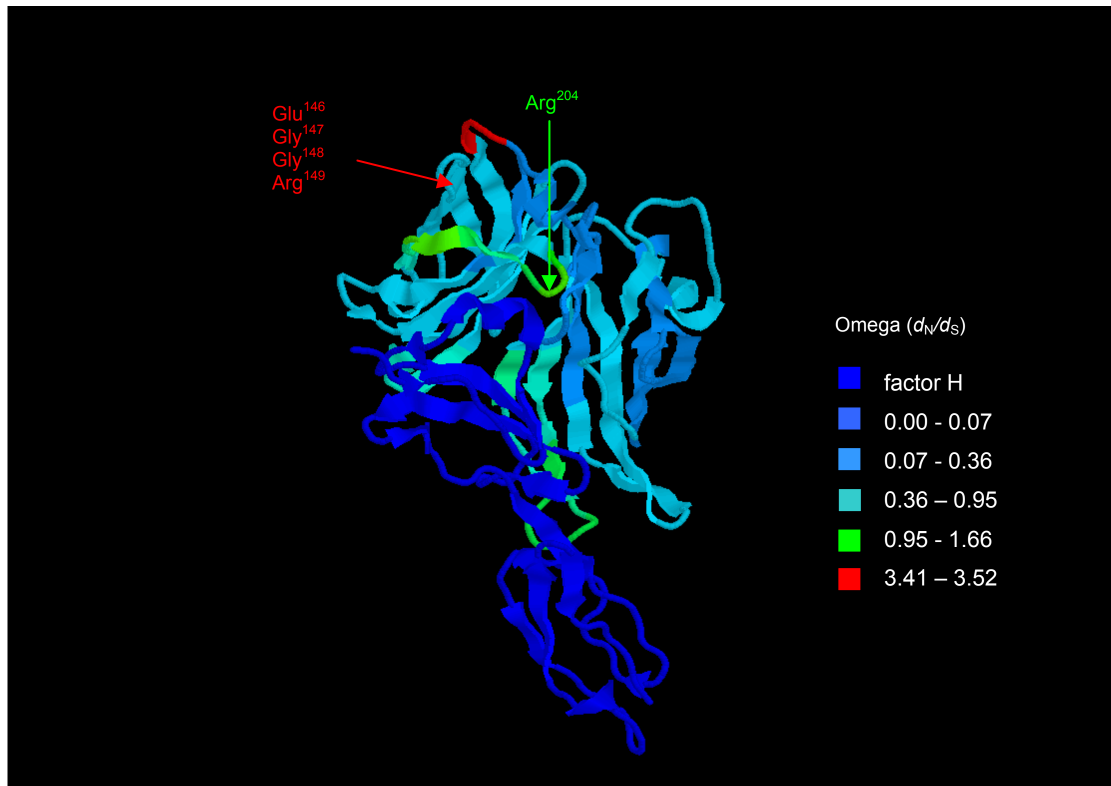
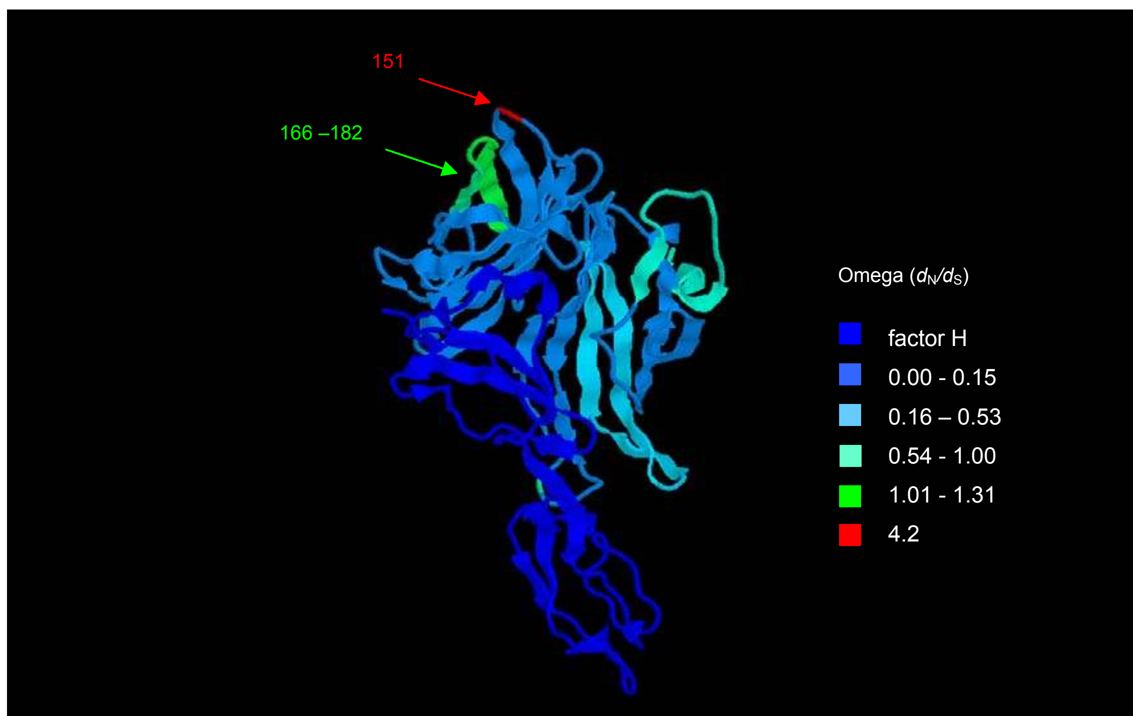


Figure 6.11: OMEGAMAP program output using subfamily A/variant 2 and subfamily B/variant 1 nucleotide sequences. Panels (a), (b) and (c) depict fireplots of the sitewise posterior distribution of log(ω) for subfamily B, subfamily A (without variant 3) and subfamily A/variant 3 sequences respectively. A fireplot visualizes the posterior on log(ω) along the sequence using a colour gradient where a higher posterior density is represented by more intense colour (closer to white), and lower posterior density is represented by less intense colour (closer to red). Panels (d), (e) and (f) depict the posterior probability of positive selection (y-axis, values 0 to 1) along the codon sequence (x-axis) for subfamily B, subfamily A (without variant 3) and subfamily A/variant 3 sequences respectively. Sites with a point estimate of ω>1 are indicated with blue braces. Note: Variants 2 and 3 differ from subfamily B/variant 1 in length by +4 and +7 bp respectively, *e.g.* residue 147 of variant 1 is at position 151 of subfamily A/variant 2 (e) and 154 of subfamily A/variant 3 (d).

(a)



(b)



(c)

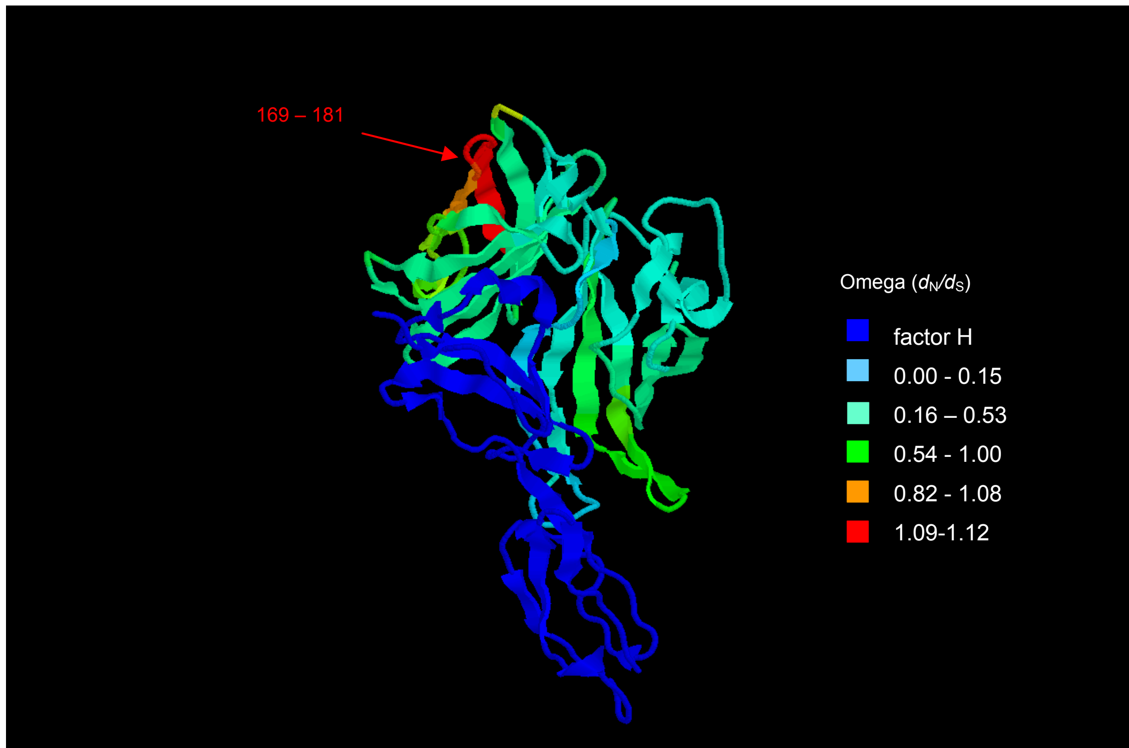


Figure 6.12: Structure of the fH-fHbp complex (Schneider *et al.*, 2009) with temperature colouring using per-site point estimate of ω for (a) subfamily B/variant 1 sequences, (b) subfamily A/variant 2 and (c) subfamily A/variant 3 sequences. Peptides indicated in (a) are putative bactericidal epitopes previously identified (Giuliani *et al.*, 2005; Scarselli *et al.*, 2008; Welsch *et al.*, 2004). In (b) and (c) positively selected sites are indicated. Note: subfamily A/variant 2 and subfamily A/variant 3 differ from subfamily B/variant 1 in length by +4 bp (*e.g.* Glu^{151} is equivalent to Glu^{147} in variant 1) and +7 bp (*e.g.* Glu^{154} is equivalent to Glu^{147} in variant 1) in length respectively.

6.3 Discussion

An ideal vaccine candidate provides cross-protection against all variants of a targeted pathogen. To date, proteins suggested as components of meningococcal vaccines either do not elicit protective immune responses or, like fHbp, are variable (Jodar *et al.*, 2002). Consequently, it is important to catalogue this diversity before a vaccine formulation is finalised to ensure maximum vaccine coverage. For fHbp, a number of studies have been performed to achieve this, (Bambini *et al.*, 2009; Murphy *et al.*, 2009), however, a universal agreed nomenclature is essential to enable comparisons among different studies. Two different fHbp classification schemes have been proposed; one classifies the protein variants of fHbp (referred to as GNA1870) into three variant families, named variants 1, 2, and 3 (Masignani *et al.*, 2003), while the other group's variants of the same protein (referred to as LP2086) are classified into subfamilies A and B (Fletcher *et al.* , 2004). Here, a unified nomenclature was proposed in which unique fHbp peptide and nucleotide sequences are assigned numbers arbitrarily and entered into a database that can be queried and new sequences deposited (http://neisseria.org/nm/typing/fhbp). Using this nomenclature as a basis, higher order classifications can be applied without confusion.

Understanding diversity also requires appropriate isolate collections, with a sample frame appropriate to the question addressed. For this reason, the present study investigated the 107 meningococci used to establish MLST, which includes the globally important disease-associated hyperinvasive meningococcal lineages of all serogroups from the latter half of the twentieth century, which have been extensively characterised (Callaghan *et al.* , 2006; Maiden *et al.* , 1998; Thompson *et al.* , 2003; Urwin *et al.* , 2004). The number of fHbp alleles in this set, 28 encoding 27 peptides was broadly comparable to the number of other surface proteins investigated: PorA (33 alleles encoding 33 peptides); PorB (31 alleles encoding 28 peptides); FetA (33 alleles encoding 31 peptides); and Opa (90 alleles encoding 83 peptides) (Callaghan *et a l.*, 2006; Urwin *et al.*, 2004). The diversity of fHbp resolved into two major clusters by the phylogenetic approaches used, as described previously (Fletcher *et al.*, 2004; Murphy *et al.*, 2009), with evidence of a third group (variant 3) (Masignani *et al.*, 2003). While other variants may be discovered by further studies, especially of carried rather than disease-associated meningococci, comparison with published sequences from various

sources (GenBank, http://neisseria.org/nm/typing/fhbp) demonstrated that the 107 isolates included all the major variant clusters of the protein described to date. Of the previously described groups, subfamily B/variant 1 was the most prevalent (60%) among the 107 isolates. In previous studies it accounted for 54 (70% of isolates) (Beernink *et al.*, 2007; Fletcher *et al.*, 2004; Jacobsson *et al.*, 2006; Masignani *et al.*, 2003; Murphy *et al.*, 2009; Welsch *et al.*, 2004).

There was evidence that fHbp alleles and consequently variant peptides are generated by horizontal genetic exchange, as is the case for other meningococcal antigens (Bennett *et al.*, 2008; Derrick *et al.*, 1999; Harrison *et al.*, 2008). Further, the gene is also present in *N. gonorrhoeae*, with the gonococcal fHbp sequences described to date belonging to subfamily A/variant 2 group (Fletcher *et al.*, 2004; Masignani *et al.*, 2003). It is possible that subfamily A/variant 3 arose through a recombination event with a DNA fragment donated from another member of the genus *Neisseria* or that a common ancestor of the organisms had this fHbp type. This is supported by a neighbour-joining phylogenetic analysis of peptide sequences that clustered the subfamily A/variant 3 sequences between a gonococcal fHbp peptide sequence and other subfamily A/variant 2 sequences. Common gene pools have been documented for other *Neisseria* antigens such as PorB2, FetA and TbpB (Bennett *et al.*, 2008; Derrick *et al.*, 1999; Harrison *et al.*, 2008). In the case of PorB and TbpB, different variant classes are thought to have arisen due to inter-species recombination.

A recent study suggests that fHbp has a modular architecture with five variable modular segments flanked by invariant residues (Beernink & Granoff, 2009). This agrees with the ClonalFrame analysis carried out in this study which showed five main regions in the gene where lateral gene transfer events are likely to have occurred and they are in roughly the same positions. Beernink *et al.* speculate that progenitors of the pathogenic *Neisseria* recombined to form modular fHbp groups with some evidence that subfamily A/variant 2 is a hybrid of variable segments from subfamily B/variant 1 and subfamily A/variant 3. Currently there are nine known modular groups (Pajon *et al.*, 2010). This modular grouping of fHbp may have implications for vaccine design and also classification of the protein.

While *N. gonorrhoeae* appears to have a *fHbp* gene, it is known to bind fH *via* porin proteins (Ngampasutadol *et al.*, 2008). Both meningococci and gonococci have specificity for human fH, which may partly explain their restriction to humans (Granoff *et al.*, 2009; Welsch & Ram, 2008). The gene encoding fHbp has also been detected in

the commensal species *Neisseria cinerea* and *Neisseria lactamica* and a potential fHbp peptide detected by Western blot analysis (Fletcher *et al.*, 2004; Masignani *et al.*, 2003). However, the distribution of the *fHbp* gene among all the *Neisseria* species and its function in the non-pathogenic organisms is yet to be fully elucidated.

Despite the genetic and antigenic diversity of carried populations of meningococci (Caugant & Maiden, 2009; Yazdankhah & Caugant, 2004), most invasive meningococcal disease is caused by a small number of clonal complexes, known as the hyperinvasive lineages (Caugant, 2001; Yazdankhah *et a l.*, 2004). In common with other variable antigens (Callaghan *et al.* , 2008a; Harrison *et al.* , 2008; Trotter *et al.* , 2007; Urwin *et al.* , 2004), the distribution of fHbp variants was not random among clonal complexes, with certain variants more likely to be found in given hyperinvasive lineages, as seen in other isolate collections (Bambini *et al.* , 2009; Jacobsson *et al.* , 2006; Masignani *et al.*, 2003). The ST-32 complex and serogroup A were significantly associated with particular subfamily B/variant 1 fHbp peptides (1 and 5 respectively). The ST-11 complex, which can be distinguished from other hyperinvasive lineages by harbouring only *tbpB* isotype I and lacking the *opcA* gene (Claus *et al.* , 2001), was significantly associated with subfamily A/variant 2 and fHbp peptide 22. Similarly, stable associations have been observed in serogroup A, X and W-135 meningococci in Africa over a 45 year period (Beernink *et al.*, 2009a). It should be noted, however, that while these associations exist, they were not absolute. For example, while most of the ST-32 complex isolates were peptide 1, there was an isolate with the fHbp peptide 13 and several of the complexes, while they may have a dominant fHbp peptide type, can also contain others. The ST-41/44 complex in particular was heterogeneous, with multiple fHbp peptides. This has been shown previously (Bambini *et al.* , 2009; Beernink *et al.*, 2007; Jacobsson *et al*., 2006; Masignani *et al.*, 2003). The reasons for these associations are not fully understood but models of strain structure in recombining pathogens show that immune selection can, counter-intuitively, lead to the stable associations of antigenic variants characteristic of meningococcal hyperinvasive lineages (Buckee *et al.*, 2008; Callaghan *et al.*, 2008b; Gupta *et al.*, 1996).

The availability of protein structures for fHbp, including one with the protein bound to factor H (Schneider *et al.*, 2009), allowed analysis of the sequence variability of regions encoding different structural and functional domains. Variation in peptide sequence is present throughout fHbp, rather than being limited to particular variable regions as is the case in FetA and the PorA, and PorB2 (but not PorB3) porins. Most

fHbp diversity was found in the C-terminal region (~161 a.a. in length) of fHbp, while there was less in the N-terminal region (~105 a.a. in length), which contained a domain that anchors the protein to the cell membrane (Mascioni *et al.*, 2008). Other invariant regions of the protein are those involved with factor H interaction (particularly within subfamilies/variants although some of the interaction residues show some polymorphism (Schneider *et al.*, 2009)) and residues that make up hydrophobic cores of the β-barrels and the points of contact between the N and C-terminal domains (Mascioni *et al.*, 2008). Beernink *et al.* found invariant blocks between variable segments clustered together on the side of the protein that is anchored to the cell wall (Beernink & Granoff, 2009).

The selection pressures acting on fHbp were deduced by means of a Bayesian algorithm, which has a number of advantages over the maximum likelihood approaches used previously to analyse selection pressures on the PorB protein (Urwin *et al.*, 2002), and the results compared to those obtained in functional studies. Epitope mapping has identified the residue Arg^{204} as being essential for the binding of a bactericidal monoclonal antibody (MAb) (Giuliani *et al.*, 2005; Scarselli *et al.*, 2008). Residues also identified as potentially involved in a conformational epitope with Arg^{204} are residues Glu^{146} – Arg^{149} (Cantini *et al.*, 2006; Scarselli *et al.*, 2008; Welsch *et al.*, 2004). Due to their placement and clustering in structural models, it is thought that these residues could make up a bactericidal epitope in the C-terminal region with the potential cooperation of other residues (Cantini *et al.*, 2006; Scarselli *et al.*, 2008) and have been shown to be placed away from the fH recognition site and therefore may not interfere with fH binding (Schneider *et al.*, 2009). The selection analyses identified these residues as displaying evidence of immune selection on the subfamily B/variant 1 protein, underlining their potential relevance as protective epitopes. One of these residues also showed evidence of positive selection in subfamily A/variant 2; 151 (147 in subfamily B/variant 1).

Other protective epitopes identified to date include residues 121–122 present in subfamily B/variant 1 proteins, residues between 25–59 present in subfamily A/variant 2 and subfamily B/variant 1 and residues between positions 174 to 216 of variant 2 and 3 proteins (Beernink & Granoff, 2008; Beernink *et al.*, 2008; Beernink *et al.*, 2009b). Particularly in meningococci expressing fHbp at low levels, these epitopes can induce bactericidal activity by eliciting cooperative pairs of MAbs that and also inhibit fH binding thus increasing complement-mediated activity (Beernink *et al.*, 2008). The

selection analysis provided evidence for positive selection in a region that partially overlapped with one of these putative bactericidal epitopes found in the C-terminal region of the subfamily A/variant 2 and subfamily A/variant 3 proteins, *i.e.* residue 174. It is very encouraging that such analyses can be used to predict regions involved in immune interactions of other bacterial proteins.

Molecular epidemiology has played a major role in the development, implementation, and study of meningococcal vaccines (Bjune *et al.*, 1991; Maiden *et al.*, 2008; O'Hallahan *et al.*, 2005; Rodriguez *et al.*, 1999). For candidate protein components it is essential to determine the number of variants required and to identify those likely to provide the broadest possible protection, ideally before a vaccine formulation is tested in humans. Although the use of functional assays is important, nucleotide and peptide sequence diversity give important guides to this process. In the case of fHbp, the existence of multiple variants and the evidence for particular epitopes under immune selection indicates that, as for other meningococcal antigens it will be important to use vaccine formulations with multiple components to achieve broad coverage, particularly as it has been shown that cross-protection between the two major subfamilies and within, subfamily A, variants 2 and 3, is limited (Beernink *et al.*, 2007; Fletcher *et al.*, 2004; Masignani *et al.*, 2003). An alternative strategy is to create chimeric proteins containing domains from the different subfamilies/variants (Beernink & Granoff, 2008). However multivalency is achieved, the optimum number of variants to be used will depend on a combination of molecular epidemiological and functional studies. Further, the lifespan of such vaccines will depend on the dynamics of fHbp evolution in natural populations of meningococci and any possible effects of vaccination on this process. The nomenclature scheme and analytical framework described here should contribute to assembling the information required to answer these questions.

CHAPTER 7: Concluding remarks and future work

7.1 Conclusions

There are relatively few large and representative surveys of meningococcal disease and fewer still that utilise the more recently developed molecular typing techniques such as MLST. This is changing however and a number of studies have shown the relative ease by which these techniques can be employed to give a comprehensive picture of the meningococcal disease and carriage populations (Claus *et al.*, 2005; Jolley *et al.*, 2002; Maiden *et al.*, 2008; Russell *et al.*, 2008; Yazdankhah *et al.*, 2004). These types of surveys are important as they help us understand better the diversity of the meningococcal population and the dynamics of transmission. The goal of the EU-MenNet consortium was to improve the understanding of the spread of one of the most severe childhood diseases, meningococcal disease. This was achieved by integrated epidemiological and population genetics studies, including that described in this work, employing the latest molecular isolate characterisation techniques and electronic data transfer *via* the internet. This involved substantial coordination and integrated data collection among 18 countries of the project. The legacy of such work is the enhanced European expertise in molecular epidemiology for research and surveillance, improved collaboration among labs and the creation of datasets which can be analysed and interpreted to enhance our insight into the spread of meningococcal disease and how best to prevent and control it. This study is the largest comprehensive and representative molecular epidemiological survey of meningococcal disease across the continent of Europe known to date.

The meningococcus is a naturally transformable organism which resides commensally in the human nasopharynx and is therefore constantly exposed to the human immune system. Diversity of the organism is a consequence of such influences and previous studies have shown a highly heterogeneous meningococcal population. Despite this however, it is consistently found that the disease population is dominated by relatively few lineages (Caugant, 1998; Yazdankhah *et al.*, 2004). In this relatively recent European disease sample this was also the case with a handful of types dominating and spread widely across the continent. There were some regional

variations however with some types more associated with particular countries or regions of the continent, *e.g.* ST-162 complex in Greece and ST-269 complex in the north-western fringe of Europe in the UK and the Republic of Ireland. Moreover, small but significant F_{ST} values indicated a degree of geographically restricted gene-flow in Europe except between Sweden and Norway where there was no significant difference in genetic distribution. Overall, geography contributed to 4.4% of the diversity of MLST concatenated sequences. Taken together, these observations suggest that these successful lineages which have persisted worldwide for some time have maintained their presence across Europe, but there may be uneven dispersion of them and their rate spread is slow enough such that geographic structuring is evident. Consequently, the neighbouring countries Norway and Sweden which are relatively isolated physically from other European countries are indistinct from each other in terms of genetic distribution. Therefore, transmission is still likely to be relatively more important within countries than among them.

Serogroup B is still the major disease-associated serogroup in Europe (EU-IBIS, 2006; Trotter *et al.* , 2007) and also the one for which no comprehensive vaccine currently exists. Investigation into vaccine candidates has focussed on a number of OMPs which include PorA and FetA. It is important therefore to know the diversity and distribution of such antigens in a population where such a vaccine may be employed to estimate potential coverage. These antigens were found to be highly heterogeneous in the European dataset but this diversity was structured in a number of ways. Their distribution was non-random with a small number of dominant types accounting for most disease. Non-overlapping combinations of antigens, particularly the PorA VRs, were found which is consistent with selection by host immunity. These dominant antigens were also associated with clonal lineages which are associated with disease. These strain types have been shown to persist in spite of much recombination over wide geographic and temporal spread (Gupta *et a l.*, 1996; Urwin *et al.* , 2004). Together with their immunogenicity, these features make these antigens attractive vaccine candidates as a relatively small number of types could have the potential to give broad coverage that may have long range potential in terms of time and distance.

Already it has been shown that both phenotypic and genotypic typing is useful in determining the meningococcal strains responsible for outbreaks and allowed for the tailoring of specific OMV vaccines in New Zealand, Norway and Cuba (Bjune *et a l.*, 1991; O'Hallahan *et al.* , 2005; Rodriguez *et al.* , 1999). In the endemic setting, such

vaccines would be inappropriate and instead cocktail vaccines containing a number of antigens are more suitable. This study has shown the ready amenability and usefulness of molecular epidemiology in vaccine design. A proposed vaccine recipe containing four each of PorA and FetA antigens gave a theoretical coverage of almost 79% in the European dataset which compares well with any of the currently available or in development vaccines such as NonaMen. Furthermore, since the full extent of cross protection is still relatively unknown, coverage may be yet higher. An advantage of a putative vaccine such as this which is based on subcapsular antigens is that it can potentially provide protection across all serogroups not just serogroup B.

Efforts to develop serogroup B vaccines by means such as reverse vaccinology (Rappuoli, 2000) have thrown up an array of new vaccine candidates such as fHbp, a principal component of two recombinant protein vaccines in clinical trials at the time of writing (Anderson *et al.*, 2009; Biolchi *et al.*, 2009). A well characterised dataset which includes the globally important disease-associated hyperinvasive meningococcal lineages of all serogroups from the latter half of the twentieth century demonstrated that like many other surface-exposed vaccine candidates fHbp is diverse. Again like PorA and FetA, this diversity was structured with a small number of types accounting for most of the collection and there were associations with clonal lineage and serogroup. With two competing nomenclatures currently used for this antigen and with an increasing amount of molecular epidemiological work being carried out on it, it was proposed that a universal classification scheme be adopted. In this scheme fHbp peptide and nucleotide sequences are assigned numbers arbitrarily and are independent of subfamily/variant. A Web-accessible database was established to facilitate querying of sequences and submission of new allele sequences. Using this nomenclature as a basis, higher order classifications can be applied without confusion. A novel use of sequence data is the detection of selection acting on the gene by means of Bayesian algorithms. A similar approach using maximum likelihood was able to identify residues in the surface loops of PorB are under very strong positive selection (Urwin *et a l.*, 2002). The selection analyses in this study identified residues displaying evidence of immune selection which were previously identified involved in MAb binding and potentially in a conformational epitope. This therefore underlines their potential relevance as protective epitopes.

This work has shown the many applications of sequence typing technologies, from molecular epidemiology, phylogeny, evolutionary biology, surveillance and

vaccine design. With the continued development of next-generation sequencing technologies, there will be further and even greater insight into these areas of biology in the coming years.

7.2 Future work

Continued detailed surveillance of meningococcal disease in Europe is important for many purposes. It can be used for detection of shifts in distribution of types found be they natural fluctuations or due to interventions such as immunisation programs. Persistence of genotypes over relatively long time periods is known, however shifts do occur. For example, prior to the 1970s serogroup A disease was common in Europe but is now virtually non-existent and the reasons for its departure are not fully understood. Moreover, serogroup Y associated disease has shown a marked increase in the United States since the early 1990s when previously it was rarely found. MLST along with antigen type information, *e.g.* serogroup, PorA/FetA offers high discriminatory power. The ECDC's Europe-wide TESSy surveillance system nomenclature for *N. meningitidis* includes serogroup, PorA, FetA and MLST information. Surveillance tools such as EMERT have also been developed to monitor trends in real time and as of the 1st January 2010 contained information for about 1500 isolates. Molecular typing can also be used in disease cluster detection using software tools such SaTScan (Elias *et al.* , 2006).

Collaboration among countries in Europe will allow for assessment of spread of types across the continent whether they be newly emerging, newly introduced or re-emerging types. Also, comparison with data from other regions of the world will pinpoint any global spread of strains and possible routes of introduction of types, *e.g.* returning pilgrims from Mecca to various parts of the world. In this age of international travel and more porous borders, particularly in Europe, never before has it been so easy for the introduction of new strains into new populations. Introduction of invasive types into immunologically naïve populations is particularly worrisome. But also, never before has there been such ease of communication and dissemination of information amongst labs with ready electronic transfer of molecular and epidemiological data that it will be possible to have easy global collaboration. Increasing agreement on classification schemes and readily accessible public databases has aided comparison and collaboration.

Surveillance following the introduction of a vaccine can also help measure its impact in terms of direct and indirect protection and track changes in antigen distribution that may occur through natural fluctuations, as a result of a vaccine which

may include escape variants. It may over time be necessary to re-formulate the vaccine recipe. As yet, there is little information on the amount of cross-protection amongst PorA and FetA variant families. This will be important to evaluate as it will affect the potential coverage of the vaccine recipe.

Many molecular epidemiological studies of the diversity of fHbp in disease samples have focussed mainly on serogroup B (Bambini *et al.* , 2009; Murphy *et al.* , 2009). However, since the gene is found in all meningococci so far tested there is potential for protection across the serogroups. Therefore, fHbp typing of a large and well-characterised disease collection such as the EUMenNet sample would provide great insights into the diversity of this vaccine across a wide geographic region. The potential coverage of the new fHbp-containing vaccines could be assessed. Additional collections that may be of interest include the UK carriage and disease collections.

With the continuing development and progression of new sequencing technologies and their wider availability, there will be an increasing opportunity to compare the genomes of many organisms with relative ease. A population genomics approach will allow the differences amongst lineages in their invasiveness and transmissibility to be elucidated further. Also, by comparison with other commensal species, perhaps the key factors that led to the emergence of pathogenicity in the meningococcus will be clarified.

Is maith an scéalaí an aimsir.
Seanfhocal gaeilge

REFERENCES

(2007). Meningitis Research Foundation.

Achtman, M. (1990). Molecular epidemiology of epidemic bacterial meningitis. *Rev Med Microbiol* 1, 29-38.

Achtman, M. (1995). Global epidemiology of meningococcal disease. In *Meningococcal Disease*, pp. 159-175. Edited by K. A. V. Cartwright. Chichester: John Wiley and Sons.

Achtman, M. (1996). A surfeit of YATMs? *J Clin Microbiol* 34, 1870.

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

Achtman, M., van der Ende, A., Zhu, P. & other authors (2001). Molecular epidemiology of serogroup a meningitis in Moscow, 1969 to 1997. *Emerg Infect Dis* 7, 420-427.

Aguilera, J. F., Perrocheau, A., Meffre, C. & Hahne, S. (2002). Outbreak of serogroup W135 meningococcal disease after the Hajj pilgrimage, Europe, 2000. *Emerg Infect Dis* 8, 761-767.

Ahmad, K. (2004). Vaccination halts meningitis outbreak in Burkina Faso. *The Lancet* 363, 1290.

Ala'Aldeen, D. A., Davies, H. A. & Borriello, S. P. (1994). Vaccine potential of meningococcal FrpB: studies on surface exposure and functional attributes of common epitopes. *Vaccine* 12, 535-541.

Ala'Aldeen, D. A. & Borriello, S. P. (1996). The meningococcal transferrin-binding proteins 1 and 2 are both surface exposed and generate bactericidal antibodies capable of killing homologous and heterologous strains. *Vaccine* 14, 49-53.

Alcala, B., Arreaza, L., Salcedo, C., Uria, M. J., De La Fuente, L. & Vazquez, J. A. (2002). Capsule switching among C:2b:P1.2,5 meningococcal epidemic strains after mass immunization campaign, Spain. *Emerg Infect Dis* 8, 1512-1514.

Alonso, J. M., Gilmet, G., Le-Rouzic, E. M. & other authors (2007). Workshop on vaccine pressure and *Neisseria meningitidis* , Annecy, France, 9-11 March 2005. *Vaccine* 25, 4125-4129.

Andersen, J., Berthelsen, L., Bech Jensen, B. & Lind, I. (1998). Dynamics of the meningococcal carrier state and characteristics of the carrier strains: a longitudinal study within three cohorts of military recruits. *Epidemiol Infect* 121, 85-94.

Anderson, A., Jones, T., Arora, A. & other authors (2009).Development of a factor H-binding protein vaccine for broad protection against invasive *Neisseria meningitidis* serogroup B (MNB) disease. In *European Monitoring Group for Meningococci* . Manchester, UK.

Arkwright, J. A. (1907). On Variations of the Meningococcus and Its Differentiation from Other Cocci Occurring in the Cerebro-Spinal Fluid. *The Journal of Hygiene* 7, 193-204.

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

Ashton, F. E., Ryan, J. A., Borczyk, A., Caugant, D. A., Mancino, L. & Huang, D. (1991). Emergence of a virulent clone of *Neisseria meningitidis* serotype 2a that is associated with meningococcal group C disease in Canada. *J Clin Microbiol* 29, 2489-2493.

Avery, O. T. & Goebel, W. F. (1929). Chemo-immunological studies on conjugated carbohydrate-proteins: II. Immunological specificity of synthetic sugar-protein antigens. *J Exp Med* 50, 533-550.

Aycock, W. L. & Mueller, J. H. (1950). Meningococcus carrier rates and meningitis incidence. *Bacteriol Rev* 14, 115-160.

Baker, M. G., Martin, D. R., Kieft, C. E. & Lennon, D. (2001). A 10-year serogroup B meningococcal disease epidemic in New Zealand: descriptive epidemiology, 1991-2000. *J Paediatr Child Health* 37, S13-19.

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

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

Bandelt, H. J. & Dress, A. W. (1992). Split decomposition: a new and useful approach to phylogenetic analysis of distance data. *Mol Phylogenet Evol* 1, 242-252.

Beernink, P. T., Welsch, J. A., Harrison, L. H., Leipus, A., Kaplan, S. L. & Granoff, D. M. (2007). Prevalence of factor H-binding protein variants and NadA among meningococcal group B isolates from the United States: implications for the development of a multicomponent group B vaccine. *J Infect Dis* 195, 1472-1479.

Beernink, P. T. & Granoff, D. M. (2008). Bactericidal antibody responses induced by meningococcal recombinant chimeric factor H-binding protein vaccines. *Infect Immun* 76, 2568-2575.

Beernink, P. T., Welsch, J. A., Bar-Lev, M., Koeberling, O., Comanducci, M. & Granoff, D. M. (2008). Fine antigenic specificity and cooperative bactericidal activity of

monoclonal antibodies directed at the meningococcal vaccine candidate, factor H-binding protein. *Infect Immun* 76, 4232-4240.

Beernink, P. T., Caugant, D. A., Welsch, J. A., Koeberling, O. & Granoff, D. M. (2009a). Meningococcal factor H-binding protein variants expressed by epidemic capsular group A, W-135, and X strains from Africa. *The Journal of Infectious Diseases* 199, 1360-1368.

Beernink, P. T. & Granoff, D. M. (2009). The modular architecture of meningococcal factor H-binding protein. *Microbiology* 155, 2873-2883.

Beernink, P. T., LoPasso, C., Angiolillo, A., Felici, F. & Granoff, D. (2009b). A region of the N-terminal domain of meningococcal factor H-binding protein that elicits bactericidal antibody across antigenic variant groups. *Mol Immunol* 46, 1647-1653.

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

Bennett, J. S., Thompson, E. A., Kriz, P., Jolley, K. A. & Maiden, M. C. (2008). A common gene pool for the *Neisseria* FetA antigen. *Int J Med Microbiol*.

Bernfield, L., Fletcher, D. L., Howell, A., Farley, J. E., Zagursky, R., Knauf, M. & Zlotnick, G. (2002).Identification of a novel vaccine candidate for group B *Neisseria meningitidis*. In *Thirteenth International Pathogenic Conference* . Edited by D. A. Caugant & E. Wedege. Oslo, Norway.

Bilukha, O. O. & Rosenstein, N. (2005). Prevention and control of meningococcal disease. Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR Recomm Rep* 54, 1-21.

Biolchi, A., Boccadifuoco, G., Galli, B., Vienken, K., Pizza, M., Rappuoli, R., Serruto, D., Donnelly, J. & Giuliani, M. M. (2009).Measuring antigen specific bactericidal responses to a multicomponent vaccine against serogroup B meningococcus. In *European Monitoring Group for Meningococci*. Manchester, UK.

Bisharat, N., Cohen, D. I., Harding, R. M., Falush, D., Crook, D. W., Peto, T. & Maiden, M. C. (2005). Hybrid *Vibrio vulnificus*. *Emerg Infect Dis* 11, 30-35.

Bjune, G., Høiby, E. A., Grønnesby, J. K. & other authors (1991). Effect of outer membrane vesicle vaccine against group B meningococcal disease in Norway. *Lancet* 338, 1093-1096.

Bjune, G. (1992). "Herd immunity" and the meningococcal vaccine trial in Norway [letter]. *Lancet* 340, 315.

Black, J. R., Dyer, D. W., Thompson, M. K. & Sparling, P. F. (1986). Human immune response to iron-repressible outer membrane proteins of *Neisseria meningitidis*. *Infect Immun* 54, 710-713.

Blakebrough, I. S., Greenwood, B. M., Whittle, H. C., Bradley, A. K. & Gilles, H. M. (1982). The epidemiology of infections due to *Neisseria meningitidis* and *Neisseria lactamica* in a northern Nigerian community. *J Infect Dis* 146, 626-637.

Boisier, P., Nicolas, P., Djibo, S. & other authors (2007). Meningococcal meningitis: unprecedented incidence of serogroup X-related cases in 2006 in Niger. *Clin Infect Dis* 44, 657-663.

Borrow, R., Balmer, P. & Miller, E. (2005). Meningococcal surrogates of protection-serum bactericidal antibody activity. *Vaccine* 23, 2222-2227.

Bramley, J. C., Hall, T., Finn, A., Buttery, R. B., Elliman, D., Lockhart, S., Borrow, R. & Jones, I. G. (2001). Safety and immunogenicity of three lots of meningococcal serogroup C conjugate vaccine administered at 2, 3 and 4 months of age. *Vaccine* 19, 2924-2931.

Branham, S. E. (1953). Serological relationships among meningococci. *Bact Rev* 17, 175-188.

Brehony, C., Jolley, K. A. & Maiden, M. C. (2007). Multilocus sequence typing for global surveillance of meningococcal disease. *FEMS Microbiol Rev* 31, 15-26.

Broome, C. V. (1986). The carrier state: *Neisseria meningitidis* . *J Antimicr ob Chemother* 18 Suppl. A, 25-34.

Brueggemann, A. B., Pai, R., Crook, D. W. & Beall, B. (2007). Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. *PLoS Pathog* 3, e168.

Bruge, J., Bouveret-Le Cam, N., Danve, B., Rougon, G. & Schulz, D. (2004). Clinical evaluation of a group B meningococcal N-propionylated polysaccharide conjugate vaccine in adult, male volunteers. *Vaccine* 22, 1087-1096.

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

Callaghan, M. J., Jolley, K. A. & Maiden, M. C. (2006). Opacity-associated adhesin repertoire in hyperinvasive *Neisseria meningitidis*. *Infect Immun* 74, 5085-5094.

Callaghan, M. J., Buckee, C., McCarthy, N. D. & other authors (2008a). Opa protein repertoires of disease-causing and carried meningococci. *J Clin Micro biol* 46, 3033-3041.

Callaghan, M. J., Buckee, C. O., Jolley, K. A., Kriz, P., Maiden, M. C. & Gupta, S. (2008b). The effect of immune selection on the structure of the meningococcal opa protein repertoire. *PLoS Pathogens* 4, e1000020.

Campos, J., Fuste, M. C., Trujillo, G., Saez Nieto, J., Vazquez, J., Loren, J. G., Vinas, M. & Spratt, B. G. (1992). Genetic diversity of penicillin-resistant *Neisseria meningitidis*. *J Infect Dis* 166, 173-177.

Cano, R., Larrauri, A., Mateo, S., Alcala, B., Salcedo, C. & Vazquez, J. A. (2004). Impact of the meningococcal C conjugate vaccine in Spain: an epidemiological and microbiological decision. *Euro Surveill* 9, 11-15.

Cantini, F., Savino, S., Scarselli, M. & other authors (2006). Solution structure of the immunodominant domain of protective antigen GNA1870 of *Neisseria meningitidis*. *J Biol Chem* 281, 7220-7227.

Cantini, F., Veggi, D., Dragonetti, S., Savino, S., Scarselli, M., Romagnoli, G., Pizza, M., Banci, L. & Rappuoli, R. (2009). Solution structure of the factor H-binding protein, a survival factor and protective antigen of *Neisseria meningitidis*. *J Biol Chem* 284, 9022-9026.

Cartwright, K., Reilly, S., White, D. & Stuart, J. (1992). Early treatment with parenteral penicillin in meningococcal disease. *BMJ* 305, 143-147.

Cartwright, K. (1995). Introduction and historical aspects. In *Meningococcal disease*. Edited by K. Cartwright. Chichester, UK: John Wiley & Sons.

Cartwright, K. A. V., Stuart, J. M., Jones, D. M. & Noah, N. D. (1987). The Stonehouse survey: nasopharyngeal carriage of meningococci and *Neisseria lactamica*. *Epidemiol Infect* 99, 591-601.

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

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

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

Caugant, D. A., Frøholm, L. O., Bovre, K., Holten, E., Frasch, C. E., Mocca, L. F., Zollinger, W. D. & Selander, R. K. (1987a). Intercontinental spread of *Neisseria meningitidis* clones of the ET-5 complex. *Antonie Leeuwenhoek J Microbiol* 53, 389-394.

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

Caugant, D. A., Høiby, E. A., Magnus, P., Scheel, O., Hoel, T., Bjune, G., Wedege, E., Eng, J. & Frøholm, L. O. (1994). Asymptomatic carriage of *Neisseria meningitidis* in a randomly sampled population. *J Clin Microbiol* 32, 323-330.

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

Caugant, D. A. (2001). Global trends in meningococcal disease. In *Meningococcal Disease: m ethods and protocols* , pp. 273-294. Edited by A. J. Pollard & M. C. J. Maiden. Totowa, New Jersey: Humana Press.

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

Chain, E., Florey, H. W., Gardner, A. D., Heatley, N. G., Jennings, M. A., Orr-Ewing, J. & Sanders, A. G. (1940). Penicillin as a chemotherapeutic agent. . *Lancet* 24, 226-231.

Choo, S., Zuckerman, J., Goilav, C., Hatzmann, E., Everard, J. & Finn, A. (2000). Immunogenicity and reactogenicity of a group C meningococcal conjugate vaccine compared with a group A+C meningococcal polysaccharide vaccine in adolescents in a randomised observer-blind controlled trial. *Vaccine* 18, 2686-2692.

Claus, H., Vogel, U., Muhlenhoff, M., Gerardy-Schahn, R. & Frosch, M. (1997). Molecular divergence of the *sia* locus in different serogroups of *Neisseria meningitidis* expressing polysialic acid capsules. *Mol Gen Genet* 257, 28-34.

Claus, H., Stoevesandt, J., Frosch, M. & Vogel, U. (2001). Genetic isolation of meningococci of the electrophoretic type 37 complex. *J Bacteriol* 183, 2570-2575.

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

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

Claus, H., Elias, J., Meinhardt, C., Frosch, M. & Vogel, U. (2007). Deletion of the meningococcal fetA gene used for antigen sequence typing of invasive and commensal isolates from Germany: frequencies and mechanisms. *J Clin Microbiol* 45, 2960-2964.

Comanducci, M., Bambini, S., Brunelli, B. & other authors (2002). NadA, a novel vaccine candidate of *Neisseria meningitidis*. *J Exp Med* 195, 1445-1454.

Costantino, P., Viti, S., Podda, A., Velmonte, M. A., Nencioni, L. & Rappuoli, R. (1992). Development and phase 1 clinical testing of a conjugate vaccine against meningococcus A and C. *Vaccine* 10, 691-698.

de Moraes, J. C., Perkins, B. A., Camargo, M. C. & other authors (1992). Protective efficacy of a serogroup B meningococcal vaccine in Sao Paulo, Brazil. *Lancet* 340, 1074-1078.

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

Dehio, C., Gray-Owen, S. D. & Meyer, T. F. (1998). The role of neisserial Opa proteins in interactions with host cells. *Trends Microbiol* 6, 489-494.

Dellicour, S. & Greenwood, B. (2007). Systematic review: Impact of meningococcal vaccination on pharyngeal carriage of meningococci. *Tropical Medicine & International Health* 12, 1409-1421.

Dereeper, A., Guignon, V., Blanc, G. & other authors (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucl Acids Res* 36, W465-469.

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

Derrick, J. P., Heckels, J. E. & Virji, M. (2006). Major Outer Membrane Proteins of Meningococci. In *Handbook of Meningococcal Disease Infection Biology, Vaccination, Clinical Management*, pp. 181-215. Edited by M. Frosch & M. C. Maiden. Weinheim: Wiley-VCH.

Didelot, X. & Falush, D. (2007). Inference of bacterial microevolution using multilocus sequence data. *Genetics* 175, 1251-1266.

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

Diggle, M. A. & Clarke, S. C. (2005). Increased genetic diversity of *Neisseria meningitidis* isolates after the introduction of meningococcal serogroup C polysaccharide conjugate vaccines. *J Clin Microbiol* 43, 4649-4653.

Dyet, K., Devoy, A., McDowell, R. & Martin, D. (2005). New Zealand's epidemic of meningococcal disease described using molecular analysis: implications for vaccine delivery. *Vaccine* 23, 2228-2230.

Edwards, M. S. & Baker, C. J. (1981). Complications and sequelae of meningococcalinfections in children. *The Journal of Pediatrics* 99, 540-545.

Efron, A., Sorhouet Pereira, C., Salcedo, C., Vazquez, J. & Regueira, M. (2008).Significant increase of serogroup W135 invasive *Neisseria men ingitidis* strains in Argentina: A new epidemiological feature on the region. In *International Pathogenic Neisseria Conference*. Rotterdam, the Netherlands.

El Ahmer, O. R., Essery, S. D., Saadi, A. T., Raza, M. W., Ogilvie, M. M., Weir, D. M. & Blackwell, C. C. (1999). The effect of cigarette smoke on adherence of respiratory pathogens to buccal epithelial cells. *FEMS Immunol Med Microbiol* 23, 27-36.

Eldridge, J., Sutcliffe, E. M., Abbott, J. D. & Jones, D. M. (1978). Serological grouping of meningococci and detection of antigen in cerebrospinal fluid by coagglutination. *Medical Laboratory Science* 35, 63-66.

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

EU-IBIS (2006).Invasive *Neisseria meningitidis* in Europe 2003/2004: EU-IBIS.

EU-IBIS (2007). EU-IBIS Network. Invasive *Neisseria meningitidis* in Europe 2006. Health Protection Agency, London.

EU-IBIS, Chandra, M. & Ramsay, M. (2007).EU-IBIS: Surveillance and epidemiology of invasive meningococcal disease in Europe 1999 – 2006. Presentation given at the 9th EMGM, Rome, Italy: EU-IBIS.

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

Excoffier, L., Laval, G. & Schneider, S. (2005). ARLEQUIN (version 3.0): An integrated software package for population genetics data analysis. *Evol Bioinform Online* 1, 47-50.

Feavers, I. M. & Maiden, M. C. J. (1998). A gonococcal *porA* pseudogene: implications for understanding the evolution and pathogenicity of *Neisseria gonorrhoeae* . *Mol Microbiol* 30, 647-656.

Feavers, I. M. (2001). Meningococcal vaccines and vaccine developments. In *Meningococcal Vaccines*, pp. 1-22. Edited by A. J. Pollard & M. C. J. Maiden. Totowa, NJ: Humana Press.

Feil, E., Carpenter, G. & Spratt, B. G. (1995). Electrophoretic variation in adenylate kinase of *Neisseria meningitidis* is due to inter- and intraspecies recombination. *Proc Natl Acad Sci USA* 92, 10535-10539.

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

Feil, E. J., Holmes, E. C., Bessen, D. E. & other authors (2001). Recombination within natural populations of pathogenic bacteria: Short- term empirical estimates and long-term phylogenetic consequences. *Proc Natl Acad Sci USA* 98, 182-187.

Feil, E. J. & Spratt, B. G. (2001). Recombination and the population structures of bacterial pathogens. *Annu Rev Microbiol* 55, 561-590.

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

Findlow, J., Lowe, A., Deane, S., Balmer, P., van den Dobbelsteen, G., Dawson, M., Andrews, N. & Borrow, R. (2005). Effect of sequence variation in meningococcal PorA outer membrane protein on the effectiveness of a hexavalent PorA outer membrane vesicle vaccine in toddlers and school children. *Vaccine* 23, 2623-2627.

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

Finne, J., Bitter Suermann, D., Goridis, C. & Finne, U. (1987). An IgG monoclonal antibody to group B meningococci cross-reacts with developmentally regulated polysialic acid units of glycoproteins in neural and extraneural tissues. *J Immunol* 138, 4402-4407.

Finney, M., Vaughan, T., Taylor, S. & other authors (2008). Characterization of the key antigenic components and pre-clinical immune responses to a meningococcal disease vaccine based on *Neisseria lactamica* outer membrane vesicles. *Hum Vaccin* 4, 23-30.

Fletcher, L. D., Bernfield, L., Barniak, V. & other authors (2004). Vaccine potential of the *Neisseria meningitidis* 2086 lipoprotein. *Infect Immun* 72, 2088-2100.

Frasch, C. E. & Chapman, S. (1972). Classification of *Neisseria meningitidis* Group B into Distinct Serotypes. *Infect Immun* 5, 98-102.

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

Frasch, C. E. (1995). Meningococcal vaccines: past, present, and future. In *Meningococcal disease* , pp. 246-283. Edited by K. Cartwright. Chichester, England: John Wiley & Sons.

Frasch, C. E., van Alphen, L., Poolman, J. T. & Rosenqvist, E. (2001). Outer membrane protein vesicle vaccines for meningococcal disease. In *Meningococcal vaccines*, pp. 81-108. Edited by A. J. Pollard & M. C. J. Maiden. Totowa, New Jersey: Humana press.

Fraser, C., Hanage, W. P. & Spratt, B. G. (2005). Neutral microepidemic evolution of bacterial pathogens. *Proc Natl Acad Sci USA* 102, 1968-1973.

Frosch, M., Weisgerber, C. & Meyer, T. F. (1989). Molecular characterization and expression in *Escherichia coli* of the gene complex encoding the polysaccharide capsule of *Neisseria meningitidis* group B. *Proc Natl Acad Sci USA* 86, 1669-1673.

Frosch, M. & Meyer, T. F. (1992). Transformation-mediated exchange of virulence determinants by co-cultivation of pathogenic *Neisseriae*. *FEMS Micro biol Lett* 100, 345-349.

Frosch, M. & Maiden, M. C. (2006). Introduction: The meningococcus, an enigmatic pathogen. In *Handbook of meningococcal disease*, pp. 21. Edited by M. Frosch & M. C. Maiden. Weinheim: Wiley-VCH.

Frosch, M. & Vogel, U. (2006). Structure and genetics of the meningococcal capsule. In *Handbook of meningococcal disease* , pp. 145-162. Edited by M. Frosch & M. C. Maiden. Weinheim, Germany: Wiley-VCH.

Giuliani, M. M., Santini, L., Brunelli, B. & other authors (2005). The region comprising amino acids 100 to 255 of *Neisseria menin gitidis* lipoprotein GNA 1870 elicits bactericidal antibodies. *Infect Immun* 73, 1151-1160.

Giuliani, M. M., Adu-Bobie, J., Comanducci, M. & other authors (2006). A universal vaccine for serogroup B meningococcus. *Proc Natl Acad Sci USA* 103, 10834-10839.

Go, M. F., Kapur, V., Graham, D. Y. & Musser, J. M. (1996). Population genetic analysis of *Helicobacter pylori* by multilocus enzyme electrophoresis: extensive allelic diversity and recombinational population structure. *J Bacteriol* 178, 3934-3938.

Gold, R. & Wyle, F. A. (1970). New classification of *Neisseria meningitidis* by means of bactericidal reactions. *Infect Immun* 1, 479-484.

Gold, R., Goldschneider, I., Lepow, M. L., Draper, T. F. & Randolph, M. (1978). Carriage of *Neisseria m eningitidis* and *Neisseria lactamica* in infants and children. *J Infect Dis* 137, 112-121.

Goldacre, M. J., Roberts, S. E. & Yeates, D. (2003). Case fatality rates for meningococcal disease in an English population, 1963-98: database study. *BMJ* 327, 596-597.

Goldschneider, I., Gotschlich, E. C. & Artenstein, M. S. (1969a). Human Immunity to the Meningococcus. II. Development of Natural Immunity. *J Exp Med* 129, 1327-1348.

Goldschneider, I., Gotschlich, E. C. & Artenstein, M. S. (1969b). Human Immunity to the Meningococcus. I. The Role of Humoral Antibodies. *J Exp Med* 129, 1307-1326.

Gorringe, A., Halliwell, D., Matheson, M., Reddin, K., Finney, M. & Hudson, M. (2005). The development of a meningococcal disease vaccine based on *Neisseria lactamica* outer membrane vesicles. *Vaccine* 23, 2210-2213.

Gorringe, A. R., Taylor, S., Brookes, C. & other authors (2009). Phase I safety and immunogenicity study of a candidate meningococcal disease vaccine based on *Neisseria lactamica* outer membrane vesicles. *Clin Vaccine Immunol* 16, 1113-1120.

Gotschlich, E. C., Goldschneider, I. & Artenstein, M. S. (1969). Human immunity to the meningococcus IV. Immunogenicity of group A and group C meningococcal polysaccharides. *J Exp Med* 129, 1367-1384.

Gottfredsson, M., Diggle, M. A., Lawrie, D. I., Erlensdottir, H., Hardardottir, H., Kristinsson, K. G. & Clarke, S. (2006). *Neisseria meningitidis* sequence type and risk for death, Iceland. *Emerg Infect Dis* 12, 1066-1073.

Granoff, D. M., Welsch, J. A. & Ram, S. (2009). Binding of complement factor H (fH) to *Neisseria meningitidis* is specific for human fH and inhibits complement activation by rat and rabbit sera. *Infect Immun* 77, 764-769.

Gray, S. J., Sobanski, M. A., Kaczmarski, E. B., Guiver, M., Marsh, W. J., Borrow, R., Barnes, R. A. & Coakley, W. T. (1999). Ultrasound-enhanced latex immunoagglutination and PCR as complementary methods for non-culture-based confirmation of meningococcal disease. *J Clin Microbiol* 37, 1797-1801.

Gray, S. J., Borrow, R. & Kaczmarski, E. B. (2001). Meningococcal serology. In *Meningococcal Disease*, pp. 61-87. Edited by A. J. Pollard & M. C. J. Maiden. Totowa, NJ: Humana Press.

Greenwood, B. (2006). Editorial: 100 years of epidemic meningitis in West Africa - has anything changed? *Trop Med Int Health* 11, 773-780.

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

Greenwood, B. M., Blakebrough, I. S., Bradley, A. K., Wali, S. & Whittle, H. C. (1984). Meningococcal disease and season in sub-Saharan Africa. *Lancet* 1, 1339-1342.

Griffiss, J. M. (1995). Mechanisms of host immunity. In *Meningococcal disease*, pp. 35-70. Edited by K. Cartwright. Chichester: John Wiley & Sons Ltd.

Grundmann, H., Hori, S. & Tanner, G. (2001). Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. *J Clin Microbiol* 39, 4190-4192.

Guindon, S. & Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Syst Biol* 52, 696-704.

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

Gupta, S. & Maiden, M. C. J. (2001). Exploring the evolution of diversity in pathogen populations. *Trends Microbiol* 9, 181-192.

Guttormsen, H. K., Wetzler, L. M. & Naess, A. (1993). Humoral immune response to the class 3 outer membrane protein during the course of meningococcal disease. *Infect Immun* 61, 4734-4742.

Hammerschmidt, S., Birkholz, C., Zahringer, U., Robertson, B. D., van Putten, J., Ebeling, O. & Frosch, M. (1994). Contribution of genes from the capsule gene complex (*cps*) to lipooligosaccharide biosynthesis and serum resistance in *Neisseria meningitidis*. *Mol Microbiol* 11, 885-896.

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

Haralambous, E., Dolly, S. O., Hibberd, M. L., Litt, D. J., Udalova, I. A., O'Dwyer, C., Langford, P. R., Kroll, J. S. & Levin, M. (2006). Factor H, a regulator of complement activity, is a major determinant of meningococcal disease susceptibility in UK Caucasian patients. *Scand J Infect Dis* 38, 764 - 771.

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

Harrison, L. H., Trotter, C. L. & Ramsay, M. E. (2009). Global epidemiology of meningococcal disease. *Vaccine* 27, B51-B63.

Harrison, O. B., Maiden, M. C. & Rokbi, B. (2008). Distribution of transferrin binding protein B gene (tbpB) variants among *Neisseria* species. *BMC Microbiology* 8, 66.

Hassan-King, M. K. A., Wall, R. A. & Greenwood, B. M. (1988). Meningococcal carriage, meningococcal disease and vaccination. *J Infect* 16, 55-59.

Heath, P. T. & McVernon, J. (2002). The UK Hib vaccine experience. *Arch Dis Child* 86, 396-399.

Heckels, J. E. (1989). Structure and function of pili of pathogenic *Neisseria* species. *Clin Microbiol Rev* 2S, S66-S73.

Heckenberg, S. G. B., de Gans, J., Brouwer, M. C., Weisfelt, M., Piet, J. R., Spanjaard, L., van der Ende, A. & van de Beek, D. (2008). Clinical features, outcome, and meningococcal genotype in 258 adults with meningococcal meningitis - A prospective cohort study. *Medicine (Baltimore)* 87, 185-192.

Hitchcock, P. J. (1989). Unified nomenclature for pathogenic *Neisseria* species. *Clin Microbiol Rev* 2, S64-S65.

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

Hudson, R. R., Slatkin, M. & Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics* 132, 583-589.

Hudson, R. R. (1994). Analytical results concerning linkage disequilibrium in models with genetic transformation and conjugation. *J Evol Biol* 7, 535-548.

Hunter, P. R. & Gaston, M. A. (1988). Numerical index of discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol* 26, 2465-2466.

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

Jacobsson, S., Thulin, S., Molling, P., Unemo, M., Comanducci, M., Rappuoli, R. & Olcen, P. (2006). Sequence constancies and variations in genes encoding three new meningococcal vaccine candidate antigens. *Vaccine* 24, 2161-2168.

Jansen, C., Wiese, A., Reubsaet, L., Dekker, N., de Cock, H., Seydel, U. & Tommassen, J. (2000). Biochemical and biophysical characterization of in vitro folded outer membrane porin PorA of *Neisseria meningitidis*. *Biochim Biophys Acta* 1464, 284-298.

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

Jelfs, J., Munro, R., Wedege, E. & Caugant, D. A. (2000b). Sequence variation in the *porA* gene of a clone of *Neisseria meningitidis* during epidemic spread. *Clin Diagn Lab Immunol* 7, 390-395.

Jennings, H. J. & Lugowski, C. (1981). Immunochemistry of groups A, B, and C meningococcal polysaccharide-tetanus toxoid conjugates. *J Immunol* 127, 1011-1018.

Jennings, H. J., Roy, R. & Gamian, A. (1986). Induction of meningococcal group B polysaccharide-specific IgG antibodies in mice by using an N-propionylated B polysaccharide-tetanus toxoid conjugate vaccine. *J Immunol* 137, 1708-1713.

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

Jodar, L., LaForce, F. M., Ceccarini, C., Aguado, T. & Granoff, D. M. (2003). Meningococcal conjugate vaccine for Africa: a model for development of new vaccines for the poorest countries. *Lancet* 361, 1902-1904.

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

Jolley, K. A., Feil, E. J., Chan, M. S. & Maiden, M. C. (2001). Sequence type analysis and recombinational tests (START). *Bioinformatics* 17, 1230-1231.

Jolley, K. A. & Urwin, R. (2001). Computational methods for meningococcal population studies. In *Methods in Molecular Medicine, vol 67: Meningococcal Disease: Methods and Protocols* . Edited by A. J. Pollard & M. C. J. Maiden. Totowa, New Jersey: Humana Press.

Jolley, K. A., Kalmusova, J., Feil, E. J., Gupta, S., Musilek, M., Kriz, P. & Maiden, M. C. (2002). Carried Meningococci in the Czech Republic: a Diverse Recombining Population. *J Clin Microbiol* 40, 3549-3550.

Jolley, K. A., Chan, M. S. & Maiden, M. C. (2004). mlstdbNet - distributed multi-locus sequence typing (MLST) databases. *BMC Bioinformatics* 5, 86.

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

Jolley, K. A. & Maiden, M. C. (2006). AgdbNet - antigen sequence database software for bacterial typing. *BMC Bioinformatics* 7, 314.

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

Jones, D. M. (1995). Epidemiology of meningococcal disease in Europe and the USA. In *Meningococcal disease*, pp. 147-157. Edited by K. Cartwright. Chichester, England: John Wiley and Sons Ltd.

Kaczmarski, E. B., Ragunathan, P. L., Marsh, J., Gray, S. J. & Guiver, M. (1998). Creating a national service for the diagnosis of meningococcal disease by polymerase chain reaction. *Commun Dis Public Health* 1, 54-56.

Kafetzis, D. A., Stamboulidis, K. N., Tzanakaki, G., Kourea Kremastinou, J., Skevaki, C. L., Konstantopoulos, A. & Tsolia, M. (2007). Meningococcal group C disease in Greece during 1993-2006: the impact of an unofficial single-dose vaccination scheme adopted by most paediatricians. *Clinical Microbiology and Infection* 13, 550-552.

Kirkwood, B. R. & Sterne, J. A. C. (2003). *Essentia l M edical Sta tistics*, 2nd edn. Oxford, UK: Blackwell Science.

Kirsch, E. A., Barton, R. P., Kitchen, L. & Giroir, B. P. (1996). Pathophysiology, treatment and outcome of meningococcemia: a review and recent experience. *Pediatr Infect Dis J* 15, 967-978; quiz 979.

Koroleva, I. S., Platonov, A. E., van der Ende, A., Kuijper, E. & Dankert, J. (1998). Characteristics of pathogenic *Neisseria mening itidis* in Moscow: prevalence of 'non-European' strains. *Clinical Microbiology and Infection* 4, 123-128.

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

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

Kremastinou, J., Tzanakaki, G., Levidiotou, S., Markou, F., Themeli, E., Voyiatzi, A., Psoma, E., Theodoridou, M. & Blackwell, C. C. (2003). Carriage of *Neisseria meningitidis* and *Neisseria lactam ica* in northern Greece. *FEMS Immunol Med Microbiol* 39, 23-29.

Kristiansen, B. E., Ask, E., Jenkins, A., Fermer, C., Radstrom, P. & Sköld, O. (1991). Rapid diagnosis of meningococcal meningitis by polymerase chain reaction. *Lancet* 337, 1568-1569.

Kuipers, B., van den Dobbelsteen, G., Wedege, E. & van Alphen, L. (2001). Serological Characterization. In *Meningococcal Disease: methods and protocols* , pp. 131-145. Edited by A. J. Pollard & M. C. Maiden. Totowa, New Jersey: Humana Press.

Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5, 150-163.

LaForce, F. M., Konde, K., Viviani, S. & Preziosi, M. P. (2007). The Meningitis Vaccine Project. *Vaccine* 25, A97-A100.

Lambris, J. D., Ricklin, D. & Geisbrecht, B. V. (2008). Complement evasion by human pathogens. 6, 132-142.

Lapeyssonnie, L. (1963). La méningite cérébrospinale en Afrique. *BullWorld He alth Organ* 28 (suppl), 53-114.

Law, D. K., Lorange, M., Ringuette, L. & other authors (2006). Invasive meningococcal disease in Quebec, Canada, due to an emerging clone of ST-269 serogroup B meningococci with serotype antigen 17 and serosubtype antigen P1.19 (B:17:P1.19). *J Clin Microbiol* 44, 2743-2749.

Levin, B. R. (1981). Periodic selection, infectious gene exchange and the genetic structure of *E. coli* populations. *Genetics* 99, 1-23.

Levin, B. R. & Bergstrom, C. T. (2000). Bacteria are different: observations, interpretations, speculations, and opinions about the mechanisms of adaptive evolution in prokaryotes. *Proc Natl Acad Sci USA* 97, 6981-6985.

Linz, B., Schenker, M., Zhu, P. & Achtman, M. (2000). Frequent interspecific genetic exchange between commensal *neisseriae* and *Neisseria meningitidis*. *Mol Microbiol* 36, 1049-1058.

Lorenz, M. G. & Wackernagel, W. (1994). Bacterial Gene Transfer by Natural Genetic Transformation in the Environment. *Microbiol Rev* 58, 563-602.

Lucidarme, J., Comanducci, M., Findlow, J. & other authors (2009). Characterisation of fHbp, nhba (gna2132), nadA, porA, Sequence Type and the genomic presence of IS1301 in group B meningococcal ST269 clonal complex case-isolates from England and Wales. *J Clin Microbiol* 47, 3577-3585.

Luijkx, T. A., van Dijken, H., Hamstra, H. J., Kuipers, B., van der Ley, P., van Alphen, L. & van den Dobbelsteen, G. (2003). Relative immunogenicity of PorA Subtypes in a

multivalent *Neisseria meningitidis* vaccine is not dependent on presentation form. *Infect Immun* 71, 6367-6371.

MacLennan, J., Kafatos, G., Neal, K. & other authors (2006). Social behavior and meningococcal carriage in British teenagers. *Emerg Infect Dis* 12, 950-957.

Madico, G., Welsch, J. A., Lewis, L. A. & other authors (2006). The meningococcal vaccine candidate GNA1870 binds the complement regulatory protein factor H and enhances serum resistance. *J Immunol* 177, 501-510.

Maiden, M. C., Stuart, J. M. & Group, U. M. C. (2002). Carriage of serogroup C meningococci 1 year after meningococcal C conjugate polysaccharide vaccination. *Lancet* 359, 1829-1831.

Maiden, M. C. (2006). Multilocus Sequence Typing of Bacteria. *Annu Rev Micro biol* 60, 561-588.

Maiden, M. C. & Urwin, R. (2006). Population dynamics of bacterial pathogens. In *Evolution of microbial pathogens*, pp. 35-53. Edited by H. S. Seifert. Washington D.C.: ASM Press.

Maiden, M. C. (2007).Molecular typing of meningococci: defining an elusive pathogen. In *EMGM*. Rome, Italy.

Maiden, M. C., Ibarz-Pavon, A. B., Urwin, R. & other authors (2008). Impact of Meningococcal Serogroup C Conjugate Vaccines on Carriage and Herd Immunity. *J Infect Dis* 197, 737-743.

Maiden, M. C. J. (1993). Population genetics of a transformable bacterium: the influence of horizontal genetical exchange on the biology of *Neisseria meningitid is*. *FEMS Microbiol Lett* 112, 243-250.

Maiden, M. C. J. & Feavers, I. M. (1995). Population genetics and global epidemiology of the human pathogen *Neisseria meningitidis*. In *Population genetics of bacteria* , pp. 269-293. Edited by S. Baumberg, J. P. W. Young, E. M. H. Wellington & J. R. Saunders. Cambridge, England: Cambridge University Press.
Maiden, M. C. J., Bygraves, J. A., Feil, E. & other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 95, 3140-3145.

Makela, P. H., Kayhty, H., Weckstrom, P., Sivonen, A. & Renkonen, O. V. (1975). Effect of group-A meningococcal vaccine in army recruits in Finland. *Lancet* 2, 883-886.

Mandrell, R. E. & Zollinger, W. D. (1977). Lipopolysaccharide serotyping of *Neisseria meningitidis* by hemagglutination inhibition. *Infect Immun* 16, 471-475.

Marsh, J. W., O'Leary, M. M., Shutt, K. A. & Harrison, L. H. (2007). Deletion of *fetA* gene sequences in serogroup B and C *Neisseria meningitidis* isolates. *J Clin Microb iol* 45, 1333-1335.

Martin, D., Cadieux, N., Hamel, J. & Brodeur, B. R. (1997). Highly conserved *Neisseria meningitidis* surface protein confers protection against experimental infection. *J Exp Med* 185, 1173-1183.

Martin, D., Brodeur, B. R., Hamel, J., Couture, F., de Alwis, U., Lian, Z., Martin, S., Andrews, D. & Ellis, R. W. (2000a). Candidate *Neisseria meningitidis* NspA vaccine. *J Biotechnol* 83, 27-31.

Martin, S. L., Borrow, R., van der Ley, P., Dawson, M., Fox, A. J. & Cartwright, K. A. (2000b). Effect of sequence variation in meningococcal PorA outer membrane protein on the effectiveness of a hexavalent PorA outer membrane vesicle vaccine. *Vaccine* 18, 2476-2481.

Mascioni, A., Bentley, B. E., Camarda, R. & other authors (2008). Structural basis for the immunogenic properties of the meningococcal vaccine candidate LP2086. *J Biol Chem* 18, 2476-2481.

Masignani, V., Comanducci, M., Giuliani, M. M. & other authors (2003). Vaccination against *Neisseria meningitidis* using three variants of the lipoprotein GNA1870. *J Exp Med* 197, 789-799.

Maynard Smith, J. (1989). Trees, bundles or nets. *Trends Ecol Evol* 4, 302-304.

Maynard Smith, J., Dowson, C. G. & Spratt, B. G. (1991). Localized sex in bacteria. *Nature* 349, 29-31.

Maynard Smith, J. (1992). Analysing the mosaic structure of genes. *J Mol Evol* 34, 126-129.

Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proc Natl Acad Sci USA* 90, 4384-4388.

McCarthy, N. D., Colles, F. M., Dingle, K. E., Bagnall, M. C., Manning, G., Maiden, M. C. & Falush, D. (2007). Host-associated genetic import in *Campylobacter jejuni*. *Emerg Infect Dis* 13, 267-272.

McEllistrem, M. C., Kolano, J. A., Pass, M. A., Caugant, D. A., Mendelsohn, A. B., Fonseca Pacheco, A. G., Shutt, K. A., Razeq, J. & Harrison, L. H. (2004). Correlating epidemiologic trends with the genotypes causing meningococcal disease, Maryland. *Emerg Infect Dis* 10, 451-456.

McGuinness, B., Barlow, A. K., Clarke, I. N., Farley, J. E., Anilionis, A., Poolman, J. T. & Heckels, J. E. (1990). Deduced amino acid sequences of class 1 protein (PorA) from three strains of *Neisseria menin gitidis.* Synthetic peptides define the epitopes responsible for serosubtype specificity. *J Exp Med* 171, 1871-1882.

McGuinness, B. T., Clarke, I. N., Lambden, P. R., Barlow, A. K., Poolman, J. T., Jones, D. M. & Heckels, J. E. (1991). Point mutation in meningococcal *porA* gene associated with increased endemic disease. *Lancet* 337, 514-517.

McIntyre, P. (2005). Should dexamethasone be part of routine therapy of bacterial meningitis in industrialised countries? *Adv Exp Med Biol* 568, 189-197.

Merz, A. J. & So, M. (2000). Interactions of pathogenic *Neisseriae* with epithelial cell membranes. *Annu Rev Cell Dev Biol* 16, 423-457.

Michaelsen, T. E., Aase, A., Kolberg, J., Wedege, E. & Rosenqvist, E. (2001). PorB3 outer membrane protein on *Neisseria meningitidis* is poorly accessible for antibody binding on live bacteria. *Vaccine* 19, 1526-1533.

Milagres, L. G., Ramos, S. R., Sacchi, C. T., Melles, C. E. A., Vieira, V. S. D., Sato, H., Brito, G. S., Moraes, J. C. & Frasch, C. E. (1994). Immune response of Brazilian children to a *Neisseria meningitid is* serogroup B outer membrane protein vaccine: Comparison with efficacy. *Infect Immun* 62, 4419-4424.

Millar, J. W., Siess, E. E., Feldman, H. A., Silverman, C. & Frank, P. (1963). *In vivo* and *in vitro* resistance to sulfadiazine in strains of *Neisseria meningitidis*. *Jama* 186, 139-141.

Miller, E., Salisbury, D. & Ramsay, M. (2001). Planning, registration, and implementation of an immunisation campaign against meningococcal serogroup C disease in the UK: a success story. *Vaccine* 20, S58-67.

Moe, G. R., Tan, S. & Granoff, D. M. (1999). Differences in surface expression of NspA among *Neisseria meningitidis* group B strains. *Infect Immun* 67, 5664-5675.

Molesworth, A. M., Thomson, M. C., Connor, S. J., Cresswell, M. P., Morse, A. P., Shears, P., Hart, C. A. & Cuevas, L. E. (2002). Where is the meningitis belt? Defining an area at risk of epidemic meningitis in Africa. *Trans R Soc Trop Med Hyg* 96, 242-249.

Molesworth, A. M., Cuevas, L. E., Connor, S. J., Morse, A. P. & Thomson, M. C. (2003). Environmental risk and meningitis epidemics in Africa. *Emerg Infect Dis* 9, 1287-1293.

Moore, P. S., Reeves, M. W., Schwartz, B., Gellin, B. G. & Broome, C. V. (1989). Intercontinental spread of an epidemic group A *Neisseria meningitidis* strain. *Lancet* ii, 260-262.

Morand, P. E. & Rudel, T. (2006). Genetics, structure and function of pili. In *Handbook of meningococcal disease*, pp. 235-254. Edited by M. Frosch & M. C. Maiden. Weinheim, Germany: Wiley-VCH.

Morelli, G., Malorny, B., Muller, K., Seiler, A., Wang, J. F., del Valle, J. & Achtman, M. (1997). Clonal descent and microevolution of *Neisseria meningitidis* during 30 years of epidemic spread. *Mol Microbiol* 25, 1047-1064.

Morley, S. L. & Pollard, A. J. (2001). Vaccine prevention of meningococcal disease, coming soon? *Vaccine* 20, 666-687.

Morse, S. A. (1996). *Neisseria*, *Moraxella*, *Kingella* and *Eikenella* . In *Medical Microbiology*. Edited by S. Baron. Galveston: The University of Texas Medical Branch at Galveston.

Moulton, V. (2003). SplitsTree: A network-based tool for exploring evolutionary relationships in molecular data. In *The Phylogenetic Handbook: A Practical Approach to DNA and Protein Phylogeny*, pp. 312-328. Edited by M. Salemi & A. M. Vandamme. Cambridge, UK: Cambridge University Press.

Muros-Le Rouzic, E., Doemland, M., Gilmet, G. & Teyssou, R. (2006). Geographical differences in invasive meningoccal disease rates and serogroup distribution. *Poster presented at International Pathogenic Neisseria Conference, 2006; Cairns, Queensland, Australia*.

Murphy, E., Andrew, L., Lee, K. & other authors (2009). Sequence diversity of the factor H-binding protein vaccine candidate in epidemiologically relevant strains of serogroup B *Neisseria meningitidis*. *J Infect Dis* 200, 379-389.

Muzzi, A., Masignani, V. & Rappuoli, R. (2007). The pan-genome: towards a knowledge-based discovery of novel targets for vaccines and antibacterials. *Drug Discovery Today* 12, 429-439.

Muzzi, A., Jolley, K., Bambini, S., Comandi, S., Rappuoli, R., Pizza, M., Maiden, M. C. J. & Comanducci, M. (2009).Variability of the antigens of the Novartis investigational MenB vaccine in a carrier/disease strain panel. In *European Monitoring Group for Meningococci (EMGM)*. Manchester, UK.

Nassif, X., Bourdoulous, S., Eugene, E. & Couraud, P. O. (2002). How do extracellular pathogens cross the blood-brain barrier? *Trends Microbiol* 10, 227-232.

Neal, K. R. (2000). Changing carriage rate of *Neisseria meningitidis* among university students during the first week of term: cross sectional study. *BMJ* 320, 846-849.

Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3, 418-426.

Newcombe, J., Cartwright, K. A. V., Palmer, W. H. & McFadden, J. (1996). Polymerase chain reaction of peripheral blood for the diagnosis of meningococcal disease. *J Clin Microbiol* 34, 1637-1640.

Newcombe, J., Cartwright, K., Dyer, S. & McFadden, J. (1998). Naturally occurring insertional inactivation of the porA gene of *Neisseria meningitidis* by integration of IS1301. *Mol Microbiol* 30, 453-454.

Ngampasutadol, J., Ram, S., Gulati, S., Agarwal, S., Li, C., Visintin, A., Monks, B., Madico, G. & Rice, P. A. (2008). Human factor H interacts selectively with *Neisseria gonorrhoeae* and results in species-specific complement evasion. *J Immunol* 180, 3426-3435.

Ni, H., Knight, A. I., Cartwright, K., Palmer, W. H. & McFadden, J. (1992). Polymerase chain reaction for diagnosis of meningococcal meningitis. *Lancet* 340, 1432-1434.

Nicolas, P., Decousset, L., Riglet, V., Castelli, P., Stor, R. & Blanchet, G. (2001). Clonal expansion of sequence type (ST-)5 and emergence of ST-7 in serogroup A meningococci, Africa. *Emerg Infect Dis* 7, 849-854.

Nicolas, P., Norheim, G., Garnotel, E., Djibo, S. & Caugant, D. A. (2005). Molecular epidemiology of *Neisseria meningitidis* isolated in the African Meningitis Belt between 1988 and 2003 shows dominance of sequence type 5 (ST-5) and ST-11 complexes. *J Clin Microbiol* 43, 5129-5135.

Nicolas, P., Agbenoko, K., Badziklou, K. & other authors (2008).Dominance of the sequence type (ST-)5 complex in 5 countries of the African Meningitis Belt between 2004 and 2007. In *International Pathogenic Neisseria Conference*. Rotterdam, the Netherlands.

Njanpop-Lafourcade, B. M., Tamekloe, T. A., Sanou, O. & other authors (2008).Serogroup X meningococcal meningitis in Togo during 2007 and 2008. In *International Pathogenic Neisseria Conference*. Rotterdam, the Netherlands.

Norheim, G., Rosenqvist, E., Aseffa, A. & other authors (2006). Characterization of *Neisseria meningitidis* isolates from recent outbreaks in Ethiopia and comparison with those recovered during the epidemic of 1988 to 1989. *J Clin Microbiol* 44, 861-871.

Norheim, G., Rosenqvist, E., Hoiby, E. A. & Caugant, D. A. (2008).Comparison of serogroup A ST-2859 emerging in Burkina Faso in 2003 and 2004 with predominating ST-7 strains. In *International Pathogenic Neisseria Conference*. Rotterdam, the Netherlands.

O'Hallahan, J., Lennon, D., Oster, P. & other authors (2005). From secondary prevention to primary prevention: a unique strategy that gives hope to a country ravaged by meningococcal disease. *Vaccine* 23, 2197-2201.

O'Rourke, M. & Spratt, B. G. (1994). Further evidence for the non-clonal population structure of *Neisseria gonorrhoeae*: extensive genetic diversity within isolates of the same electrophoretic type. *Microbiology* 140, 1285-1290.

Oliver, K. J., Reddin, K. M., Bracegirdle, P., Hudson, M. J., Borrow, R., Feavers, I. M., Robinson, A., Cartwright, K. & Gorringe, A. R. (2002). *Neisseria lactamica* protects against experimental meningococcal infection. *Infect Immun* 70, 3621-3626.

Olyhoek, A. J., Sarkari, J., Bopp, M., Morelli, G. & Achtman, M. (1991). Cloning and expression in *Escherichia coli* of *opc*, the gene for an unusual class 5 outer membrane protein from *Neisseria meningitidis*. *Microb Pathogen* 11, 249-257.

Olyhoek, T., Crowe, B. A. & Achtman, M. (1987). Clonal population structure of *Neisseria meningitidis* serogroup A isolated from epidemics and pandemics between 1915 and 1983. *Rev Infect Dis* 9, 665-682.

Oster, P., Lennon, D., O'Hallahan, J., Mulholland, K., Reid, S. & Martin, D. (2005). MeNZB: a safe and highly immunogenic tailor-made vaccine against the New Zealand *Neisseria meningitidis* serogroup B disease epidemic strain. *Vaccine* 23, 2191-2196.

Pajon, R., Beernink, P. T., Harrison, L. H. & Granoff, D. M. (2010). Frequency of factor H-binding protein modular groups and susceptibility to cross-reactive bactericidal activity in invasive meningococcal isolates. *Vaccine* 28, 2122-2129.

Parkhill, J., Achtman, M., James, K. D. & other authors (2000). Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* 404, 502-506.

Pecoul, B., Varaine, F., Keita, M., Soga, G., Djibo, A., Soula, G., Abdou, A., Etienne, J. & Rey, M. (1991). Long-acting chloramphenicol versus intravenous ampicillin for treatment of bacterial meningitis. *Lancet* 338, 862-866.

Peltola, H., Jonsdottir, K., Lystad, A., Sievers, C. J. & Kallings, I. (1982). Meningococcal disease in Scandinavia. *BMJ* 284, 1618-1621.

Perea, W. (2007).Meningococcal meningitis in Africa: overview and current challenges. Presentation at the International Workshop on *Neisseria* Vaccines, Varadero, Cuba.: WHO/CDS.

Perez-Trallero, E., Vicente, D., Montes, M. & Cisterna, R. (2002). Positive effect of meningococcal C vaccination on serogroup replacement in *Neisseria meningitidis* . *Lancet* 360, 953.

Petering, H., Hammerschmidt, S., Frosch, M., van Putten, J. P., Ison, C. A. & Robertson, B. D. (1996). Genes associated with meningococcal capsule complex are also found in *Neisseria gonorrhoeae*. *J Bacteriol* 178, 3342-3345.

Pettersson, A., Kuipers, B., Pelzer, M., Verhagen, E., Tiesjema, R. H., Tommassen, J. & Poolman, J. T. (1990). Monoclonal antibodies against the 70-kilodalton iron-regulated protein of *Neisseria meningitidis* are bactericidal and strain specific. *Infect Immun* 58, 3036-3041.

Pettersson, A., Maas, A., van Wassenaar, D., van der Ley, P. & Tommassen, J. (1995). Molecular characterization of FrpB, the 70-kilodalton iron-regulated outer membrane protein of *Neisseria meningitidis*. *Infect Immun* 63, 4181-4184.

Pizza, M., Scarlato, V., Masignani, V. & other authors (2000). Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287, 1816-1820.

Plested, J. S., Harris, S. L., Wright, J. C. & other authors (2003). Highly conserved *Neisseria meningitidis* inner-core lipopolysaccharide epitope confers protection against experimental meningococcal bacteremia. *J Infect Dis* 187, 1223-1234.
Pollard, A. J. & Frasch, C. (2001). Development of natural immunity to *Neisseria meningitidis*. *Vaccine* 19, 1327-1346.

Pollard, A. J., Scheifele, D. & Rosenstein, N. (2001). Epidemiology of Meningococcal Disease in North America. In *Meningococcal Disease: Methods and Protocols*, pp. 341-356. Edited by A. J. Pollard & M. C. Maiden. Totowa, New Jersey: Humana Press.

Poolman, J. & Berthet, F. X. (2001). Alternative vaccine strategies to prevent serogroup B meningococcal diseases. *Vaccine* 20, S24-S26.

Poolman, J. T., De Marie, S. & Zanen, H. C. (1980). Variability of low-molecular-weight, heat-modifiable outer membrane proteins of *Neisseria meningitidis* . *Infect Immun* 30, 642-648.

Poolman, J. T. & Abdillahi, H. (1988). Outer membrane protein serotyping of *Neisseria meningitidis*. *Eur J Clin Microbiol Infect Dis* 7, 291-292.

Prinz, T. & Tommassen, J. (2000). Association of iron-regulated outer membrane proteins of *Neisseria meningitidis* with the RmpM (class 4) protein. *FEMS Microbiol Lett* 183, 49-53.

Pullinger, G. D., Coffey, T. J., Maiden, M. C. & Leigh, J. A. (2007). Multilocus-sequence typing analysis reveals similar populations of *Streptococcus uberis* are responsible for bovine intramammary infections of short and long duration. *Vet Microbiol* 119, 194-204.

R Core Team, R. D. & Team, C. (2008).R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

Ramsay, M. E., Andrews, N. J., Trotter, C. L., Kaczmarski, E. B. & Miller, E. (2003). Herd immunity from meningococcal serogroup C conjugate vaccination in England: database analysis. *BMJ* 326, 365-366.

Rappuoli, R. (2000). Reverse vaccinology. *Curr Opin Microbiol* 3, 445-450.
Reingold, A. L., Broome, C. V., Hightower, A. W. & other authors (1985). Age-specific differences in duration of clinical protection after vaccination with meningococcal polysaccharide A vaccine. *Lancet* ii, 114-118.

Reinhardt, M., Elias, J., Albert, J., Frosch, M., Harmsen, D. & Vogel, U. (2008). EpiScanGIS: an online geographic surveillance system for meningococcal disease. *Int J Health Geogr* 7, 33.

Ribot, E. M., Fitzgerald, C., Kubota, K., Swaminathan, B. & Barrett, T. J. (2001). Rapid pulsed-field gel electrophoresis protocol for subtyping of *Campylobacter jejuni*. *J Clin Microbiol* 39, 1889-1894.

Richmond, P., Borrow, R., Miller, E., Clark, S., Sadler, F., Fox, A., Begg, N., Morris, R. & Cartwright, K. (1999). Meningococcal Serogroup C Conjugate Vaccine Is Immunogenic in Infancy and Primes for Memory. *J Infect Dis* 179, 1569-1572.

Riordan, T., Cartwright, K., Andrews, N. & other authors (1998). Acquisition and carriage of meningococci in marine commando recruits. *Epidemiol Infect* 121, 495-505.

Robbins, J. B. & Schneerson, R. (1990). Polysaccharide-protein conjugates: A new generation of vaccines. *The Journal of Infectious Diseases* 161, 821-832.

Roberts, L. (2008). INFECTIOUS DISEASE: An Ill Wind, Bringing Meningitis. *Science* 320, 1710-1715.

Robinson, D. A., Sutcliffe, J. A., Tewodros, W., Manoharan, A. & Bessen, D. E. (2006). Evolution and Global Dissemination of Macrolide-Resistant Group A Streptococci. *Antimicrob Agents Chemother* 50, 2903-2911.

Rodriguez, A. P., Dickinson, F., Baly, A. & Martinez, R. (1999). The epidemiological impact of antimeningococcal B vaccination in Cuba. *Mem Inst Oswaldo Cruz* 94, 433-440.

Rodríguez de Córdoba, S., Esparza-Gordillo, J., Goicoechea de Jorge, E., Lopez-Trascasa, M. & Sánchez-Corral, P. (2004). The human complement factor H: functional roles, genetic variations and disease associations. *Mol Immunol* 41, 355-367.

Rokbi, B., Renauld-Mongenie, G., Mignon, M., Danve, B., Poncet, D., Chabanel, C., Caugant, D. A. & Quentin-Millet, M. J. (2000). Allelic diversity of the two transferrin binding protein B gene isotypes among a collection of *Neisseria meningitidis* strains representative of serogroup B disease: implication for the composition of a recombinant TbpB-based vaccine. *Infect Immun* 68, 4938-4947.

Rosenstein, N. E., Perkins, B. A., Stephens, D. S. & other authors (1999). The Changing Epidemiology of Meningococcal Disease in the United States 1992-1996. *J Infect Dis* 180, 1894-1901.

Rosenstein, N. E., Perkins, B. A., Stephens, D. S., Popovic, T. & Hughes, J. M. (2001). Meningococcal disease. *N Engl J Med* 344, 1378-1388.

Russell, J. E. (2001).Variation in the PorA protein, and clonal diversity within the UK *Neisseria meningitidis* population over a twenty year period (1975-1995), pp. 235: Open University.

Russell, J. E., Jolley, K. A., Feavers, I. M., Maiden, M. C. & Suker, J. (2004). PorA variable regions of *Neisseria meningitidis*. *Emerg Infect Dis* 10, 674-678.

Russell, J. E., Urwin, R., Gray, S. J., Fox, A. J., Feavers, I. M. & Maiden, M. C. (2008). Molecular epidemiology of meningococcal disease in England and Wales 1975-1995, before the introduction of serogroup C conjugate vaccines. *Microbiology* 154, 1170-1177.

Sacchi, C. T., Lemos, A. P. S., Whitney, A. M., Solari, C. A., Brandt, M. E., Melles, C. E. A., Frasch, C. E. & Mayer, L. W. (1998). Correlation between serological and sequencing analyses of the PorB outer membrane protein in the *Neisseria meningitidis* serotyping system. *Clin Diagn Lab Immunol* 5, 348-354.

Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4, 406-425.

Salit, I. E. (1981). Meningococcemia caused by serogroup W135. Association with hypogammaglobulinemia. *Arch Intern Med* 141, 664-665.

Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Acadamy of Sciences of the USA* 74, 5463-5467.

Scarselli, M., Cantini, F., Santini, L. & other authors (2008). Epitope mapping of a bactericidal monoclonal antibody against the factor H-binding protein of *Neisseria meningitidis*. *J Mol Biol* 386, 97-108.

Scherp, H. W. & Rake, G. (1935). Studies on meningococcus infection: VIII. The type I specific substance. *J Exp Med* 61, 753-769.

Scherp, H. W. & Rake, G. (1945). Studies on meningococcal infection: XIII. Correlation between antipolysaccharide and the antibody which protects mice against infection with type I meningococci. *J Exp Med* 81, 85-92.

Schildkamp, R. L., Lodder, M. C., Bijlmer, H. A., Dankert, J. & Scholten, R. J. (1996). Clinical manifestations and course of meningococcal disease in 562 patients. *Scand J Infect Dis* 28, 47-51.

Schneider, M. C., Exley, R. M., Chan, H., Feavers, I., Kang, Y.-H., Sim, R. B. & Tang, C. M. (2006). Functional significance of factor H-binding to *Neisseria meningitidis* . *J Immunol* 176, 7566-7575.

Schneider, M. C., Prosser, B. E., Caesar, J. J. E. & other authors (2009). *Neisseria meningitidis* recruits factor H using protein mimicry of host carbohydrates. *Nature* 458, 890-893.

Scholten, R. J., Kuipers, B., Valkenburg, H. A., Dankert, J., Zollinger, W. D. & Poolman, J. T. (1994). Lipo-oligosaccharide immunotyping of *Neisseria meningitidis* by a whole-cell ELISA with monoclonal antibodies. *J Med Microbiol* 41, 236-243.

Schouls, L. M., van der Ende, A., Damen, M. & van de Pol, I. (2006). Multiple-locus variable-number tandem repeat analysis of *Neisseria meningitidis* yields groupings similar to those obtained by multilocus sequence typing. *J Clin Microbiol* 44, 1509-1518.

Schwentker, F. F., Gelman, S. & Long, P. H. (1937). Landmark article April 24, 1937. The treatment of meningococcic meningitis with sulfanilamide. Preliminary report. By Francis F. Schwentker, Sidney Gelman, and Perrin H. Long. *Jama* 108, 1407-1408.

Seib, K. L., Serruto, D., Oriente, F., Delany, I., Adu-Bobie, J., Veggi, D., Arico, B., Rappuoli, R. & Pizza, M. (2009). Factor H-binding protein is important for meningococcal survival in human whole blood and serum and in the presence of the antimicrobial peptide LL-37. *Infect Immun* 77, 292-299.

Selander, R. K., Caugant, D. A., Ochman, H., Musser, J. M., Gilmour, M. N. & Whittam, T. S. (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol* 51, 837-884.

Sierra, G. V. G., Campa, H. C., Varcacel, N. M. & other authors (1991). Vaccine against group B *Neisseria meningitidis*: protection trial and mass vaccination results in Cuba. *NIPH Ann* 14, 195-207.

Simpson, E. H. (1949). Measurement of Diversity. *Nature* 163, 688.

Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., Heiner, C., Kent, S. B. & Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature* 321, 674-679.

Sobanski, M. A., Barnes, R. A. & Coakley, W. T. (2001). Detection of meningococcal antigen by latex agglutination. In *Meningococcal Disease*, pp. 41-59. Edited by A. J. Pollard & M. C. J. Maiden. Totowa, NJ: Humana Press.

Spratt, B. G. & Maiden, M. C. J. (1999). Bacterial population genetics, evolution and epidemiology. *Proc R Soc Lond B Biol Sci* 354, 701-710.

Staden, R. (1996). The Staden sequence analysis package. *Mol Biotechnol* 5, 233-241.

Statacorp (2007).Stata Statistical Software: Release 10. College Station, TX: StataCorp LP.

Stefanelli, P., Fazio, C., Neri, A., Sofia, T. & Mastrantonio, P. (2003). First report of capsule replacement among electrophoretic type 37 *Neisseria meningitidis* strains in Italy. *J Clin Microbiol* 41, 5783-5786.

Stein, D. M., Robbins, J., Miller, M. A., Lin, F. Y. & Schneerson, R. (2006). Are antibodies to the capsular polysaccharide of *Neisseria meningitidis* group B and *Escherichia coli* K1 associated with immunopathology? *Vaccine* 24, 221-228.

Stephens, D. S. (2007). Conquering the meningococcus. *FEMS Microbiol Rev* 31, 3-14.

Stiehm, E. R. & Damrosch, D. S. (1966). Factors in the prognosis of meningococcal infection. Review of 63 cases with emphasis on recognition and management of the severely ill patient. *J Pediatr* 68, 457-467.

Suker, J., Feavers, I. M., Achtman, M., Morelli, G., Wang, J. F. & Maiden, M. C. J. (1994). The *porA* gene in serogroup A meningococci: evolutionary stability and mechanism of genetic variation. *Mol Microbiol* 12, 253-265.

Sultan, B., Labadi, K., Guégan, J. F. & Janicot, S. (2005). Climate drives the meningitis epidemics onset in West Africa. *PLoS Med* 2, e6.

Swain, C. L. (2006). Survival of meningococci outside of the host: implications for acquisition. *Epidemiol Infect*, 1-6.

Swartley, J. S., Marfin, A. A., Edupuganti, S., Liu, L. J., Cieslak, P., Perkins, B., Wenger, J. D. & Stephens, D. S. (1997). Capsule switching of *Neisseria meningitidis*. *Proc Natl Acad Sci USA* 94, 271-276.

Swartley, J. S., Liu, L. J., Miller, Y. K., Martin, L. E., Edupuganti, S. & Stephens, D. S. (1998). 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*. *J Bacteriol* 180, 1533-1539.

Taha, M. K., Achtman, M., Alonso, J. M., Greenwood, B., Ramsay, M., Fox, A., Gray, S. & Kaczmarski, E. (2000). Serogroup W135 meningococcal disease in Hajj pilgrims. *Lancet* 356, 2159.

Taha, M. K., Bichier, E., Perrocheau, A. & Alonso, J. M. (2001). Circumvention of herd immunity during an outbreak of meningococcal disease could be correlated to escape mutation in the *porA* gene of *Neisseria meningitidis*. *Infect Immun* 69, 1971-1973.

Takahashi, H., Kuroki, T., Watanabe, Y., Tanaka, H., Inouye, H., Yamai, S. & Watanabe, H. (2004). Characterization of *Neisseria meningitidis* isolates collected from 1974 to 2003 in Japan by multilocus sequence typing. *J Med Microbiol* 53, 657-662.

Tettelin, H., Saunders, N. J., Heidelberg, J. & other authors (2000). Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* 287, 1809-1815.

Teyssou, R. & Muros-Le Rouzic, E. (2007). Meningitis epidemics in Africa: A brief overview. *Vaccine* 25, A3-A7.

Thompson, E. A. L. (2001).Antigenic variation in the potential meningococcal vaccine candidate FetA, pp. 183. Oxford: University of Oxford.

Thompson, E. A. L., Feavers, I. M. & Maiden, M. C. J. (2003). Antigenic diversity of meningococcal enterobactin receptor FetA, a vaccine component. *Microbiology* 149, 1849-1858.

Tommassen, J., Vermeij, P., Struyvé, M., Benz, R. & Poolman, J. T. (1990). Isolation of *Neisseria meningitidis* mutants deficient in class 1 (PorA) and class 3 (PorB) outer membrane proteins. *Infect Immun* 58, 1355-1359.

Troncoso, G., Sanchez, S., Moreda, M., Criado, M. T. & Ferreiros, C. M. (2000). Antigenic cross-reactivity between outer membrane proteins of *Neisseria meningitidis* and commensal *Neisseria* species. *FEMS Immunol Med Microbiol* 27, 103-109.

Trotter, C. & Ramsay, M. (2007). Vaccination against meningococcal disease in Europe: review and recommendations for the use of conjugate vaccines. *FEMS Microbiol Rev* 31, 101 - 107.

Trotter, C. L., Andrews, N. J., Kaczmarski, E. B., Miller, E. & Ramsay, M. E. (2004). Effectiveness of meningococcal serogroup C conjugate vaccine 4 years after introduction. *Lancet* 364, 365-367.

Trotter, C. L., Gay, N. J. & Edmunds, W. J. (2006a). The natural history of meningococcal carriage and disease. *Epidemiol Infect* 134, 556-566.

Trotter, C. L., Ramsay, M. E., Gray, S., Fox, A. & Kaczmarski, E. (2006b). No evidence for capsule replacement following mass immunisation with meningococcal serogroup C conjugate vaccines in England and Wales. *Lancet Infect Dis* 6, 616-617; author reply 617-618.

Trotter, C. L., Chandra, M., Cano, R. & other authors (2007). A surveillance network for meningococcal disease in Europe. *FEMS Microbiol Rev* 31, 27-36.

Trotter, C. L. & Greenwood, B. M. (2007). Meningococcal carriage in the African meningitis belt. *Lancet Infect Dis* 7, 797-803.

Tsai, C. M., Frasch, C. E. & Mocca, L. F. (1981). Five structural classes of major outer membrane proteins in *Neisseria meningitidis*. *J Bacteriol* 146, 69-78.

Urwin, R., Holmes, E. C., Fox, A. J., Derrick, J. P. & Maiden, M. C. (2002). Phylogenetic evidence for frequent positive selection and recombination in the meningococcal surface antigen PorB. *Mol Biol Evol* 19, 1686-1694.

Urwin, R., Russell, J. E., Thompson, E. A., Holmes, E. C., Feavers, I. M. & Maiden, M. C. (2004). Distribution of Surface Protein Variants among Hyperinvasive Meningococci: Implications for Vaccine Design. *Infect Immun* 72, 5955-5962.

van Belkum, A., Tassios, P. T., Dijkshoorn, L. & other authors (2007). Guidelines for the validation and application of typing methods for use in bacterial epidemiology. *Clinical Microbiology and Infection* 13, 1-46.

van Brakel, M. J., van Vught, A. J. & Gemke, R. J. (2000). Pediatric risk of mortality (PRISM) score in meningococcal disease. *Eur J Pediatr* 159, 232-236.

van den Dobbelsteen, G., van Dijken, H., Hamstra, H. J., Ummels, R., van Alphen, A. J. W. & van der Ley, P. (2004).From HexaMen to NonaMen: expanding a multivalent PorA-based meningococcal outer membrane vesicle vaccine. In *IPNC 2006* . Milwaukee, USA.

van den Dobbelsteen, G. P. J. M., van Dijken, H. H., Pillai, S. & van Alphen, L. (2007). Immunogenicity of a combination vaccine containing pneumococcal conjugates and meningococcal PorA OMVs. *Vaccine* 25, 2491-2496.

van der Ende, A., Hopman, C. T. P., Zaat, S., Oude Essink, B. B., Berkhout, B. & Dankert, J. (1995). Variable expression of class 1 outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the -10 and -35 regions of the promoter. *J Bacteriol* 177, 2475-2480.

van der Ende, A., Hopman, C. T. P. & Dankert, J. (1999). Deletion of *porA* by recombination between clusters of repetitive extragenic palindromic sequences in *Neisseria meningitidis*. *Infect Immun* 67, 2928-2934.

van der Ende, A., Hopman, C. T. P. & Dankert, J. (2000). Multiple mechanisms of phase variation of PorA in *Neisseria meningitidis*. *Infect Immun* 68, 6685-6690.

van der Ende, A., Hopman, C. T., Keijzers, W. C., Spanjaard, L., Lodder, E. B., van Keulen, P. H. & Dankert, J. (2003). Outbreak of meningococcal disease caused by PorA-deficient meningococci. *The Journal of Infectious Diseases* 187, 869-871.

van der Ley, P., van der Biezen, J., Sutmuller, R., Hoogerhout, P. & Poolman, J. T. (1996). Sequence variability of FrpB, a major iron-regulated outer-membrane protein in the pathogenic *Neisseriae*. *Microbiology* 142, 3269-3274.

van Deuren, M., Brandtzaeg, P. & van der Meer, J. W. (2000). Update on meningococcal disease with emphasis on pathogenesis and clinical management. *Clinical Microbiology Reviews* 13, 144-166, table of contents.

Vazquez, J. (2001). The resistance of *Neisseria meningitidis* to the antimicrobial agents: an issue still in evolution. *Rev Med Microbiol* 12, 39-45.

Verheul, A. F. M., Snippe, H. & Poolman, J. T. (1993). Meningococcal lipopolysaccharides: Virulence Factor and Potential Vaccine Component. *Microbiol Rev* 57,1, 34-45.

Vermont, C. L., van Dijken, H. H., Kuipers, A. J., van Limpt, C. J. P., Keijzers, W. C. M., van der Ende, A., de Groot, R., van Alphen, L. & van den Dobbelsteen, G. P. J. M. (2003). Cross-Reactivity of Antibodies against PorA after Vaccination with a Meningococcal B Outer Membrane Vesicle Vaccine. *Infect Immun* 71, 1650-1655.

Virji, M., Makepeace, K., Ferguson, D. J., Achtman, M. & Moxon, E. R. (1993). Meningococcal Opa and Opc proteins: their role in colonization and invasion of human epithelial and endothelial cells. *Mol Microbiol* 10, 499-510.

Vogel, U. (2001). Meningitis. Outcome of a complex interaction. *Lancet* 358 Suppl, S43.

Wang, J. F., Caugant, D. A., Li, X., Hu, X., Poolman, J. T., Crowe, B. A. & Achtman, M. (1992). Clonal and antigenic analysis of serogroup A *Neisseria meningitidis* with particular reference to epidemiological features of epidemic meningitis in China. *Infect Immun* 60, 5267-5282.

Wang, J. F., Caugant, D. A., Morelli, G., Koumaré, B. & Achtman, M. (1993). Antigenic and epidemiological properties of the ET-37 complex of *Neisseria meningitidis*. *J Infect Dis* 167, 1320-1329.

Weber, M. V. R., Claus, H., Maiden, M. C. J., Frosch, M. & Vogel, U. (2006). Genetic mechanisms for loss of encapsulation in polysialyltransferase-gene-positive meningococci isolated from healthy carriers. *Int J Med Microbiol* 296, 475-484.

Wedege, E., Høiby, E. A., Rosenqvist, E. & Bjune, G. (1998). Immune responses against major outer membrane antigens of *Neisseria men ingitidis* in vaccinees and controls who contracted meningococcal disease during the Norwegian serogroup B protection trial. *Infect Immun* 66, 3223-3231.

Welsch, J., Ram, S., Koeberling, O. & Granoff, D. (2008). Complement-dependent synergistic bactericidal activity of antibodies against factor H-binding protein, a

sparsely distributed meningococcal vaccine antigen. *The Journal of Infectious Diseases* 197, 1053-1061.

Welsch, J. A., Rossi, R., Comanducci, M. & Granoff, D. M. (2004). Protective activity of monoclonal antibodies to genome-derived neisserial antigen 1870, a *Neisseria meningitidis* candidate vaccine. *J Immunol* 172, 5606-5615.

Welsch, J. A. & Ram, S. (2008). Factor H and Neisserial pathogenesis. *Vaccine* 26, I40-I45.

West, D., Reddin, K., Matheson, M., Heath, R., Funnell, S., Hudson, M., Robinson, A. & Gorringe, A. (2001). Recombinant *Neisseria meningitidis* transferrin binding protein A protects against experimental meningococcal infection. *Infect Immun* 69, 1561-1567.

Whalen, C. M., Hockin, J. C., Ryan, A. & Ashton, F. (1995). The changing epidemiology of invasive meningococcal disease in Canada, 1985 through 1992. Emergence of a virulent clone of *Neisseria meningitidis*. *Jama* 273, 390-394.

WHO (2003).Prevention and control of epidemic meningococcal disease in Africa - Report of a WHO technical consultation meeting. Geneva: WHO - Department of Communicable Disease Surveillance and Response.

WHO (2007). Risk of epidemic meningitis in Africa: a cause for concern. *WHO-Weekly epidemiological record* 10.

Wilson, D. J. & McVean, G. (2006). Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics* 172, 1411-1425.

Womble, D. D. (2000). GCG: The Wisconsin Package of sequence analysis programs. *Methods Mol Biol* 132, 3-22.

Woods, T. C., Helsel, L. O., Swaminathan, B. & other authors (1992). Characterization of *Neisseria meningitidis* serogroup C by multilocus enzyme electrophoresis and ribosomal DNA restriction profiles (ribotyping). *J Clin Microbiol* 30, 132-137.

Wright, J. C., Plested, J. S. & Moxon, E. R. (2006). Genetics, structure and function of lipopolysaccharide. In *Handbook of meningococcal disease*, pp. 163-179. Edited by M. Frosch & M. C. Maiden. Weinheim, Germany: Wiley-VCH.

Wright, S. (1950). Genetical structure of populations. *Nature* 166, 247-249.

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19, 395-420.

Wyle, F. A., Artenstein, M. S., Brandt, B. L., Tramont, E. C., Kasper, D. L., Altieri, P. L., Berman, S. L. & Lowenthal, J. P. (1972). Immunologic response of man to group B meningococcal polysaccharide vaccines. *J Infect Dis* 126, 514-521.

Yazdankhah, S. P. & Caugant, D. A. (2004). *Neisseria meningitidis*: an overview of the carriage state. *J Med Microbiol* 53, 821-832.

Yazdankhah, S. P., Kriz, P., Tzanakaki, G. & other authors (2004). Distribution of serogroups and genotypes among disease-associated and carried isolates of *Neisseria meningitidis* from the Czech Republic, Greece, and Norway. *J Clin Microbiol* 42, 5146-5153.

Yazdankhah, S. P., Kesanopoulos, K., Tzanakaki, G., Kremastinou, J. & Caugant, D. A. (2005a). Variable-number tandem repeat analysis of meningococcal isolates belonging to the sequence type 162 complex. *J Clin Microbiol* 43, 4865-4867.

Yazdankhah, S. P., Lindstedt, B. A. & Caugant, D. A. (2005b). Use of variable-number tandem repeats to examine genetic diversity of *Neisseria meningitidis*. *J Clin Microbiol* 43, 1699-1705.

Zhang, X., Shao, Z., Zhu, Y., Xu, L., Xu, X., Mayer, L. W., Xu, J. & Jin, Q. (2008). Genetic characteristics of serogroup A meningococci circulating in China 1956-2005. *Clinical Microbiology and Infection* 14, 555-561.

Zhou, J., Bowler, L. D. & Spratt, B. G. (1997). Interspecies recombination, and phylogenetic distortions, within the glutamine synthetase and shikimate dehydrogenase genes of *Neisseria meningitidis* and commensal *Neisseria* species. *Mol Microbiol* 23, 799-812.

APPENDICES

APPENDIX 1

Table 1: Summary of isolates with completed MLST for each country and year

country	year			
	2000	**2001**	**2002**	**total**
Austria	54	76	57	187
Belgium	80	117	72	269
Czech Rep	53	67	70	190
Denmark	40	50	84	174
E & W	77	150	98	325
Finland	40	47	45	132
France	93	122	125	340
Germany	141	171	213	525
Greece	42	39	42	123
Iceland	0	17	0	17
Ireland	49	27	31	107
Italy	31	23	26	80
Netherlands	175	238	204	617
Norway	74	66	48	188
Portugal	0	36	0	36
Scotland	78	78	80	236
Spain	145	118	146	409
Sweden	0	58	35	93
total	1172	1500	1376	4048

Table 2: PorA typing submitted and completed by country

country	boilate	DNA	data	incomplete	total
Austria	188				188
Belgium	262			21	283
Czech Rep.	89		19	86	194
Denmark	173			2	175
E & W	14		241	81	336
Finland	132			7	139
France	19		355	6	380
Germany	393			144	537
Greece			68	51	119
Iceland	17				17
Ireland	112			1	113
Italy	77			4	81
Netherlands		562	1	77	640
Norway			181	7	188
Portugal	36				36
Scotland			235	1	236
Spain	382		1	41	424
Sweden			91	2	93
	1894	562	1101	529	4086

Table 3: FetA typing submitted and completed by country

country	boilate	DNA	data	incomplete	total
Austria	134			54	188
Belgium	220			63	283
Czech Rep.	83		62	49	194
Denmark	137			38	175
E & W	15		162	159	336
Finland	129			10	139
France	19		337	24	380
Germany	336			201	537
Greece			50	69	119
Iceland	17				17
Ireland	107			6	113
Italy	65			16	81
Netherlands		508		132	640
Norway			180	8	188
Portugal	34			2	36
Scotland				236	236
Spain	349		1	74	424
Sweden				93	93
	1645	508	792	1141	4086

Table 4: Breakdown of clonal complexes found in Europe with ST make up and diversity index (*D*) of each complex in terms of ST. For each complex, only STs with a frequency of more than 3 are included in the table. *D* values near 1 indicate high diversity while values near 0 indicate low diversity. Calculations of 95% confidence intervals performed as described by Grundmann *et al* (Grundmann *et al.*, 2001). Non-overlapping CIs indicate significant difference in *D*.

clonal complex (% of dataset)	total no. of isolates	no. of STs	STs present > 3 times	no. of isolates	Index of Diversity (*D*)	95% CIs
ST-41/44 complex (25%)	1014	259	41	297	0.9007	0.8848-0.9165
			42	85		
			40	56		
			154	31		
			1403	26		
			1194	20		
			44	17		
			43	14		
			280	14		
			1374	13		
			136	11		
			409	11		
			303	10		
			414	9		
			571	9		
			1127	9		
			340	8		
			1960	8		
			112	6		
			180	6		
			839	6		
			1097	6		
			2016	6		
			3488	6		
			170	5		
			318	5		
			1103	5		
			2708	5		
			482	4		
			1196	4		
			1788	4		
			2288	4		
			2925	4		
			3515	4		
ST-11 complex (22%)	903	66	11	813	0.1878	0.1520-0.2235

clonal complex	total no. of isolates	no. of STs	STs present > 3 times	no. of isolates	Index of Diversity (*D*)	95% CIs
			247	7		
			2704	4		
ST-32 complex (17%)	707	94	32	312	0.7697	0.7408-0.7985
			33	92		
			34	79		
			749	48		
			259	17		
			1130	16		
			1249	13		
			2264	7		
			265	4		
			1332	4		
			2685	4		
			3324	4		
ST-8 complex (7%)	273	46	8	112	0.7677	0.7267-0.8088
			66	67		
			2680	15		
			153	8		
			2711	6		
			2757	5		
			3524	5		
			1094	4		
			1372	4		
			2699	4		
ST-269 complex (6%)	256	62	269	94	0.8475	0.8069-0.8882
			1163	21		
			275	18		
			1049	11		
			283	10		
			479	10		
			1195	10		
			13	6		
			352	5		
			492	5		
			1214	4		
ST-213 complex (2%)	74	14	213	58	0.4032	0.2572-0.5493
ST-23 complex (2%)	68	15	23	48	0.5004	0.3509-0.6500
			2692	4		
ST-22 complex (2%)	67	21	22	24	0.8381	0.7671-0.9091
			184	11		
			1224	6		
			1158	4		

clonal complex	total no. of isolates	no. of STs	STs present > 3 times	no. of isolates	Index of Diversity (*D*)	95% CIs
ST-60 complex (2%)	65	28	60	25	0.8428	0.7580-0.9276
			1940	6		
			1243	4		
ST-35 complex (1%)	43	16	35	21	0.7530	0.6186-0.8875
			160	5		
ST-461 complex (1%)	43	12	461	26	0.6268	0.4625-0.7911
			1946	5		
ST-162 complex (1%)	41	10	162	32	0.4038	0.1422-0.6655
ST-18 complex (1%)	37	13	18	16	0.7639	0.6562-0.8716
			145	10		
ST-174 complex (1%)	32	14	2495	6	0.9073	0.8632-0.9514
			174	5		
			1060	5		
			1466	5		
ST-334 complex (0.5%)	21	7	1031	13	0.6143	0.3859-0.8427
ST-167 complex (0.5%)	19	8	167	6	0.8187	0.7081-0.9293
			168	6		
ST-364 complex (0.3%)	14	13	n/a	n/a	0.9890	0.9623-1.0157
ST-254 complex (0.3%)	13	13	n/a	n/a	0.9890	0.9623-1.0157
ST-103 complex (0.3%)	12	4	103	9	0.4545	0.1212-0.7879
ST-865 complex (0.3%)	12	8	3327	4	0.8939	0.7646-1.0232
ST-231 complex (0.3%)	11	5	1909	4	0.8182	0.6961-0.9403
ST-750 complex (0.3%)	11	6	750	6	0.7273	0.4543-1.0002
ST-1157 complex (0.2%)	8	2	1157	7	0.2500	-0.1008-0.6008
ST-53 complex (0.2%)	7	2	53	6	0.2857	-0.0922-0.6636
ST-5 complex (0.1%)	5	2	7	4	0.4000	-0.0293-0.8293
ST-226 complex (0.1%)	4	3	n/a	n/a	0.8333	0.5833-1.0833
ST-198 complex (0.07%)	3	2	n/a	n/a	0.6667	0.3038-1.0296
ST-92 complex (0.07%)	3	3	n/a	n/a	n/a	
ST-37 complex (0.04%)	2	2	n/a	n/a	n/a	
ST-376 complex (0.04%)	2	2	n/a	n/a	n/a	
ST-116 complex (0.02%)	1	1	n/a	n/a	n/a	
ST-549 complex (0.02%)	1	1	n/a	n/a	n/a	
No value/unassigned (7%)	278	224	n/a	n/a	n/a	
total	**4048**					

Table 5: Diversity and evenness of each country in terms of ST.

country	*D*	*E*
Finland	0.9661	0.4611
Spain	0.9641	0.1618
Greece	0.9474	0.3524
Ireland	0.9455	0.4266
Austria	0.9433	0.2484
Germany	0.9423	0.0947
Italy	0.9402	0.4180
Scotland	0.9356	0.1670
E & W	0.9354	0.1407
Sweden	0.9278	0.3377
Netherlands	0.9251	0.0839
Norway	0.9139	0.1590
Portugal	0.9079	0.4723
Belgium	0.9022	0.1136
Czech Rep	0.8780	0.1223
France	0.8408	0.0641
Denmark	0.8367	0.1004
Iceland	0.5147	0.3434

Table 6: F_{ST} values of pairwise comparisons of STs in 18 European countries.

	Austria	Belgium	Czech Rep	Denmark	E & W	Finland	France	Germany	Greece	Iceland	Ireland	Italy	Netherlands	Norway	Portugal	Scotland	Spain
Belgium	0.02008																
Czech Rep	0.03177	0.07081															
Denmark	0.07132	0.14635	0.08768														
E & W	0.01219	0.02305	0.04516	0.09514													
Finland	0.02022	0.06368	0.04283	0.07221	0.02914												
France	0.03070	0.04154	0.02251	0.12358	0.05122	0.06844											
Germany	0.00900	0.02648	0.03080	0.05882	0.01707	0.02298	0.02938										
Greece	0.05062	0.09300	0.07899	0.12729	0.06481	0.03845	0.09130	0.05630									
Iceland	0.09779	0.13315	0.03769	0.17553	0.13061	0.12605	0.02537	0.09359	0.14263								
Ireland	0.02259	0.03053	0.06869	0.11365	0.00601	0.03222	0.07412	0.02927	0.07376	0.17122							
Italy	0.02108	0.01626	0.07699	0.13145	0.02487	0.04849	0.07531	0.03208	0.08250	0.17910	0.02741						
Netherlands	0.01332	0.00597	0.05894	0.11380	0.01586	0.04885	0.04974	0.02235	0.08943	0.13332	0.01890	0.00398					
Norway	0.01841	0.06800	0.03302	0.02173	0.03451	0.01919	0.05411	0.01449	0.05978	0.10055	0.04977	0.05841	0.05045				
Portugal	0.08296	0.09032	0.16752	0.22565	0.11800	0.11575	0.12847	0.08085	0.11875	0.23413	0.13301	0.12508	0.11501	0.14348			
Scotland	0.01470	0.03945	0.03681	0.09234	0.01806	0.02343	0.03927	0.01707	0.05214	0.08761	0.03158	0.04874	0.03932	0.03282	0.07632		
Spain	0.02173	0.03829	0.06982	0.10391	0.04592	0.05511	0.05970	0.02110	0.08765	0.13022	0.05632	0.04921	0.04187	0.05834	0.04003	0.03445	
Sweden	0.01367	0.05423	0.05175	0.03613	0.02426	0.01852	0.06296	0.01090	0.06045	0.12970	0.02989	0.04493	0.03978	**0.00533***	0.11547	0.02723	0.04049

bold* $p > 0.05$ can accept null hypothesis of no difference between populations *i.e.* presence of gene flow between two countries

Table 7: F_{ST} values of pairwise comparisons of allelic profiles in 18 European countries.

	Austria	Belgium	Czech Rep	Denmark	E & W	Finland	France	Germany	Greece	Iceland	Ireland	Italy	Netherlands	Norway	Portugal	Scotland	Spain
Belgium	0.02196																
Czech Rep.	0.03311	0.02229															
Denmark	0.05637	0.08906	0.10133														
E & W	0.02198	0.01014	0.02334	0.07912													
Finland	0.01639	0.02650	0.04435	0.05046	0.01695												
France	0.03911	0.01602	0.00829	0.10909	0.02775	0.05255											
Germany	0.00935	0.01918	0.02868	0.04010	0.02069	0.01251	0.03201										
Greece	0.02443	0.03417	0.04434	0.06557	0.02039	0.01923	0.05484	0.02049									
Iceland	0.13043	0.09679	0.05977	0.21529	0.12019	0.15844	0.04455	0.12145	0.15986								
Ireland	0.02645	0.02889	0.03835	0.08212	0.01436	0.02118	0.05015	0.02336	0.02849	0.14790							
Italy	0.01578	0.02529	0.02684	0.07499	0.02071	0.02161	0.03666	0.01680	0.02860	0.13423	0.02249						
Netherlands	0.01751	0.00409	0.02144	0.06885	0.00806	0.01809	0.02198	0.01309	0.02627	0.10331	0.02084	0.01577					
Norway	0.01658	0.03249	0.03464	0.02210	0.02868	0.01792	0.04037	0.00701	0.02664	0.12553	0.03452	0.02777	0.02228				
Portugal	0.06420	0.08219	0.09933	0.11866	0.07104	0.05611	0.10355	0.04611	0.06055	0.24978	0.06778	0.06426	0.07396	0.07605			
Scotland	0.02136	0.01586	0.01455	0.08668	0.00772	0.02209	0.02182	0.01943	0.02504	0.09849	0.01996	0.01862	0.01335	0.02875	0.06800		
Spain	0.02454	0.02962	0.04106	0.07749	0.02039	0.01809	0.04824	0.01888	0.02464	0.15671	0.02555	0.02255	0.02459	0.03537	0.02697	0.02314	
Sweden	0.01825	0.03234	0.04689	0.02474	0.02568	0.00619	0.05204	0.00985	0.02480	0.15425	0.03205	0.02871	0.02188	**0.00551***	0.07252	0.03256	0.03009

bold* $p > 0.05$ can accept null hypothesis of no difference between populations *i.e.* presence of gene flow between two countries

Table 8: Temporal structuring in Europe. F_{ST} values of pairwise comparisons of each year using STs in 18 European countries.

country		F_{ST} values	
		2000	**2001**
Austria	**2001**	0.00881	
	2002	0.00198	0.00168
Belgium		0.00504	
		0.00450	0.01216*
Czech Rep.		-0.00309	0.00504
		0.0045*	0.01216
Denmark		0.00178	
		0.00306	0.00040
E & W		0.00018	
		0.00596	0.00366
Finland		0.00728	
		0.00451	0.0104*
France		0.00091	
		0.00028	0.00023
Germany		0.00208	
		0.0071**	0.0091**
Greece		0.00238	
		0.00103	0.00911
Ireland		0.02247*	
		0.03244*	0.00318
Italy		0.00095	
		0.01175	0.00251
Netherlands		0.02862**	0.01688**
		0.01715**	0.00102
Norway		0.00004	
		0.00133	0.00797
Scotland		0.01719	
		0.06854**	0.01706*
Spain		0.00446*	
		0.01242**	0.00667*
Sweden		n/a	
		n/a	0.00281

$p < 0.05$ ** $p < 0.005$
n/a data not available for Sweden year 2000
note: Data not available for 2000 and 2002 for Iceland and Portugal

Table 9: Temporal structuring in Europe. F_{ST} values of pairwise comparisons of each year using allelic profiles in 18 European countries.

country		F_{ST} values	
		2000	**2001**
Austria	**2001**	0.02303*	
	2002	0.00984	0.00045
Belgium		0.00492	
		-0.00071	0.01650*
Czech Rep.		-0.00309	
		0.01655	0.00266
Denmark		0.00049	
		-0.00571	-0.00304
E & W		-0.00448	
		-0.00084	0.00246
Finland		0.01072	
		0.00097	0.00206
France		0.00519	
		0.00065	0.0010
Germany		-0.0001	
		0.00669	0.00473
Greece		-0.00630*	
		0.01790	0.01477
Ireland		0.05620*	
		0.04287*	-0.00726
Italy		0.00040	
		-0.01460	-0.01012
Netherlands		0.02862**	
		0.02483**	-0.00219
Norway		-0.00761	
		-0.00776	-0.01057
Scotland		0.01045	
		0.06654**	0.02009**
Spain		0.01115*	
		0.02262**	-0.0002
Sweden		n/a	
		n/a	0.00412

p < 0.05 ** p < 0.005
n/a data not available for Sweden year 2000
note: Data not available for 2000 and 2002 for Iceland and Portugal

Table 10: F_{ST} values of pairwise comparisons of straintypes (serogroup:PorA:FetA:ST) in 16 European countries.

	Austria	Belgium	Czech Rep.	Denmark	E & W	Finland	France	Germany	Greece	Iceland	Ireland	Italy	Netherlands	Norway	Portugal
Belgium	0.02282														
Czech Rep.	0.01805	0.04290													
Denmark	0.05007	0.11012	0.08107												
E & W	0.02389	0.02823	0.02456	0.08877											
Finland	0.01493	0.04716	0.02555	0.05507	0.02457										
France	0.03992	0.03828	0.04354	0.12155	0.05322	0.05969									
Germany	0.00555	0.02728	0.02546	0.04603	0.02384	0.01584	0.03946								
Greece	0.02507	0.07165	0.03689	0.05920	0.03091	0.03162	0.08851	0.02881							
Iceland	0.20177	0.20122	0.18401	0.27244	0.21079	0.20366	0.17943	0.20474	0.26777						
Ireland	0.03048	0.03864	0.04281	0.09507	0.00872	0.03146	0.08121	0.03158	0.03945	0.25730					
Italy	0.02567	0.03161	0.04806	0.09657	0.02268	0.02941	0.07903	0.03353	0.05395	0.25613	0.01495				
Netherlands	0.01727	0.01266	0.03921	0.08632	0.01334	0.02465	0.04905	0.01728	0.04578	0.19584	0.01961	0.01070			
Norway	0.00950	0.04616	0.02746	0.02374	0.02773	0.01157	0.05094	0.00760	0.02337	0.20393	0.03669	0.03688	0.02849		
Portugal	0.03875	0.06317	0.05120	0.12500	0.07531	0.06754	0.05670	0.03875	0.06820	0.29227	0.08661	0.09493	0.07076	0.06432	
Spain	0.03053	0.04953	0.01664	0.08699	0.02980	0.02852	0.06578	0.02987	0.04791	0.19602	0.04088	0.04559	0.03907	0.03715	0.04802

bold* $p > 0.05$ can accept null hypothesis of no difference between populations *i.e.* presence of gene flow between two countries

Table 11: Straintype (serogroup:PorA:FetA:ST(cc)) association with country.

finetype	Austria	Belgium	Czech Rep	Denmark	E&W	Finland	France	Germany	Greece	Iceland	Ireland	Italy	Netherlands	Norway	Portugal	Spain	total
B:P1.7-2,4:F1-5:ST-41(cc41/44)	4	**25**	0	4	14	4	6	8	3	0	2	1	39	3	0	12	125
C:P1.5,2:F3-6:ST-11(cc11)	13	**33**	**13**	5	9	0	6	3	1	0	4	2	6	9	0	1	105
C:P1.5-1,10-8:F3-6:ST-11(cc11)	1	3	1	0	7	6	8	2	0	**11**	0	4	33	4	0	20	100
B:P1.7,16:F3-3:ST-32(cc32)	10	0	1	**33**	0	0	2	18	5	1	2	1	**2**	**20**	0	3	98
W-135:P1.5,2:F1-1:ST-11(cc11)	1	0	0	1	7	2	**28**	8	0	0	0	0	5	12	1	0	65
B:P1.7-2,4:F1-5:ST-42(cc41/44)	0	7	0	0	1	0	1	15	0	0	0	1	**25**	1	0	0	51
C:P1.5,2:F5-8:ST-8(cc8)	0	2	0	0	0	0	1	11	0	0	0	1	0	0	**4**	**22**	41
B:P1.19,15:F5-1:ST-33(cc32)	0	1	**7**	1	0	0	0	2	0	0	1	2	2	1	0	**16**	33
B:P1.19,15:F5-1:ST-749(cc32)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**30**	30
B:P1.5-2,10:F5-1:ST-34(cc32)	0	0	0	1	0	0	0	1	0	0	3	0	**21**	0	0	0	26
B:P1.7-2,13-2:F1-5:ST-40(cc41/44)	2	1	0	0	0	0	0	1	0	0	0	**6**	**14**	0	0	1	25
B:P1.7-2,16:F3-3:ST-32(cc32)	0	0	0	2	0	0	0	6	0	0	0	0	10	5	0	2	25
C:P1.5,2:F3-3:ST-11(cc11)	0	0	0	0	0	0	0	**25**	0	0	0	0	0	0	0	0	25
B:P1.19-1,15-11:F5-1:ST-269(cc269)	1	0	2	1	6	0	1	0	3	0	3	1	1	5	0	0	24
C:P1.5,2:F1-1:ST-11(cc11)	0	1	0	0	0	0	0	0	0	0	0	0	**19**	0	0	1	21
B:P1.22,14:F5-5:ST-213(cc213)	0	1	0	2	**6**	0	0	1	0	0	1	1	5	0	0	0	17
C:P1.5-1,10-8:F4-1:ST-11(cc11)	2	0	0	1	1	0	0	7	0	0	0	0	6	0	0	0	17
B:P1.7-2,4:F1-5:ST-154(cc41/44)	0	0	0	0	0	0	0	0	0	0	**14**	0	2	0	0	0	16
B:P1.5,2:F3-6:ST-11(cc11)	0	1	**3**	0	1	3	0	1	0	0	3	1	0	0	0	0	13
C:P1.5,2-1:F5-5:ST-11(cc11)	0	**4**	0	0	0	0	**6**	0	0	0	0	1	0	0	0	1	12
C:P1.5,2:F1-7:ST-8(cc8)	0	0	0	0	0	0	0	6	0	0	0	0	2	1	**3**	0	12
B:P1.7,16-6:F3-3:ST-32(cc32)	1	0	0	0	0	**8**	0	1	0	0	0	0	1	0	0	0	11
C:P1.5,2:F3-9:ST-66(cc8)	**8**	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	11
B:P1.7-2,4:F1-5:ST-1403(cc41/44)	0	0	1	0	0	0	0	1	0	0	2	**5**	1	0	0	0	10
C:P1.5,2:F1-30:ST-11(cc11)	0	0	0	3	1	0	1	1	0	0	0	0	4	0	0	0	10
B:P1.18-1,3:F1-5:ST-1194(cc41/44)	0	0	0	0	0	0	0	4	0	0	3	2	0	0	0	0	9
B:P1.22-1,14:F4-1:ST-35(cc35)	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	5	9
C:P1.5,2:F3-9:ST-2680(cc8)	0	2	0	0	0	0	0	0	0	0	0	0	7	0	0	0	9
C:P1.7-4,14-6:F3-9:ST-1031(cc334)	0	**4**	0	0	1	0	0	3	0	0	0	0	0	1	0	0	9

Y:P1.5-2,10-1:F4-1:ST-23(cc23)	0	0	0	0	0	0	1	**7**	0	0	0	0	0	1	0	0	9
B:P1.18-1,34:F1-5:ST-409(cc41/44)	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	**7**	8
B:P1.22,14:F5-9:ST-162(cc162)	1	0	0	0	0	0	0	1	**6**	0	0	0	0	0	0	0	8
B:P1.7,16-32:F3-3:ST-32(cc32)	0	0	0	0	0	0	0	0	0	0	0	0	**7**	1	0	0	8
B:P1.19,15:F5-1:ST-34(cc32)	0	1	**2**	0	3	0	0	0	0	0	0	0	0	0	0	1	7
B:P1.22,9:F1-55:ST-1163(cc269)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	**7**	7
B:P1.5,2:F5-1:ST-66(cc8)	0	0	0	0	0	0	0	**7**	0	0	0	0	0	0	0	0	7
B:P1.5-2,10-2:F1-5:ST-1127(cc41/44)	0	0	0	0	0	0	0	0	0	0	0	0	0	**7**	0	0	7
C:P1.5-1,10-4:F3-6:ST-11(cc11)	0	0	0	0	2	0	1	1	0	0	0	0	3	0	0	0	7
Y:P1.5-1,2-2:F5-8:ST-23(cc23)	0	0	0	0	0	2	0	0	0	0	0	0	1	**4**	0	0	7
B:P1.18-1,3:F5-1:ST-41(cc41/44)	0	0	0	**3**	0	0	0	0	0	0	0	0	0	3	0	0	6
B:P1.19,15-1:F1-5:ST-43(cc41/44)	0	0	0	2	1	0	0	0	0	0	0	0	0	3	0	0	6
B:P1.19-1,15-11:F5-1:ST-1049(cc269)	0	2	0	0	**4**	0	0	0	0	0	0	0	0	0	0	0	6
B:P1.22,9:F5-1:ST-1195(cc269)	0	0	0	0	3	0	0	0	0	0	**3**	0	0	0	0	0	6
B:P1.5-1,10-8:F3-6:ST-11(cc11)	0	0	0	0	1	0	0	0	0	1	1	1	0	0	0	2	6
B:P1.7-2,13-2:F4-1:ST-1374(cc41/44)	0	0	0	0	0	0	0	0	0	0	0	0	5	1	0	0	6
C:P1.5-1,10-1:F3-6:ST-11(cc11)	0	1	0	0	0	0	0	0	0	0	0	0	3	2	0	0	6
< 5	84	4	5	73	76	51	74	162	27	4	56	33	255	93	16	198	1380
total	128	208	90	132	144	76	136	304	45	17	98	63	483	178	25	329	2456

red significant association; blue significant dissociation with $p<0.0007$ with Bonferroni correction

APPENDIX 2

Copies of the published papers