Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Biologia Cellulare e Molecolare

Ciclo XXVI

Settore concorsuale di afferenza: 05/E2 - BIOLOGIA MOLECOLARE
Settore scientifico disciplinare: BIO/11 - BIOLOGIA MOLECOLARE

TITOLO TESI

Molecular variability of Meningococcal antigens in carriage and disease strains and in other *Neisseria* species

Presentata da: Alessandro Muzzi

Coordinatore Dottorato
Chiar.mo Prof.
Vincenzo Scarlato

Relatore
Chiar.mo Prof.
Vincenzo Scarlato

Esame finale anno 2014

Table of Contents

Outline

My PhD project, described in this thesis, is focused on the study of the molecular variability of some specific proteins, part of the outer membrane of the pathogen *Neisseria meningitidis*, and described as protective antigens and important virulence factors. These antigens have been employed as components of the vaccine developed by Novartis Vaccines against *N. meningitidis* of serogroup B, and their variability in the meningococcal population is a key aspect when the effect of the vaccine is evaluated.

The PhD project has led to complete three major studies described in three different manuscritps, of which two have been published and the third is in preparation. The thesis is structured in three main chapters, each of them dedicated to the three studies.

i. The first study, described in Chapter 1, is specifically dedicated to the analysis of the molecular conservation of meningococcal antigens in the genomes of all species classified in the genus *Neisseria,* whose complete genome was available (*Conservation of Meningococcal Antigens in the Genus Neisseria.* A. Muzzi, M. Mora, M. Pizza, R. Rappuoli, C. Donati. 2013. mBio 4 (3)).
ii. The second study, described in Chapter 2, focuses on the analysis of the presence, distribution and conservation of antigens in a panel of bacterial isolates obtained from cases of the disease and from healthy individuals, and collected in the same year and in the same geographical area, the Czech Republic (*Conservation of fHbp, NadA, and NHBA in carrier and pathogenic isolates of Neisseria meningitidis collected in the Czech Republic in 1993*. A. Muzzi, S. Bambini, K. Jolley, M.C. Maiden, P. Kriz, R. Borrow, J. Lucidarme, *et al.*. Manuscript in preparation).
iii. Finally, Chapter 3 describes the molecular features of antigens in a panel of bacterial isolates collected over a period of 50 years, and representatives of the epidemiological history of meningococcal disease in the Netherlands (*An Analysis of the Sequence Variability of Meningococcal fHbp, NadA and NHBA over a 50-Year Period in the Netherlands*. S. Bambini, J. Piet, A. Muzzi, W. Keijzers, S. Comandi, L. De Tora, M. Pizza, *et al.*. 2013. PloS one 8 (5), e65043).

Furthermore, the study of the variability of these molecules is of importance in predicting the impact of the vaccine after its introduction and has led to two scientific papers, where I have contributed to the analysis of the level of conservation of the meningococcal antigens (references [4] and [5]) .

Chapter 1. Conservation of Meningococcal antigens in the genus *Neisseria*

1.1 Abstract

Neisseria meningitidis, one of the major causes of bacterial meningitis and sepsis, is a member of the *Neisseria* genus, which includes species that colonize the mucosae of many animals. Three meningococcal proteins, factor-H binding protein (fHbp), Neisserial Heparin Binding Antigen (NHBA) and *Neisseria meningitidis* adhesin A (NadA) have been described as protective antigens against *N. meningitidis* of serogroup B and they have been employed as vaccine components in preclinical and clinical studies. In the vaccine formulation, fHbp and NHBA were fused to GNA2091 and GNA1030 proteins, respectively to enhance protein stability and immunogenicity. To determine the possible impact of vaccination on commensal neisseriae, we determined the presence, distribution and conservation of these antigens in the available genome sequences of the *Neisseria* genus, finding that fHbp, NHBA and NadA were conserved only in species colonizing humans, while GNA1030 and GNA2091 were conserved in many human and non-human neisseriae. Sequence analysis showed that homologous recombination contributed to shape the evolution and the distribution of both NHBA and fHbp, for which three major variants have been defined. fHbp variant 3 was probably the ancestral form of meningococcal fHbp, while fHbp variant 1 was introduced into *N. meningitidis* by a recombination event from *N. cinerea*. fHbp variant 2 was the result of a recombination event inserting a stretch of 483 bp from a variant 1 on a variant 3 background. These data indicate that a high rate of exchange of genetic materials exists between those neisseriae that colonize the upper respiratory tract of humans.

The upper respiratory tract of healthy individuals is a complex ecosystem colonized by many bacterial species. Amongst those, there are representatives of the *Neisseria* genus, including *N. meningitidis*, a major cause of bacterial meningitis and sepsis. Given the close relationship between commensal and pathogenic species, a protein-based vaccine against *N. meningitidis* has the potential to impact the other commensal species of *Neisseria*. For this reason, we have studied the distribution and evolutionary history of the antigen components of a recombinant vaccine, 4CMenB that recently received approval in Europe under the commercial name of Bexsero®. We found that fHbp, NHBA and NadA can be

found in some of the human commensal species, and that the evolution of these antigens has been essentially shaped by genetic exchange that occurs at high rate between strains of neisseriae that co-colonize the same environment.

1.2 Introduction

The genus *Neisseria* is a large group of gram-negative bacteria. Beside several human species that are only rarely associated with disease, this genus also contains two major human pathogens, *Neisseria gonorrhoeae* and *Neisseria meningitidis*. The latter generally asymptomatically colonize humans, but in a small percentage of the cases can cause systemic disease.

The taxonomy of the genus *Neisseria* based on sequence analysis is problematic. Studies based on three genes, *argF*, *recA* and *rho* and on 16S rRNA have shown that five phylogenetic groups can be identified, each including more than one named species [6]. However, probably due to the high level of recombination between these organisms, the species within these groups are less distinct than what is normally accepted in molecular taxonomy, as in the case of *Neisseria cinerea* that is often placed within the *N. meningitidis* species. Recently the complete phylogeny of the neisseriae has been reconstructed by ribosomal multilocus sequence typing (rMLST) and an accurate interpretation of the relationships between different species has been proposed [7]. Moreover, many of the genes that are associated with virulence in *N. meningitidis* are present also in commensal neisseriae, and it has been speculated that the latter might constitute a reservoir of virulence factors for the pathogenic species [8]. In some cases, genetic exchange is less frequent due to the different ecology, as for instance between *N. meningitidis* and *N. gonorrhoeae* [9].

Currently, capsular polysaccharide-based vaccines for the A, C, Y and W-135 serogroups of *N. meningitidis* are available, while the development of a polysaccharide vaccine for serogroup B was not possible due to the similarity to a human carbohydrate thus inducing a weak immunoresponse and with the risk of autoimmunity. For this reason, a multicomponent protein based vaccine known as 4CMenB has been proposed [1]. 4CMenB was recently approved in Europe under the commercial name of Bexsero®. 4CMenB includes three main antigens: the factor H binding protein (fHbp), the Neisserial Heparin-Binding Antigen (NHBA) and the *Neisseria meningitidis* adhesin A (NadA). Two additional proteins, GNA2091 and GNA1030, were incorporated into fusion proteins with fHbp

(GNA2091-fHbp) and NHBA (NHBA–GNA1030), respectively, to enhance protein stability and increase immunogenicity [1]. NadA was included as a single protein antigen.

fHbp is a surface-exposed lipoprotein that binds factor H (fH), a key inhibitor of the complement alternative pathway, leading to evasion of killing by the innate immune system [10]. The expression of fHbp by *N. meningitidis* strains is important for survival in human blood and human serum [11]. With very few exceptions [12, 13], the gene coding for fHbp is present in *N. meningitidis* of serogroup B strains in three major genetic variants (variant 1, 2 and 3) and in *N. gonorrhoeae* in one of the three variants [14]. The three variants are not fully cross-protective [3, 15], with variant 1 that does not show cross-protection activity versus variants 2 and 3 and vice versa.

NHBA, predicted to be a surface exposed lipoprotein, was selected as vaccine candidate with the name of GNA2132 (*i.e.* Genome-derived Neisseria Antigen 2132), in a genome-wide discovery program of vaccine targets [16]. GNA2132 was shown to be able to induce bactericidal antibodies in humans, and to bind heparin, improving the survival of the meningococcus in human serum. For this reason, it was later renamed NHBA (Neisserial Heparin Binding Antigen) [17]. NHBA is ubiquitous in meningococci of serogroup B [4, 13, 18].

NadA is a meningococcal surface-anchored protein from the family of trimeric autotransporters (TAs) [19] which are understood to export themselves to the bacterial surface with no external energy source or auxiliary proteins. Recent data suggest that other proteins might be involved in their secretion, and further investigation is ongoing [20]. The gene coding for NadA is present in three out of the four hypervirulent lineages of *N. meningitidis* of serogroup B, and is absent in the pathogenic *N. gonorrhoeae* and in the commensals *N. lactamica* and *N. cinerea* [21, 22]. Although their functional activity is not known, GNA1030 and GNA2091 were shown to induce protective immunity in serum bactericidal essays in mice [1].

The *fHbp*, *nhba* and *nadA* genes are also known to be present in invasive strains of *N. meningitidis* of other serogroups like serogroup C [4, 23], X [24], Y [4, 23] and W-135 [18] and in particular fHbp was investigated in Africa in epidemic strains of serogroup A, X and W-135 [25, 26]. These studies demonstrated the higher level of clonality of the non-B isolates, with the exception of serogroup C, not only in terms of genetic differences of strains but also in terms of molecular variability of these important antigens.

To evaluate whether these antigens are present only in pathogenic neisseriae or are also present in other species belonging to this genus, we analysed their presence, distribution and conservation in the complete and draft genome sequences of 80 strains of

pathogenic and commensal neisseriae, that were available in public databases. We used the Single Nucleotide Polymorphisms (SNPs) retrieved from the alignments of these sequences against the genome of *N. meningitidis* MC58 to resolve the relationships between the species closely related *to N. meningitidis* that could not be clearly distinguished using 16S rRNA. By comparing the phylogenetic relationships between the genes to the whole-genome phylogenetic tree and by analyzing the conservation of the genomic loci where the genes were inserted, we demonstrated the effect of the inter-species transfer of genetic material in their evolution.

1.3 Methods

1.3.1 Phylogenetic analysis of the *Neisseria* genus.

1080 aligned 16S rRNA sequences of the species belonging to the *Neisseria* genus were downloaded (January 2012) from the Ribosomal Database Project web site (http://rdp.cme.msu.edu/index.jsp). The sequences were pruned by keeping only sequences that were more than 1% divergent, obtaining a sample of 48 sequences that were representative of the diversity of the genus. From these, a Maximum Likelihood phylogenetic tree was computed based on the Tamura 3 parameter model [27]. Site variability was modelled with a Gamma distribution (5 categories, G=0.1273) plus invariants (I=66.74% of the sites). The values of the parameters were optimized on the data. For each branch, bootstrap consensus was inferred from 500 replicates. The analysis was conducted using Mega5 [28]. Phylogenetic networks were computed using SplitsTree [29, 30].

1.3.2 Estimate of the recombination and mutation rate.

The population scaled recombination rate r and the population scaled mutation rate q were estimated with the LDHat method [31] and with the Watterson estimator [32], respectively, using the software package RDP3 [33].

1.3.3 Genome sequences.

The 18 complete and 49 draft genome sequences of 67 isolates belonging to the *Neisseria* genus were retrieved from GenBank (December 2011). To improve the representativeness of the dataset, we downloaded (December 2012) 18 additional genome sequences of *N. cinerea* (5 strains), *N. polysaccharea* (7 strains), and *N. lactamica* (6 strains) from the

Neisseria PubMLST database (http://pubmlst.org/perl/bigsdb/bigsdb.pl?db=pubmlst_neisseria_isolates). Of these, 5 (1 *N. cinerea*, 3 *N. lactamica* and 1 *N. polysaccharea*) were re-sequencing of strains already present in the GenBank dataset that we included to correct potential sequencing error in the draft genomes. The draft genomes were composed by between 23 and 574 contigs. A complete list of the isolates is available in Table S1 in the supplemental material. As indicated in Table S1, of the 27 strains of *N. meningitidis*, 2 strains were characterized as strains of serogroup A, 17 of serogroup B, 5 of serogroup C, 1 of serogroup X, 1 of serogroup Y and one strain was not typeable because capsule null locus.

1.3.4 Genome alignments and phylogenetic analysis of the sequenced strains.

Pair-wise genome alignments of all sequences against the genome sequence of *N. meningitidis* MC58 were computed using Nucmer [34]. From these, 223369 Single Nucleotide Polymorphysms (SNPs) were identified in the core genome and the corresponding alleles in all strains were extracted. Phylogenetic analysis was conducted using the Neighbor-Joining method [35]. Evolutionary distances were calculated using the Maximum Composite Likelihood method [36]. The analysis was conducted using Mega5 [28]. Phylogenetic networks were calculated using SplitsTree4 [29, 30].

1.3.5 Genome annotation.

Annotation transfer for draft genomes was performed using RATT [37]. In particular, for *N. polysaccharea* ATCC 43768 (accession number ADBE00000000) we used the annotation of *N. gonorrhoeae* FA 1090 (accession number AE004969) as template. For *N. cinerea* ATCC 14685 (accession number ACDY00000000) we used the annotation of *N. meningitidis* MC58 (accession number AE02098) as template.

1.3.6 Criteria for genes presence/absence.

The presence of the *fHbp*, *nhba* and *nadA* coding genes was verified by nucleotide alignments against predicted coding genes, and finally checked with TFASTY [38] to find possible pseudogenes or coding sequences not predicted in the genome annotation. fHbp query sequences were representative of the three major molecular variants [3]. For NHBA, nucleotide sequences coding for peptides from 1 to 21, representative of the diversity of the molecule in *N. meningitidis* [4, 39], were used as query sequences. For NadA, the query sequences were representative of the five major protein forms [39]. To identify possible

distant homologues, the BLAST E-score cut-off was set to 10^-5, and the sequence alignments were then manually checked to identify true homologues. Gene loci were also inspected by considering the whole genome alignment performed by Mauve [40] and the level of conservation of each Locally Collinear Block (LCB) containing the genes.

1.3.7 Phylogenetic analysis of *fHbp*, *nhba* and *nadA*.

Maximum Likelihood phylogenetic trees were built for each antigen based on the Kimura 2 parameters model with Gamma correction. Branch stability was tested by bootstrap consensus of 500 replicates. The analysis was conducted using Mega5 [28]. Phylogenetic networks were built using SplitTree4 [29]. Phylogenetic distances were computed using the Kimura 2 parameters model with Gamma correction, using the values of the parameters obtained by Mega5.

1.3.8 Identification of recombination events.

To test the presence of recombination, a multiple alignment of each locus was built by extending the *fHbp*, *nhba* and *nadA* loci by 10 kbps upstream and downstream (or to the edge of the contig in the unfinished genomes if this was closer than 10kb). The draft genomes that were fragmented in the neighborhood of the gene locus were not included in the alignment. Recombination events were inferred with RDP3 [33] using the RDP, Geneconv, MaxChi, Bootscan, Chimaera, 3Seq, and Siscan methods.

1.4 Results

1.4.1 Phylogeny of the genus *Neisseria*.

A phylogenetic analysis of the *Neisseria* genus based on 48 16S rRNA sequences that are representative of the diversity of the genus resulted in a tree with generally low values of bootstrap support (Figure 1.1).

However, we could distinguish one branch containing the two species that are pathogenic in humans, *i.e. N. meningitidis* and *N. gonorrhoeae*, together with the human commensals *N. cinerea*, and *N. polysaccharea*. Within this branch, different strains of the same named species did not form monophyletic groups. This inability to resolve the relations between the different species was due both to the weak phylogenetic signal in 16S rRNA sequences and to the high level of homologous recombination within and between these species, that can be visualized by constructing a phylogenetic network (Figure 1.2), where it is almost impossible to clearly distinguish branches corresponding to named species. We estimated the relative contributions of recombination and mutation to the diversification of this set

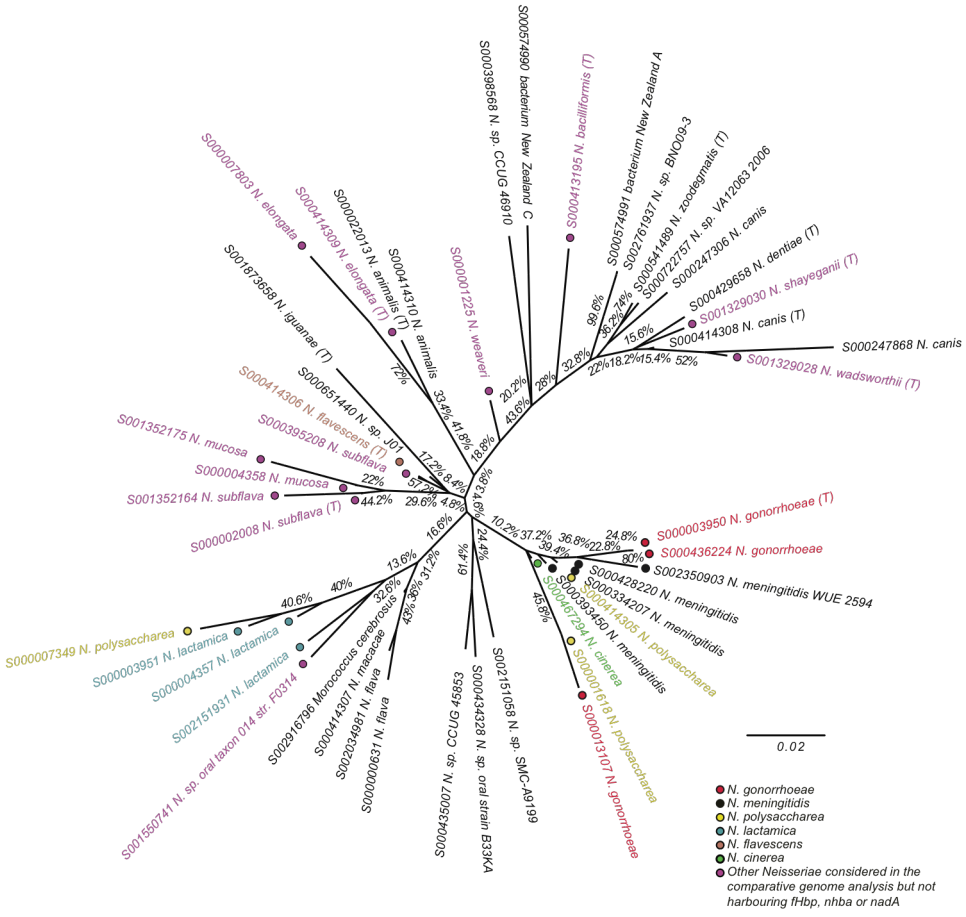


Figure 1.1 Maximum Likelihood phylogenetic tree obtained from 16S rRNA sequences. The bootstrap values are reported. In this tree, strains of the same named species do not form monophyletic branches. A branch containing sequences from N. meningitidis, *N. gonorrhoeae*, *N. polysaccharea* and *N. cinerea* can be identified. Coloured circles and labels indicate species for which a whole genome sequence was available. In particular, purple indicates those sequences where none of the three antigens was conserved, while the other sequences are coloured according to the species. (T) indicates the Type strain for the species.

of sequences by measuring the population scaled recombination rate r and the mutation rate q. We obtained r=0.042 and q=0.015, yielding a value for the ratio r/q=2.8. This value, computed across the different neisseriae species, is only marginally smaller than the estimates of the r/q ratio (r/q=3.1) obtained from strains belonging only to *N. meningitidis* using the sequences of fragments of 7 housekeeping genes [41, 42]. The fact that the value obtained by analysing exclusively sequences of *N. meningitidis* is comparable to the value obtained by analysing sequences of isolates belonging to different species of neisseriae suggests that homologous recombination plays a fundamental role not only in the diversification of lineages within the *N. meningitidis* species [42], but also in driving the speciation process within the *Neisseria* genus [43].

In Figure 1.1 we also indicated the species for which at least one genome sequence was available (see Table 1.1). Although the sample was biased towards those species that colonize humans, the available genomes allowed the sampling of all the branches of the tree. On these genomes, we verified the conservation of the three major antigens contained in the 4CMenB vaccine, namely *fHbp*, *nhba* and *nadA*, and highlighted (in purple in Figure 1.1) those sequences where none of the three antigens was present, while the

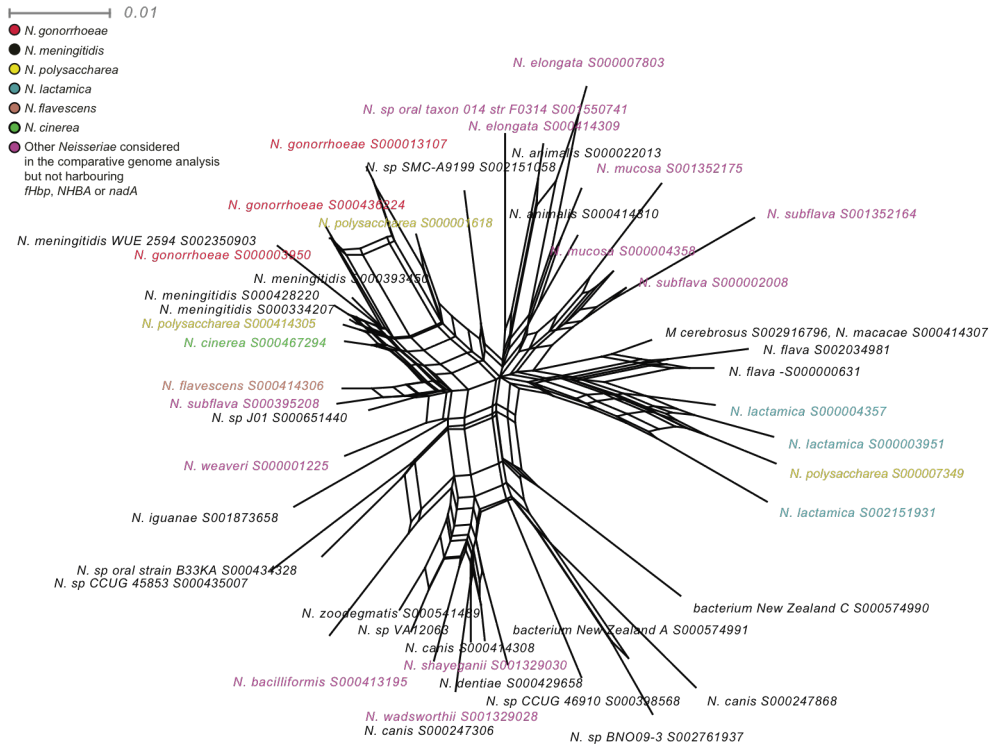


Figure 1.2 Phylogenetic network obtained using the 16*S* rRNA sequences of 48 isolates representative of the diversity of the Neisseria genus. The non tree-like structure of the network is probably the consequence of widespread recombination within the Neisseria genus. Coloured labels indicate the different species of *Neisseria*.

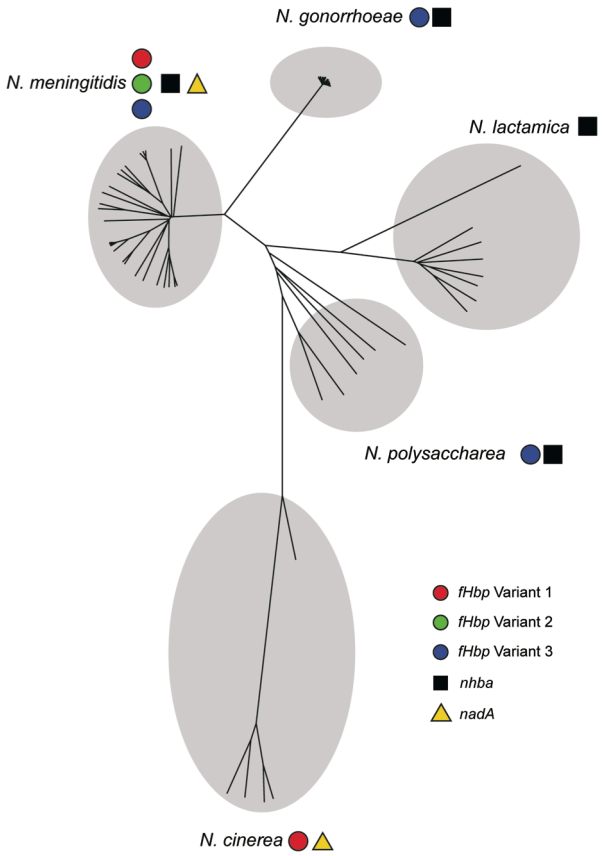


Figure 1.3 Neighbour Joining phylogenetic tree with evolutionary distances calculated using the Maximum Composite Likelihood method obtained from the SNPs in the portion of the multiple alignment shared by all strains. The alignment includes 51 strains of *N. meningitidis*, *N. gonorrhoeae*, *N. polysaccharea* and *N. lactamica*. For clarity, the branches including the *N. meningitidis*, *N. gonorrhoeae, and N. lactamica* strains are grouped in distinct grey areas. The presence of *fHbp*, *nhba* and *nadA* is indicated. For *fHbp* the colours indicate the major allelic variants. In the case of *N. meningitidis*, *fHbp* is present in all strains in one of the three major variants. *nadA* is present in *N. meningitidis* in 7 out of 27 strains with the correct translation frame. *nhba* is ubiquitous with the exception of *N. cinerea*.

other sequences are coloured according to the species. With the exception of *N. flavescens*, all the genomes where at least one antigen was present were concentrated in one branch of the tree, including *N. meningitidis, N. gonorrhoeae, N. polysaccharea, N. lactamica, N. cinerea, N. flava, M. cerebrosus*, *N. macacae,* and three unidentified neisseriae. To resolve the ambiguities of the taxonomy of this branch, we performed a genome wide phylogenetic analysis on the available complete and draft genome sequences (Figure 1.3). This analysis included all the sequences where at least one of the three genes, *fHbp*, *nhba* or *nadA* was found. *N. flavescens* was excluded from this analysis because the available genome was too divergent from the other sequences to allow a reliable alignment. The different species were well separated with no ambiguities, and all the branches were supported by high values of bootstrap (for clarity, the branches containing more than one strain of the same species were grouped in grey areas; the complete tree is reported in

Species	# of strains	fHbp			NHBA			NadA	
		# of strains harbouring a fHbp	fHbp % identity vs 1.1 - species average	notes	# of strains harbouring a NHBA	NHBA % identity vs peptide 2 - species average	notes	# of strains harbouring a NadA	NadA % identity vs NadA-3.8 - species average
Neisseria bacilliformis	1	0	-		0	-		0	-
Neisseria cinerea	5	5	93.9	Variant 1	0	-		2	54.4
Neisseria elongate glycolytica	1	0	-		0	-		0	-
Neisseria flavescens	2	0	-		1	61.2	FS	0	-
Neisseria gonorrhoeae	17 (3 closed genomes)	17	63.3	4 are FS, Variant 3	17	81.2	2 are FS, 1 at contig edge	0	-
Neisseria lactamica	8 (1 closed genome)	0	-	OPA like protein substitutes fHbp	8	83.8	2 at contig edge	0	-
Neisseria macacae	1	0	-		0	-		0	-
Neisseria meningitidis	27 (14 closed genomes)	27	83.7	Variant 1, 2 and 3	27	82.6		12	89.2
Neisseria mucosa	2	0	-		0	-		0	-
Neisseria oral taxon	1	0	-		0	-		0	-
Neisseria polysaccharea	7	7	69.3	5 are FS, 1 at contig edge, Variant 3	7	84.3	2 at contig edge	0	-
Neisseria shayeganii	1	0	-		0	-		0	-
Neisseria sicca	3	0	-		0	-		0	-
Neisseria subflava	1	0	-		0	-		0	-
Neisseria wadsworthii	1	0	-		0	-		0	-
Neisseria weaveri	2	0	-		0	-		0	-

Table 1.1 Summary of the *fHbp*, *nhba*, and *nadA* genes level of conservation in 80 neisserial complete and draft genomes. The % of identity is calculated with respect to the form of the same proteins that are included in the 4CMenB vaccine [1]. Beside intact genes, we included frameshifted (FS) or interrupted by a contig edge *fHbp* and *nhba* genes in *N. polysaccharea, N. lactamica, N. flavescens* and *N. gonorrhoeae*, a 200aa fragment of *NadA* in *N. cinerea* close to a contig edge, two frameshifted *NadA* genes and one interrupted by IS4 in *N. meningitidis*.

Figure 1.4). Amongst these species, *N. gonorrhoeae* was the one most closely related to *N. meningitidis*, and showing the smallest variability within the sequenced strains, suggesting that this species differentiated recently from an ancestor of contemporary *N. meningitidis*. In this reconstruction, the sequenced strains of *N. cinerea* were distantly related to *N. polysaccharea*, and clearly distinct from *N. meningitidis,* confirming recent results based on the analysis of the concatenated sequences of core genes [8].

1.4.2 Conservation of fHbp

Amongst the sequenced genomes of the *Neisseria* genus, the gene encoding for *fHbp* was present only in species closely related to *N. meningitidis*, i.e. *N. gonorrhoeae, N. cinerea* and *N. polysaccharea*, while it was absent, amongst others, from all *N. lactamica* strains (Table

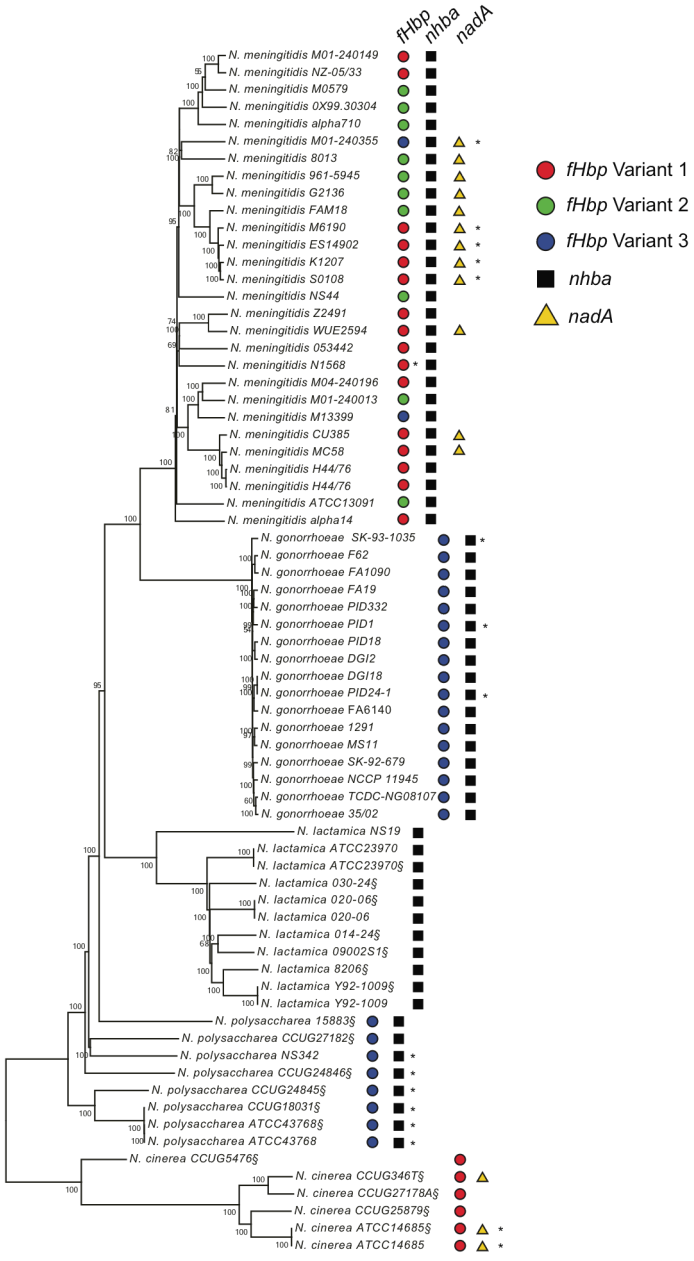


Figure 1.4 Neighbour Joining phylogenetic tree obtained from the SNPs in the portion of the multiple alignment shared by all strains. The alignment includes 64 strains of *N. meningitidis*, *N. gonorrhoeae*, *N. polysaccharea* and *N. lactamica* of which 6 have a duplicated sequence retrieved from multiple genome sequencing projects. The (§) symbol indicates sequences that were downloaded from the BigsDB database, otherwise from GenBank. The presence of *fHbp*, *nhba* and *nadA* is indicated. For *fHbp* the colours indicate the major allelic variants, asterisks (*) indicate genes that were frameshifted, at the border of a contig or interrupted by the an IS.

1.1). Interestingly, in 6/7 of the available strains of *N. polysaccharea*, the gene was frameshifted in different positions, suggesting that this gene does not confer a significant selective advantage in this species. In all strains of *N. gonorrhoeae* and *N. polysaccharea,* the *fHbp* gene could be attributed to the variant 3, while in *N. cinerea* the gene could be classified as a variant 1 (Figure 1.5). The major differences between gonococcal and meningococcal *fHbp* were concentrated in the N-term portion of the molecule (Figure 1.6) where the insertion of a single base (G) at position 40 causes a frame-shift in all gonococcal strains, which results in the loss of the lipo-box motif.

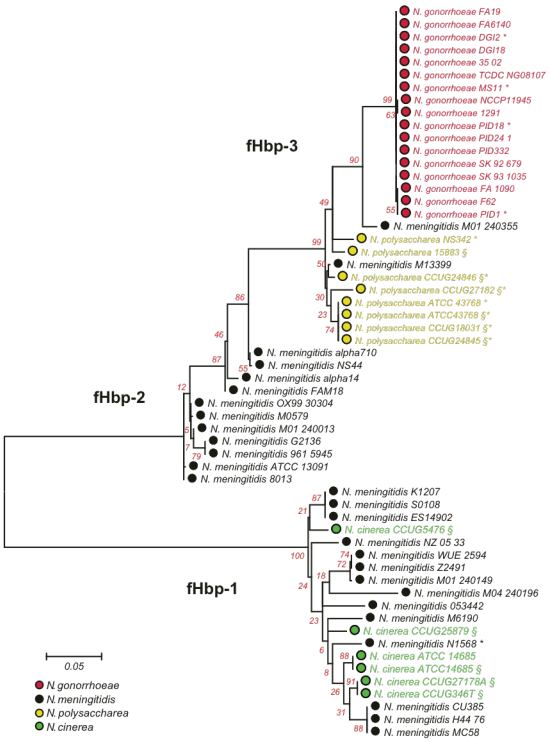


Figure 1.5 Maximum Likelihood phylogenetic tree of the *fHbp* gene**.** The tree was obtained using the Kimura 2 parameters model with Gamma correction. The average gene variability was Pi = 0.131 (s.e. 0.011). Also including sequences from other species, the tree showed the typical structuring in three branches, variant 1, 2 and 3 as already described in *N. meningitidis*. In particular *N. gonorrhoeae* strains harboured variant 3 like genes, as well as the two *N. polysaccharea* strains. *N. cinerea* harboured a *fHbp* variant 1 very similar to the *N. meningitidis* molecules. Coloured circles and labels indicate the different species of *Neisseria*. The (§) symbol indicates sequences that were downloaded from the BigsDB database, otherwise from GenBank. The (#) symbol indicates genes that were at the border of a contig. Asterisks (*) indicate genes that were frameshifted.

When *fHbp* is present, the entire locus harbouring the gene was conserved and it was located between the homologues of NMB1869 and NMB1871 (named according to strain MC58, Figure 1.7A and B). As previously reported [44] the locus was highly conserved also in *N. lactamica* where the *fHbp* gene was always substituted by a gene annotated as

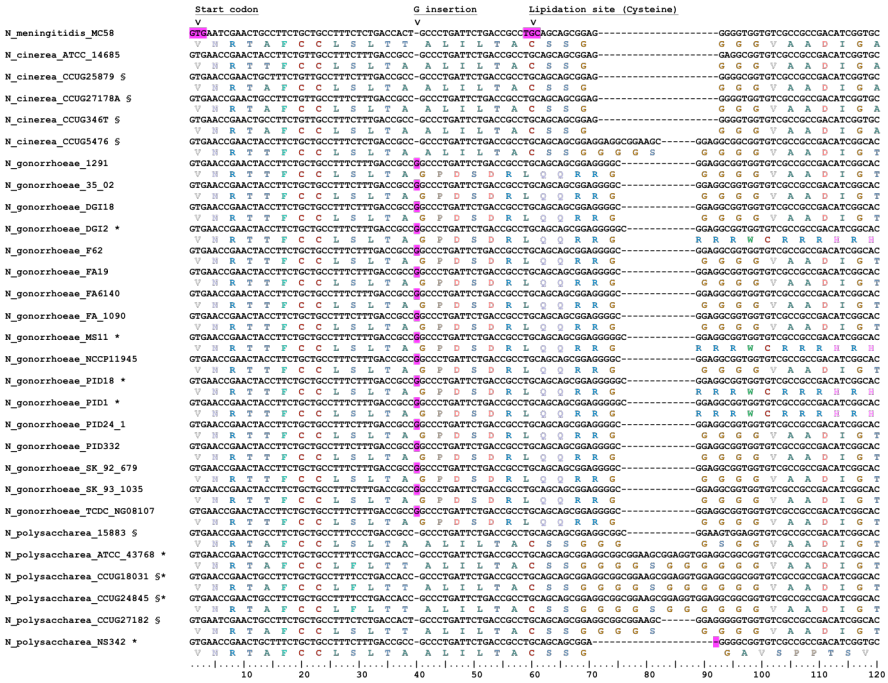


Figure 1.6 Multiple alignment of the N-term portion of the *fHbp* gene, where the major differences between gonococcal and meningococcal *fHbp* were concentrated. In both gonococcal and meningococcal strains the start codon was present in the same position, but the insertion of a single base (G) at position 40 causes a frame-shift in all gonococcal strains, which results in the loss of the lipo-box motif. The correct frame was re-established with a further eight-base insertion after position 70 [3]. In 4 cases (strains DGI2, MS11, PID18 and PID1) the insertion was of nine bases, and the correct frame of translation was not restored.

"putative opacity protein" on the opposite strand (Figure 1.7A). Finally, in most of the strains lacking a clear homologue of the *fHbp*, we observed the presence of weak similarity hits (between 30% to 40% similarity) against proteins that were mainly annotated as "fHbp proteins", or weaker hits (approximately 20% similarity) against proteins annotated as "hypothetical lipoproteins". Despite a conserved lipoprotein motif, these hits, located in a different genomic region, could not be aligned to the factor H binding proteins of *N. meningitidis*.

We then analysed a multiple alignment of a region of 10kbp including the *fHbp* gene, to identify homologous recombination events that contributed to its evolution (Figure 1.8). For clarity, a single representative for each group of strains with a similar pattern of recombination is shown, *i.e.* FA 1090 for gonococcus, CCUG27178A for *N. cinerea*, MC58 and M04-240196 for meningococcus *fHbp* variant 1, 8013 for variant 2 and M01-240355 for variant 3. A complete diagram including all the events on a selection of 22 representative sequences is shown in Figure 1.9. Two major events could be identified in Figure 1.8: i) *fHbp* variant 1 was imported into *N. meningitides* from *N. cinerea* in an event

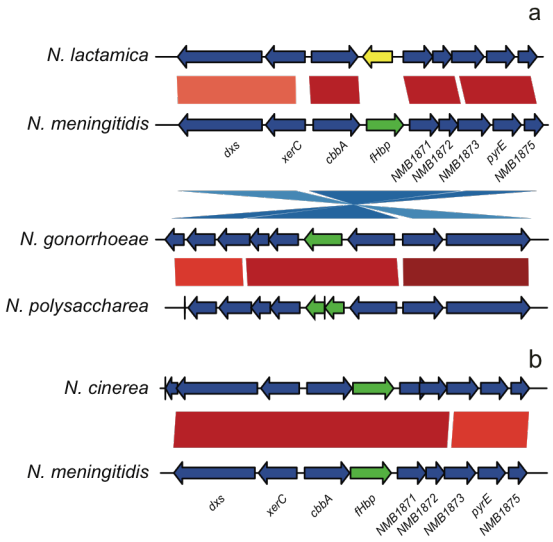


Figure 1.7 (A) Alignment of the *fHbp* locus in *N. lactamica* strain 020-06, *N. meningitidis* strain M01-240355, *N. gonorrhoeae* strain FA 1090 and *N. polysaccharea* strain ATCC 43768. The *fHbp* gene is indicated in green. In three cases (*N. meningitidis* M01-240149, WUE_2594 and Z2491) the intergenic region between the homologues of *NMB1869* and the *fHbp* showed the insertion of a highly conserved, AT-rich DNA fragment of 186-187 bp [2] corresponding to the direct repeat portion of the IS1106 (GenBank: Z11857.1) and responsible of the *fHbp* promoter displacement from its start codon. The entire locus is conserved in the reverse strand in all strains of *N. gonorrhoeae* and *N. polysaccharea*, where fHbp is always variant 3, while in both strains of *N. polysaccharea* the homologue of the *fHbp* gene is frameshifted. The *fHbp* gene is absent in all strains of *N. lactamica*, where it is substituted by a protein annotated as "putative opacity protein" in the opposite strand (in yellow). (B) *fHbp* locus in *N. meningitidis* MC58 and in *N. cinerea* strain ATCC 1468. In both cases the *fHbp* gene is variant 1.

spanning also the *cbbA* gene, that is located upstream of *fHbp*; ii) *fHbp* variant 2 originated by the substitution of a portion of a variant 3 by the homologous sequence from a variant 1 strain. The latter event could be seen clearly from a multiple alignment of the variant 2 and variant 3 *fHbp* sequences, shown in Figure 1.10. The putative donor was identified by RDP analysis as *N. cinerea* strain CCUG27178A, but, given the high similarity of the *N. cinerea* gene with the *N. meningitidis* variant 1 genes, we could not exclude that other undetected intermediate steps occurred. This complex pattern of recombination was probably the cause leading to the observed modular organization of the *fHbp* gene [45].

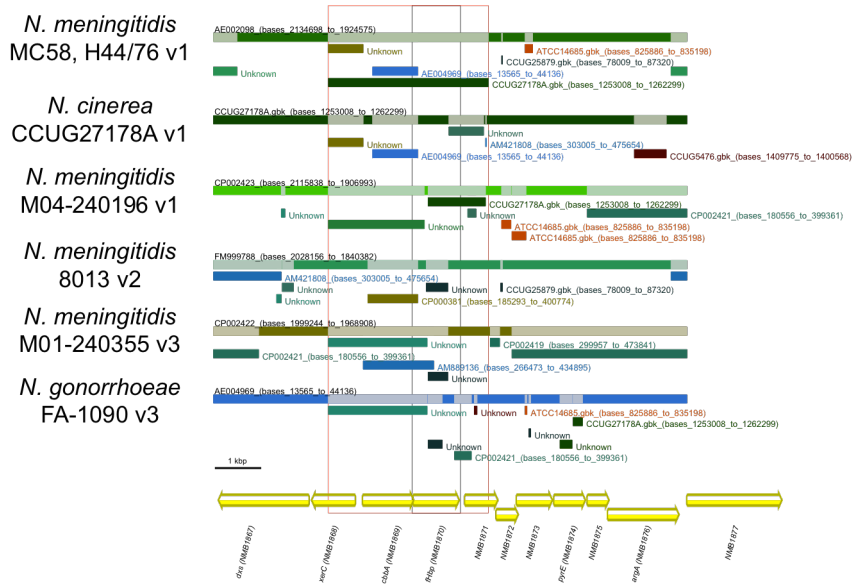


Figure 1.8 Schema of the recombination events spanning the *fHbp* gene. The position of the genes in the MC58 sequence are shown. The black box indicates the position of the *fHbp* gene. A single representative for each group of sequences showing a similar pattern of recombination events is reported. Each horizontal bar represents the locus in the genome indicated on the left. Light boxes drawn within each bar represent putative segments transferred by recombination. The imported fragments are represented by specific boxes that are drawn below each bar. The name of a putative donor strain is indicated on the right. The *fHbp* variant 1 (strain MC58) appears to have been imported into *N. meningitidis* by a recombination event including also the upstream *cbbA* gene, while the *fHbp* variant 2 was the result of the recombination of a small fragment of a variant 1 sequence on a variant 3 background.

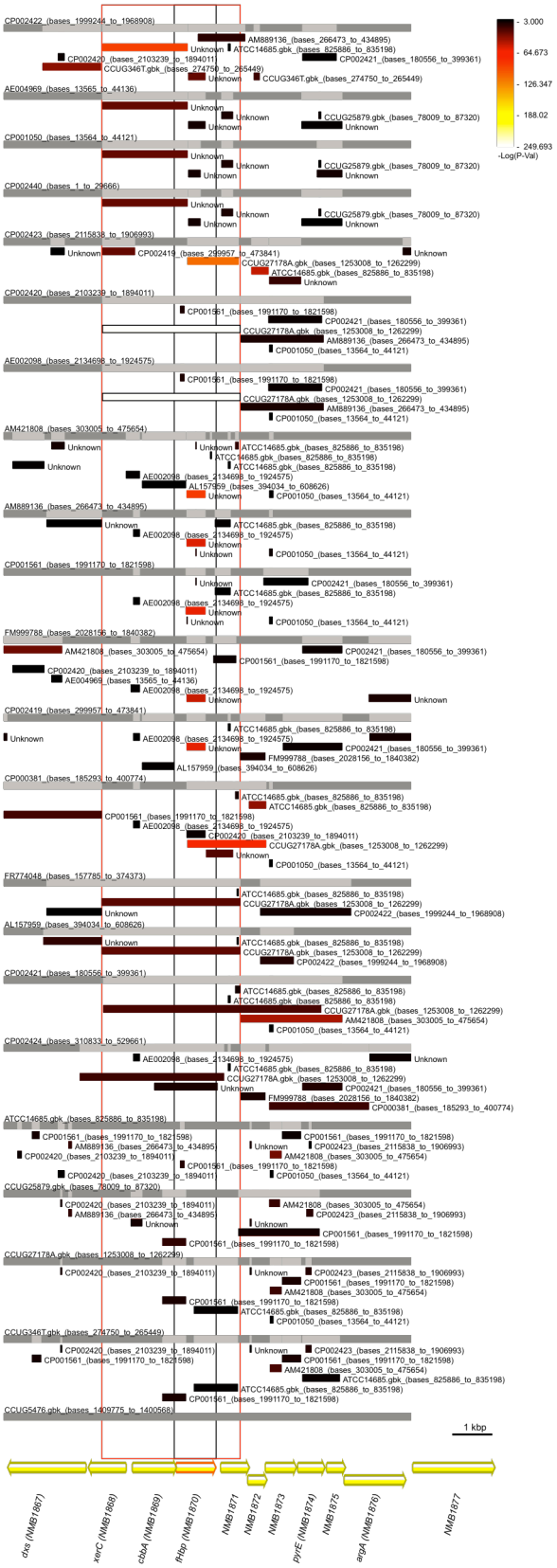


Figure 1.9 Schema of the recombination events in a region of 10kb including the *fHbp* gene. The black box highlights the position of *fHbp*.

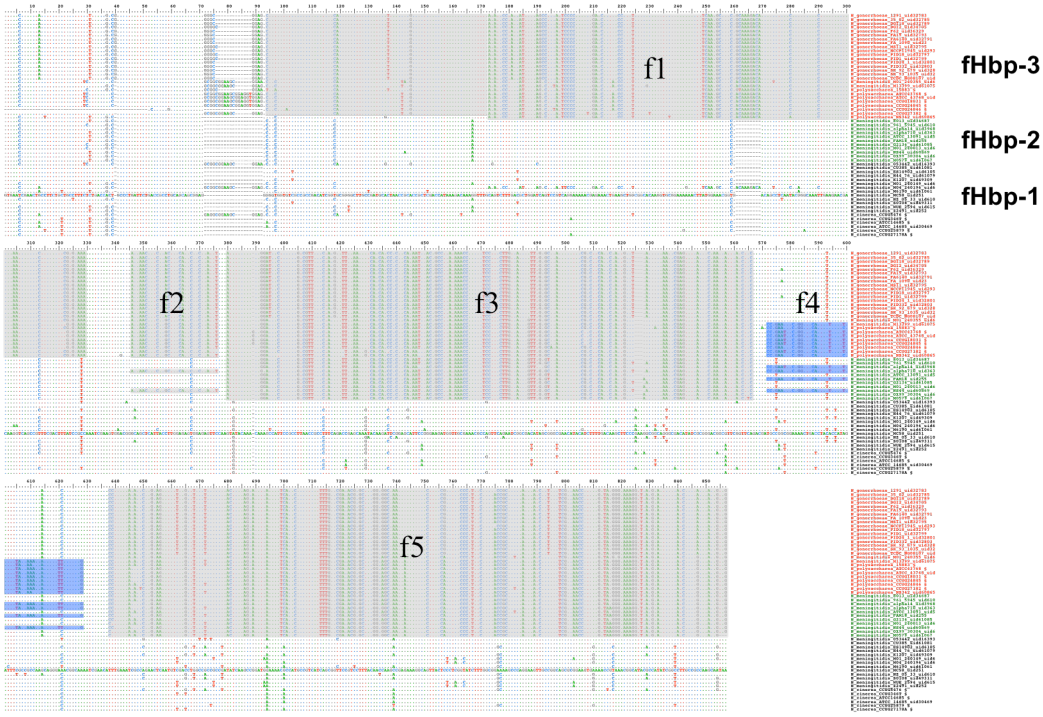


Figure 1.10 Multiple alignment of the fHbp variant 1, 2 and 3. Variant 2 sequences have originated by the substitution of a portion of a variant 3 fHbp by a foreign DNA sequence of 350 bp (length of the multiple alignment) from a variant 1 donor. Variant 1 isolates are labelled in black, variant 2 in green and variant 3 in red. The areas of the molecule variant 1-like are in white, the portions variant 3-like are highlighted in grey and the shared portion is in cyan. These different areas define 5 portions of the molecule (named f1-f5) that, by different patterns of recombination, compose all the fHbp sequences.

The structuring of the splits in a phylogenetic network of the *fHbp* sequences (Figure 1.11) was also clearly related to the modularity of the gene. Five different contiguous portions of the molecule could be highlighted in the *fHbp* sequence alignment (f1-f5 in Figure 1.10) that corresponded to the major splits in the network. f1+f2 were the major contributors of the split between variant 2 and 3, while f3+f5 determined the split between variant 1 and variant 2 and 3. All the portions segregated with variant 1 or 3 molecules (except f4 which defined a portion of the molecule under a very different pattern of recombination) and variant 2 molecules could be described as a chimera of variant 1 and 3 molecules, as previously suggested [45] and supported by our analysis of the recombination events spanning this region.

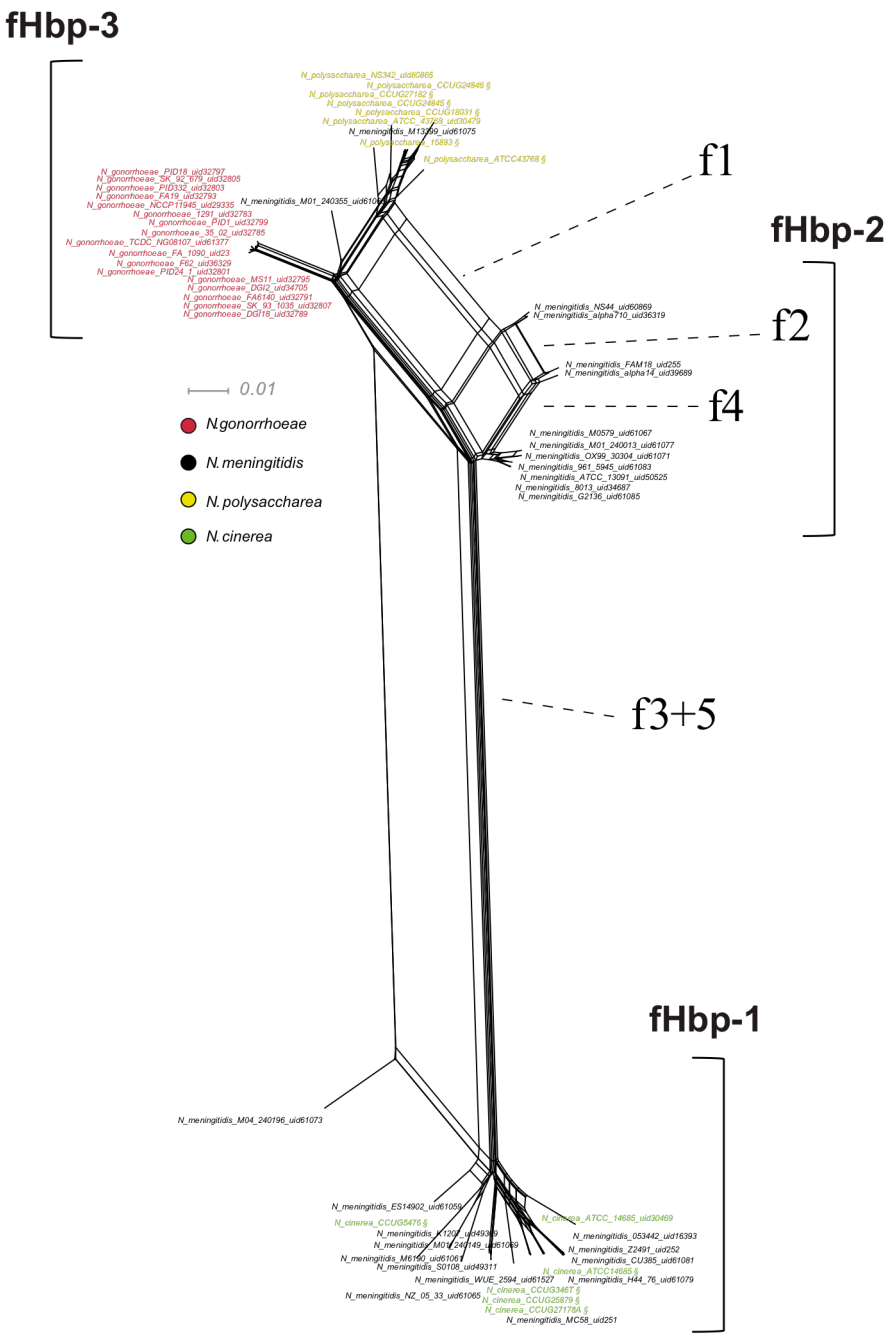


Figure 1.11 Splits network representation of the *fHbp* gene. The highlighted portions f1-f5 refer to the multiple alignment shown in Figure 1.10. Coloured labels indicate the different species of *Neisseria*.

1.4.3 Conservation of NHBA

Beside being present in all strains of *N. meningitidis*, conserved homologs of the NHBA coding gene were also found in *N. lactamica*, *N. polysaccharea, N. gonorrhoeae* and *N. flavescens* (Table 1.1). In all the species where NHBA was present, the entire locus harbouring the *nhba* gene was well conserved (). A highly conserved locus was also found in *N. flavescens* strain NRL30031 H210, but the NHBA coding gene was frameshifted. In the only other available sequenced genome of *N. flavescens* the gene was absent. However, since the latter sequence was not complete, definitive conclusions could not be drawn. In other species, where the *nhba* homolog was absent (Table 1.1), there was a very weak similarity hit, in any case inferior to 30% sequence identity.

While in most cases *nhba* sequences from different species did not segregate together in a phylogenetic tree (Figure 1.13), all gonococcal *nhba* molecules clustered in a branch of the tree that was well supported by bootstrap testing (94%). In the rest of the tree, although it was possible to identify small groups of closely related sequences with high statistical support, most inner branches had very low value of bootstrap. In addition, the conservation of the *nhba* sequences within *N. meningitidis* [4, 13, 39] was of the same order of magnitude as the conservation across the whole *Neisseria* genus. These results provided further support to the existence of a high rate of genetic exchange between the

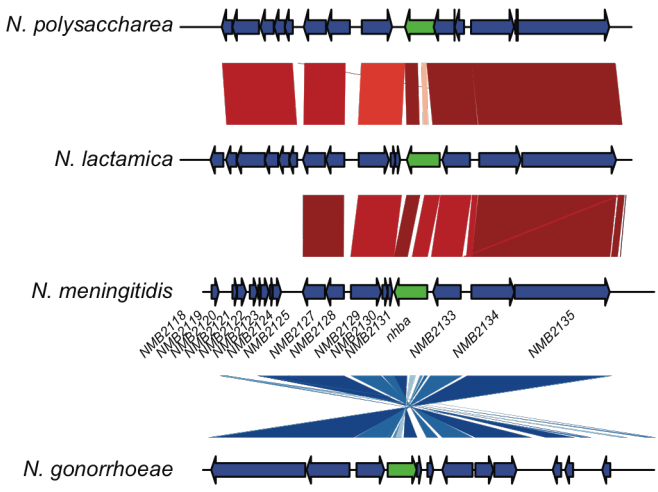


Figure 1.12 Alignment of the *nhba* locus in *N. lactamica* strain 020-06, *N. meningitidis* strain MC58, *N. gonorrhoeae* strain FA 1090 and *N. polysaccharea* strain ATCC 43768. The *nhba* gene is indicated in green.

different species that homogenize the distribution of the NHBA molecules within the *Neisseria* genus, with the possible exception of gonococcus.

A phylogenetic network analysis evidenced the influence of homologous recombination on the complex evolutionary history of this set of sequences (Figure 1.14), since it was impossible to separate phylogenetic clades corresponding to the species. The only exceptions were the strains of gonococcus that were grouped in a separated branch, although we found evidences of reticulate evolution also between the gonococcal branch and the other neisseriae. We also analysed the evidences of recombination events in the region including 10kbp upstream and downstream of the gene. In accordance with the network analysis of the gene sequences, the pattern of homologous recombination was very complex, and we found many distinct events overlapping the *nhba* gene (Figure 1.15). The level of conservation of two important functional motifs described for the NHBA, namely the lipobox motif (LXXC) and the Arg-rich heparin-binding motif (RFRRSARSRRS), was checked. The lipobox motif was conserved in all strains. The Arg-rich heparin-binding motif was well conserved in the central part (RSARSR). Where present, the three initial amino acids of the motif (RFR) were less conserved. In all Gonococcal strains, in one *N. lactamica* (strain Y92 1009), in *N. polysaccharea* and in three *N. meningitidis* strains (8013, alpha14 and N1568) a deletion of 7-8 amino acids (relative to the MC58 protein) was present, that included the three initial amino acids (RFR) of the heparin-binding motif. This region, located upstream of the heparin binding motif, is the cleavage site of the NalP protein [17] and was described as variable in gonococcus [14].

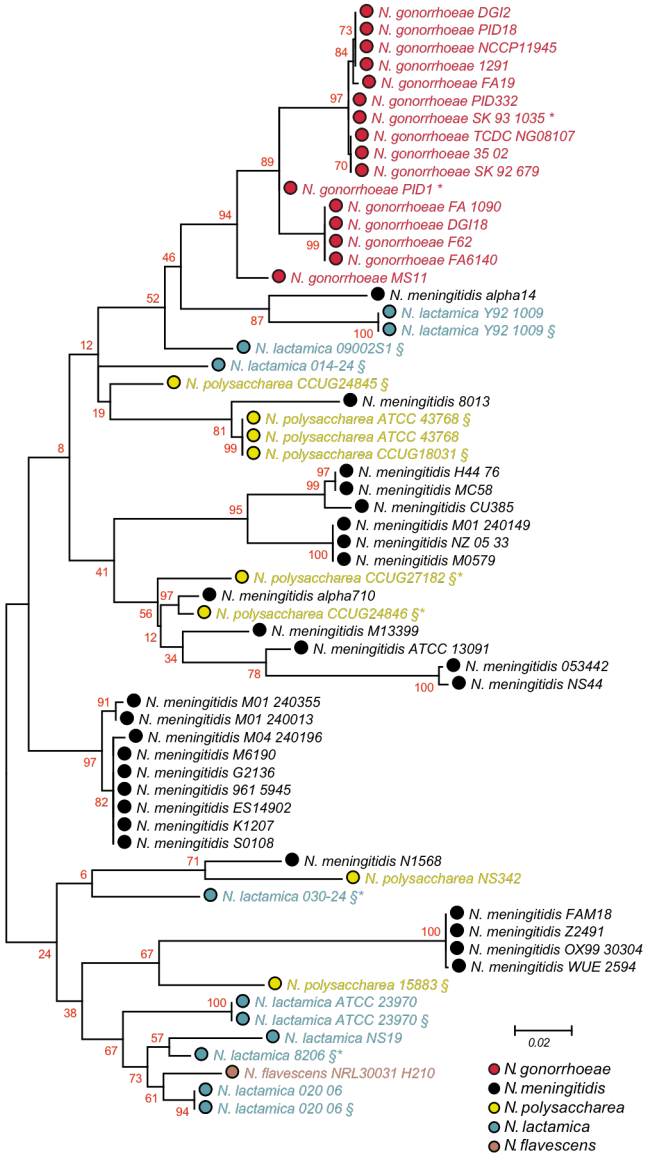


Figure 1.13 Maximum Likelihood phylogenetic tree of the *nhba* gene. The tree was obtained using the Kimura 2 parameters model with Gamma correction. The average gene variability was Pi = 0.084 (s.e. 0.004). Gonococcus formed a monophyletic branch closely related to the rest of the isolates. The other species, *N. polysaccharea*, *N. lactamica* and *N. flavescens* were scattered through the entire tree. Coloured circles and labels indicate the different species of *Neisseria*. The (§) symbol indicates sequences that were downloaded from the BigsDB database, otherwise from GenBank. The (#) symbol indicates genes that were at the border of a contig. Asterisks (*) indicate genes that were frameshifted.

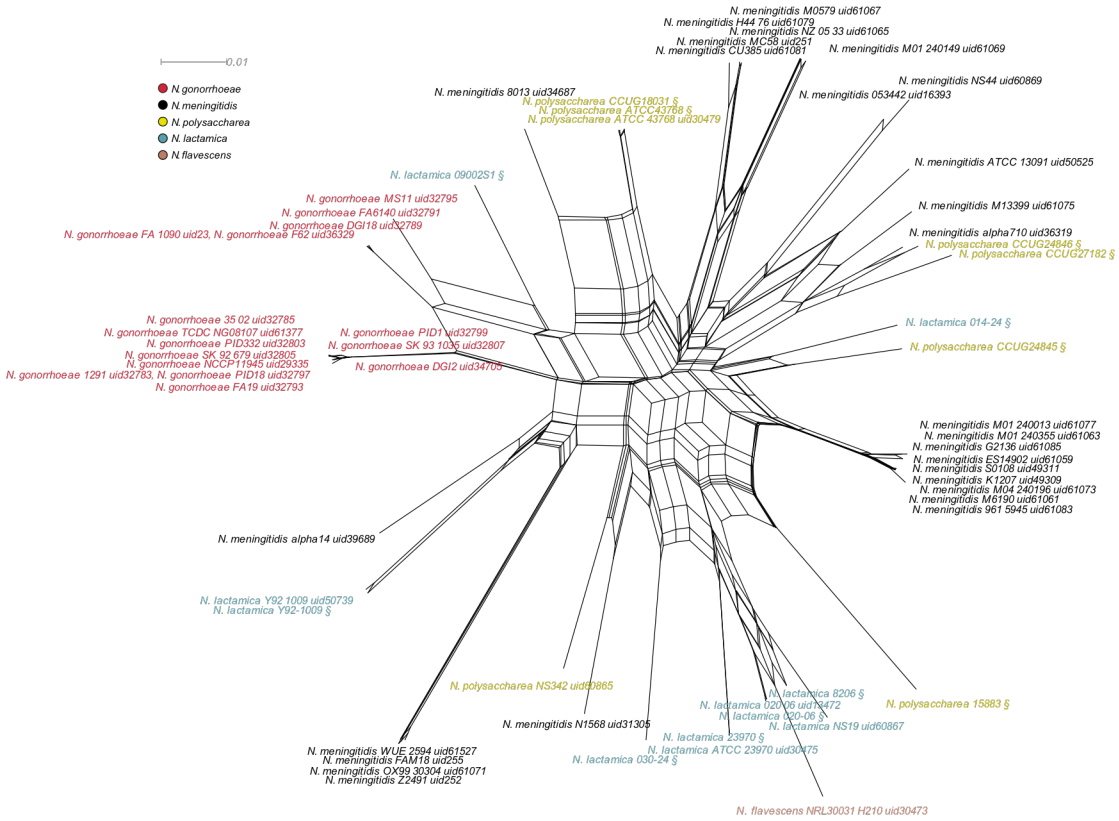


Figure 1.14 Splits network representation of the *nhba* gene. Coloured labels indicate the different species of *Neisseria*.

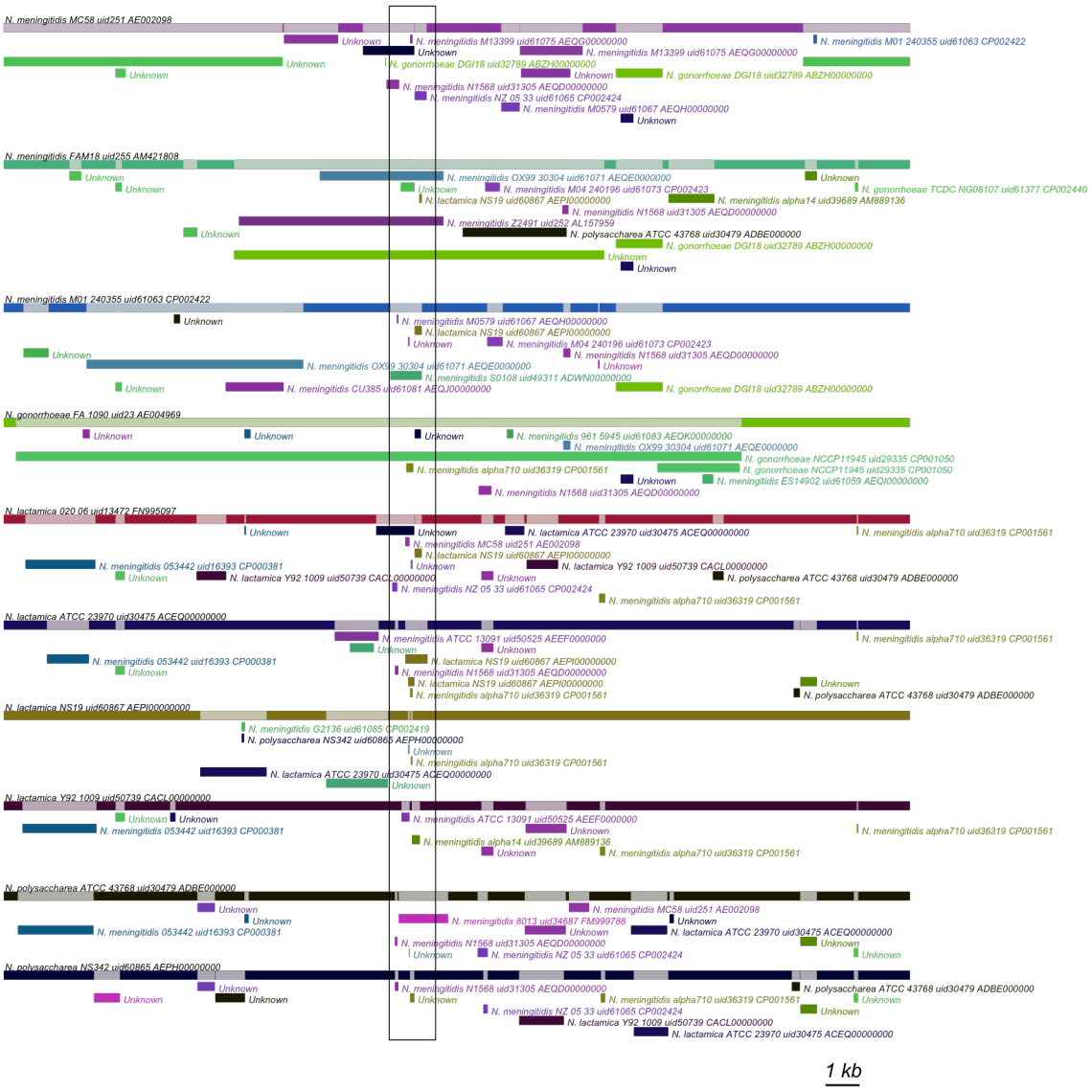


Figure 1.15 Schema of the recombination events in a region including 10kb of sequence upstream and downstream of the *nhba* gene. The black box highlights the position of *nhba*.

1.4.4 Conservation of NadA

The *nadA* gene was present in 12 of the sequenced strains of *N. meningitidis* (see Table 1.1). In two strains the coding sequence of the *nadA* homolog was interrupted by a premature stop codon introduced by a frameshift. In another strain (M6190) *nadA* was disrupted by the IS4 insertion. In two cases, the coding sequence of *nadA* was interrupted by the border of a contig, and the functionality of the gene could not be established.

We found no homolog of the *nadA* gene in *N. lactamica* and *N. polysaccharea*, although in those species, as well as in the *N. meningitidis* strains that do not harbour the gene, the locus was well conserved (Figure 1.16), suggesting that the *nadA* gene was the result of a recent insertion that left the neighbouring genes unchanged [21]. We found an intact homolog of *nadA* in one strain of *N. cinerea*, while in another strain we could reconstruct a *nadA* gene by joining two distinct contigs that contained, respectively, 850bp of the N-term and 600bp of the C-term of the molecule. However the integrity of the gene could not be established. In two *N. cinerea* strains, fragments of *nadA* could be found small contigs, but it was impossible to establish whether the gene was present. Finally, in one *N. cinerea* strain the locus was conserved but *nadA* was replaced by a gene coding for a small hypothetical ORF. When we used the concatenated pseudogene and the complete *nadA*

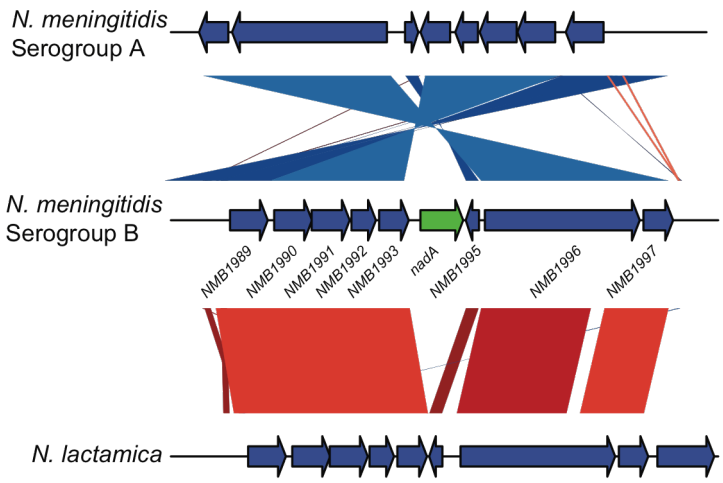


Figure 1.16 Alignment of the *nadA* locus in *N. meningitidis* of serogroup A strain Z2491, *N. meningitidis* of serogroup B strain MC58 e *N. lactamica* strain 020-06. The *nadA* gene is indicated in green. Although the locus is well conserved, the *nadA* gene is missing both in *N. meningitidis* of serogroup A and in *N. lactamica*. In *N. cinerea* ATCC_14685 the *nadA* gene is interrpupted by a contig edge, in *N. cinerea* CCUG346T the gene is intact. In both strains the gene is placed in the same locus of *N. meningitidis* of serogroup B strain MC58.

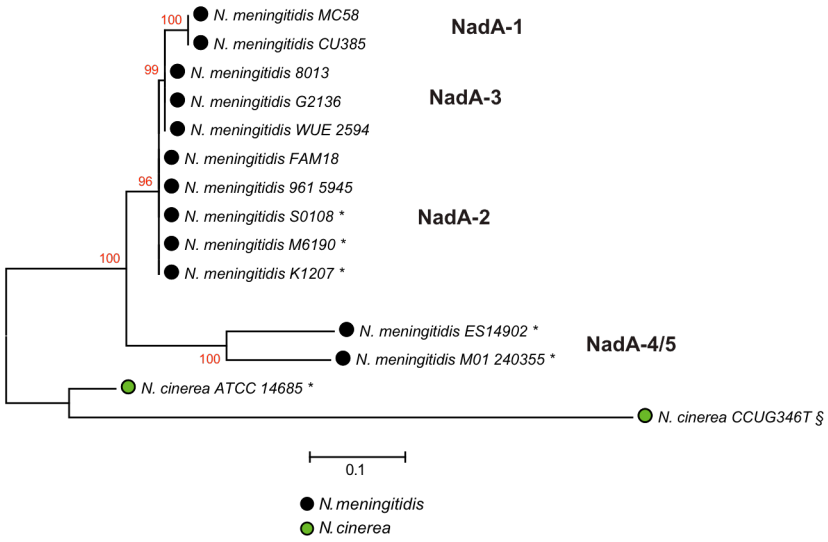


Figure 1.17 Maximum Likelihood phylogenetic tree of the *nadA* gene. The tree was obtained using the Kimura 2 parameters model with Gamma correction. The sequence from the *N. cinerea* ATCC 14685 strain was obtained by joining two fragments at the border of two distinct contigs, and it was therefore not possible to assess the integrity of the gene. Coloured circles and labels indicate the different species of *Neisseria*. The (§) symbol indicates sequences that were downloaded from the BigsDB database, otherwise from GenBank. The (#) symbol indicates genes that were at the border of a contig. The (+) symbol indicates genes that were interrupted by the IS4. Asterisks (*) indicate genes that were frameshifted.

from *N. cinerea* to build a phylogenetic tree (Figure 1.17), we found that the *N. cinerea* sequences were related to the *nadA-4/5* forms from *N. meningitidis* [13]. In all other species the locus was not conserved. In *N. lactamica*, *N. mucosa* and *N. sicca,* a weak similarity hit (less than 45% sequence identity) was present, but the homology was limited to a C-term portion of the molecule of about 120 aa, that is common to a family of YadA-like surface proteins (Table 1.1) and not specific of *nadA*.

The analysis of the presence of recombination revealed a limited number of events that related the formation of the meningococcal *nadA-4* and *5* molecules to a contribution from a donor similar to the *N. cinerea nadA* (data not shown).

1.4.5 Conservation of GNA1030

The *GNA1030* gene was present in all analyzed strains, with the exception of the draft genome of *N. wadsworthii* 9715. In *N. polysaccharea* ATCC_43768, the gene was interrupted at the border of a contig. In two strains of gonococcus (PID332 and SK_93_1035) the poly-A stretch following the ATG start codon introduced an Adenine that was responsible for a frameshift of the gene. In the other species the gene sequences were well conserved and the average gene variability of the whole *Neisseria* genus was Pi = 0.095 (s.e. 0.005). A phylogenetic tree based on the aligned sequences is shown in Figure 1.18.

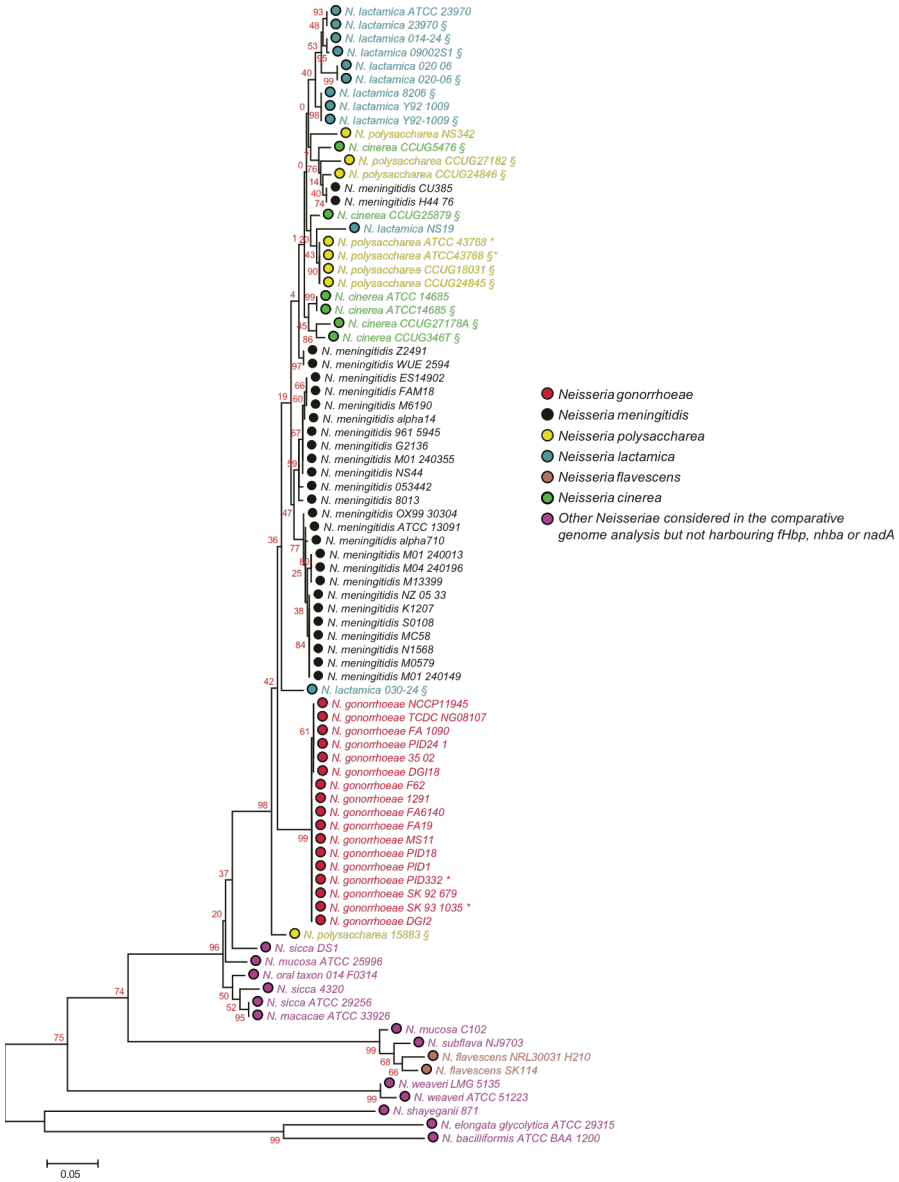


Figure 1.18 Maximum Likelihood phylogenetic tree of the *GNA1030.* Coloured labels indicate the different species of *Neisseria*. The (§) symbol indicates sequences that were downloaded from the BigsDB database, otherwise from GenBank. Genes that were frameshifted or at the border of a contig are indicated by asterisk (*).

1.4.6 Conservation of GNA2091

The *GNA2091* gene, predicted to code for a haemolysin, was present in all species. In both *N. weaveri* LMG_5135 and ATCC_51223 the gene was interrupted at the border of a contig. The predicted start codon (ATG) was not always conserved but was in some cases replaced by the GTG (Valine) or AAG (Lysine) codons. The gene was always in frame and well conserved in the *Neisseria* genus (Pi = 0.128, s.e. 0.005). The phylogenetic reconstruction of the molecule is shown in Figure 1.19.

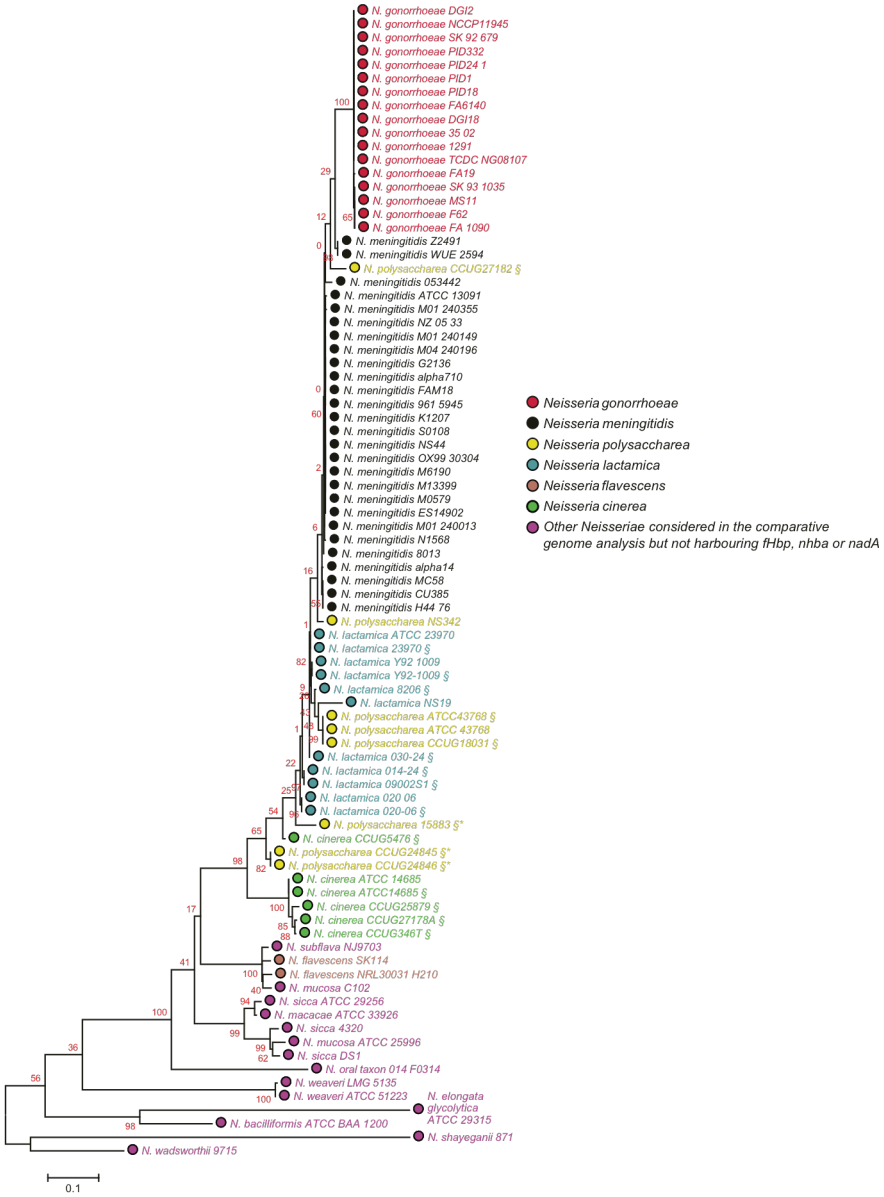


Figure 1.19 Maximum Likelihood phylogenetic tree of the *GNA2091.* Coloured labels indicate the different species of *Neisseria*. The (§) symbol indicates sequences that were downloaded from the BigsDB database, otherwise from GenBank. Genes that were frameshifted or at the border of a contig are indicated by asterisk (*).

1.5 Discussion and conclusions

The genus *Neisseria* includes a large group of bacteria, some of which are responsible for life-threatening diseases in humans, while others are mainly harmless colonizers. We have studied the distribution of *fHbp*, *nhba* and *nadA*, three antigens included in the 4CMenB vaccine, in the available genome sequences of the commensal and pathogenic neisseriae. We also studied GNA2091 and GNA1030 proteins, included in the 4CMenB vaccine as fusion proteins with fHbp (GNA2091-fHbp) and NHBA (NHBA-GNA1030). *fHbp* was present in one of its three major variants in all strains of *N. meningitidis*, in *N. gonorrhoeae* and in all commensal human neisseriae that have been sequenced to date, except *N. lactamica*. All strains of *N. gonorrhoeae* and of *N. polysaccharea* harboured a variant 3, while strains of *N. cinerea* harboured close homologs of a *N. meningitidis* variant 1. It was not possible to assess unambiguously in which species the protein evolved primarily, and all hypotheses involved at least one, but possibly two, events of inter-species recombination. Given the relatively lower recombination rate between *N. meningitidis* and *N. gonorrhoeae* [9], and the fact that a *fHbp* gene variant 3 exists also in *N. polysaccharea*, it is likely that a *fHbp* variant 3 was present in their most recent common ancestor. We also found evidences that the *fHbp* gene variant 1 evolved in *N. cinerea* and was introduced in *N. meningitidis* by a recombination event involving also the neighbouring *cbbA* gene. Finally, sequence analysis suggested that one more homologous recombination event has generated the variant 2 *fHbp* by inserting a variant 1 fragment on a variant 3 background.

A distribution similar to *fHbp* was found for *nhba*, which was absent from *N. cinerea*, while it was present in *N. lactamica.* A fully parsimonious pattern of inheritance was not compatible with the phylogenetic tree of the gene. Also, the gene was present in all strains of *N. gonorrhoeae*, and its sequences showed little variability within this species. This result was compatible with the overall low variability of the genomes of *N. gonorrhoeae* and suggested a recent diversification of the gonococcus from *N. meningitidis.* The *nhba* gene showed low variability within *N. meningitidis* [4, 13, 39], and sera of mice vaccinated with one variant were shown to be bactericidal against heterologous strains [3]. Interestingly, the heparin-binding motif was only partially conserved across all species where the gene was present. The site of cleavage of NHBA by NalP, situated upstream of the heparin binding region, was not conserved, suggesting a different protein post translation processing.

The gene coding for NadA is part of an independent genetic unit [21], which consists of a promoter region, the coding sequence and a terminator region. *nadA*, was present only in

a portion of the *N. meningitidis* strains, and was never found in any other commensal or pathogenic neisseriae, with the exception of two strains of *N. cinerea*. Analysis of the sequences immediately adjacent to this genetic unit showed that the segment was always flanked by direct repeats, indicating that a mechanism of recombination could have determined its distribution amongst the meningococcal lineages. Finally, homologues of GNA1030 and GNA2091 were present and well conserved in most isolates of the *Neisseria* genus.

Taken together, these findings highlight the complex evolutionary history of the human neisseriae. With the exception of *N. gonorrhoeae*, the human neisseriae are carried asymptomatically in the nasopharynx of healthy individuals. It has long been known that homologous recombination is the main process contributing to the genomic evolution of *N. meningitidis*, and that some antigens can be exchanged between pathogenic and commensal species [46]. We found that, when present, the sequences of *fHbp*, *nhba* and *nadA* from other species of the *Neisseria* genus do not introduce radically new variants when compared to the variability already present in *N. meningitidis*.

The distribution of the 4CMenB vaccine antigens in the commensal neisseriae suggests that vaccination could have an impact on the composition of the commensal flora. In particular, since the variability of *nhba* across the different neisseriae (*N. gonorrhoeae*, *N. lactamica* and *N. polysaccharea*) was similar to the intra-species variability within *N. meningitidis*, a recombinant vaccine based on *nhba* could have an impact on the species harbouring this protein. However, as recently shown for the 4CmenB vaccine [33], the ability of vaccine-induced antibodies to mediate killing depends on their ability to react with the variant/peptide present in the vaccine and on the amount of antigen expressed by a given strain. The latter might vary within and between species in response to yet unknown environmental factors, as recently demonstrated for NadA [34, 40]. Other factors, like surface accessibility must also be taken into account [13, 20, 47]. Therefore genetic typing is not sufficient to predict strain coverage and further studies will be needed to evaluate whether this vaccine will have any impact on commensal strains.

Chapter 2. Conservation of Meningococcal antigens in carrier and pathogenic isolates of *N. meningitidis* collected in the Czech Republic in 1993.

2.1 Abstract

The differences between *Neisseria meningitidis* strains isolated in healthy individuals and in people affected by meningococcal infection is of great interest for molecular epidemiology, in particular to understand the etiology of meningococcal disease and its relationships with genetic characteristics of meningococci. In this work, we analysed and compared the genetic variability of three surface antigens, main components of the 4CMenB vaccine, in all pathogenic isolates collected in the Czech Republic in 1993 and a larger sample of carriage strains collected over the same time period.

The genetic variability of three meningococcal antigens, fHbp, NadA, and NHBA was evaluated in 266 strains isolated in the Czech Republic. Full gene sequences were assessed for *nhba* and *fHbp*. *nadA* gene presence was determined by PCR amplification. The *nhba* and *fHbp* sequences, and the *nadA* presence-absence genotypes, were compared between isolates derived from healthy individuals and from disease cases. Also, any existing relationship between the diversity of each protein and neisserial clonal complexes was investigated. In particular, fHbp diversification was studied within the ST-41/44 clonal complex structure, where carriage and disease isolates were peculiarly distributed.

Both fHbp and NHBA showed a high degree of relatedness with the structuring in clonal complexes typical of the neisserial population and a diversified distribution in strains from healthy subjects or disease cases. In particular, the fHbp main variant 2 was particularly associated with carriage as well as specific NHBA peptide sequences. NadA was confirmed to be present mostly in pathogenic strains, rather than in carrier strains. Interesting differences in the antigenic repertoires between carriage and pathogenic strains were found in almost all clonal complexes: ST-41/44, ST-92, ST-106 and ST-116 displayed antigenic repertoires that appeared to be associated with carriage only.

2.2 Introduction

Neisseria meningitidis (*Nm*) is a human-specific pathogen that causes sudden invasive disease that can result in permanent disability or death in otherwise healthy people. Despite this association with such devastating disease, *Nm* often colonizes the oropharyngeal mucosa of humans without affecting the host. Normal colonisation of the oropharynx, commonly called carriage, occurs with variable frequency in healthy individuals, typically between 1 to 10%, and can rise to 100% in social communities living in close contact such as military recruits or students [48]. During asymptomatic carriage, the meningococcal colonisation is a complex state and possibly, it is not limited to the mucosal surface but also extends itself to the tonsillar tissue [49]. However, bacteria that cross epithelium and escape from host defences may replicate rapidly and spread into the blood, causing septicaemia, or reach and cross the vascular endothelium of the brain, infecting the meninges and the cerebrospinal fluid. Different complex mechanisms permit *Nm* to escape from the immune system, by reducing recognition of the bacterial surface by the host (i.e. the presence of the capsule, or other surface exposed phase-variable proteins), by redirecting immune response (i.e. producing outer membrane vesicles or OMV carrying important immunodominant proteins), or by inhibiting the host response, for example by reducing the complement cascade activation.

For mechanisms that are directly related to specific molecular components of the bacterium, e.g. the capsule or specific surface-exposed proteins, the monitoring of the molecular variability among the isolates can help in understanding the differences during carriage and disease (if any), both functionally and epidemiologically. Such observations become more important when considering strategies for developing vaccines against *Nm*.

Vaccines against meningococcal serogroups A, C, W-135 and Y are based on capsular polysaccharide, a strategy that fails against Serogroup B (MenB), because its capsule, a homopolymer of α2-8-linked sialic acid, resembles a human neural cell adhesion molecule, NCAM-1 [50]. As a consequence, the B capsule is poorly immunogenic, and the use of the capsule as a vaccine component, could raise an autoimmune response. Earlier MenB vaccines based on OMV successfully limited clonal outbreaks in Cuba, Norway, and New Zealand [51-53]. OMVs include surface exposed proteins such as PorA, which is the immunodominant antigen. PorA is structurally hypervariable, it has a phase-variable expression and shows only limited cross reactivity between variants [51, 52, 54], which limits the use of these vaccines.

Current MenB vaccine development relies on recombinant surface protein fusions and combinations. The 4CMenB vaccine (approved in Europe under the commercial name of

Bexsero®) [1] is a multicomponent formulation including three recombinant proteins: factor H binding protein (fHbp; genome derived Neisseria antigen (GNA)1870; lipoprotein (LP)2086) [3, 10, 39], *N. meningitidis* adhesin A (NadA; GNA1994) [22] and Neisserial Heparin Binding Antigen (NHBA; GNA2132) [55]; and the outer membrane vesicles (OMVs) derived from strain NZ98/254 [53].

fHbp is a surface-exposed lipoprotein that binds human factor H, a major negative regulator of the complement alternative pathway [10], and it allows meningococci to evade innate host defences. The *fHbp* gene is typically present in meningococcal isolates, with three main protein variants (1–3) , that exhibit a limited cross-reactivity [3, 39].

NHBA is a lipoprotein [56] able to bind heparin *in vitro* [55]. The gene is present in all meningococcal isolates analysed so far [18, 39, 57] and despite the sequence diversity observed, NHBA induces bactericidal antibodies that are cross-protective against most strains tested [1].

NadA belongs to the Oligomeric Coiled-coil Adhesin class of trimeric proteins [21]. The *nadA* gene occurs in approximately 50% of pathogenic MenB strains, but in nearly all strains of the hypervirulent clonal complexes ST-32, ST-11 and ST-8 [22]. NadA was described to have 5 molecular forms, grouped in two main groups, with the form NadA-4 generally associated to carrier strains [22].

Each of these proteins has shown immunogenicity in animal studies and clinical trials [1]. A vaccine based on the fusion of two different variants of the fHbp protein is also under development [15]. Meningococcal epidemiology is commonly monitored by molecular epidemiological tools like Multilocus Sequence Typing (MLST) [58], which is based on the sequencing of 7 specific housekeeping genes. MLST assignment is used to define groups (clonal complexes, cc) of strictly related isolates derived from a common ancestor. MLST clonal complexes have aided in the recognition and epidemiological monitoring of specific hyperinvasive meningococcal lineages. Other technique, like determination of capsular group, of strain subtype (PorA variable regions), or FetA sequencing [59] [60] can also be used to complement MLST typing [58]. More recently, sequencing of *fHbp*, has been introduced as an additional molecular tracer of meningococcal strain diversity [61].

In this study, we analysed and compared the presence and the genetic variability of the fHbp, NadA and NHBA proteins in all meningococcal isolates belonging to a Czech Republic collection, which is a reference for epidemiological studies on carriage [62-66]. In particular, we analysed all carriage strains isolated in 1993, together with all isolates causing a disease event during the same year. The association of specific clonal complexes with carriage or disease has been observed in the same strain panel [63]. We further

investigated this aspect, by looking at the sequence variability of fHbp in carrier strains belonging to clonal complex ST-41/44, and compared the results with the ones from *Nm* isolates belonging to the same clonal complex from other epidemiological studies. In total 255 strains of clonal complex ST-41/44, from carriage and from disease, were analysed.

2.3 Methods

2.3.1 Isolates panel

A systematic carrier survey was performed in the Czech Republic. Meningococci were isolated from throat swab specimens in the 1970s (610 isolates), 1980s (99 isolates), and 1990s (345 isolates) [65, 66]. The samples were from volunteers of all age groups, attending schools and workplaces from locations covering all over the country. We selected for the analysis all meningococcal isolates (213 carriage and 53 disease) collected during the year 1993.

Swabs were inoculated onto Thayer-Martin selective medium and incubated at 37°C in 5% CO_2 for 48h. Resistant colonies were subcultured onto heated blood Mueller-Hinton agar. Species confirmation was performed by Gram staining, oxidase reaction, and with the Neisseria 4H system (Sanofi Diagnostics Pasteur) or API NH system (bioMe´rieux) panels of biochemical tests. No sampled individuals had known contact with patients with invasive meningococcal disease. Chromosomal DNA was extracted from stored samples by using the Isoquick Nucleic Acid Extraction Kit (Orca Research).

To study the variability and distribution of fHbp, within clonal complex ST-41/44 (cc41/44), we composed for comparison a cc41/44 strain panel including: 32 strains collected in Czech Republic, in 1993; 14 isolates belonging to the 107 MLST Reference disease/carriage strains database [58, 61]; 17 invasive strains collected in Sweden, in 2000/2001 [18]; 171 cc41/44 invasive isolates collected in England and Wales, in 2007/2008 [5]; 21 invasive isolates from the Novartis reference strain panel, collected worldwide [39]. As for a meningococcal population analysis, distribution of 2492 cc41/44 disease/carriage isolates was evaluated from MLST database (updated on January 2012).

2.3.2 Serogrouping and molecular typing

All strains were genetically typed by MLST following standards procedures described on the Neisseria MLST website [67].

All Czech Republic strain capsular groups and sub-types were determined serologically. Capsular group was also confirmed by sequencing the *siaD* gene of 153 strains. For strains with conflicting serogroup results between these methods, we considered only data obtained by sequencing. All strains were also characterized by sequencing variable regions (VR) 1 and 2 of the *porA* gene [60] and the variable region of the *fetA* gene [59]. fHbp, NHBA and NadA genes, the latter in a limited number of cases only, were amplified and sequenced as described by Bambini *et al.* [39]. *fHbp* gene sequences of strains belonging to the 107 MLST strains database [61], to the Swedish [18], to the England and Wales [5], and to the Novartis [39] collections were already available.

Sequences were assembled and analysed using the Vector NTI Suite 9. Unique identifiers were assigned to every allelic and protein variant of each protein and sequences were deposited into the Neisseria MLST website [58]. In particular fHbp subvariants were designated as variant.oxford_peptide, NHBA subvariants were designated as peptide and NadA subvariants were designated as variant.peptide.

2.3.3 Data Analysis

Odds ratio (OR) association statistics, with the support of the Mantel-Haenszel test and the Wolf's test for sample heterogeneity, were calculated as implemented by the function *meta.MH* of the *rmeta* package of *R* [68]. Fisher's exact test to evaluate non-random association was calculated with the function *fisher.test* of the *stats* package of *R*. The strength of the association was measured with the Cramer's V index with *R* package *VCD*.

Molecular phylogenetic trees were inferred from protein sequences by using the Maximum Likelihood (ML) method based on the JTT matrix-based model, as implemented by MEGA5 [28], after performing sequences alignment with *muscle* [69].

The structure of cc41/44 was analysed with Phyloviz [70] and a minimum spanning tree was built allowing 5/7 conserved MLST loci. The complete collection of isolates from the MLST database was used. Complex founder was assigned automatically to ST-41. ST profiles were subdivided in sub-complexes ST-41, ST-44 and singlets on the basis of the diagram.

2.4 Results

2.4.1 Czech Republic carriage/disease strain panel

The meningococcal strains isolated in Czech Republic during the year 1993 and selected for this study were derived from healthy individuals or from disease cases. These two distinct groups of isolates showed different epidemiological traits. For example, isolates collected from healthy carriers and from disease cases were differently distributed with respect to the age groups of the individuals as shown in Figure 2.1. In particular, all isolates from infants (0-11 months), young children (1-4 years), and the adults (>25 years) were disease associated. In contrast, the majority of the isolates in the 15-24 years age group were carriage associated and for the 20-24 years age group the carriage/disease odds is significantly associated with carriage (p-value < 0.05, see the complete odds ratio statistics in Table 2.1).

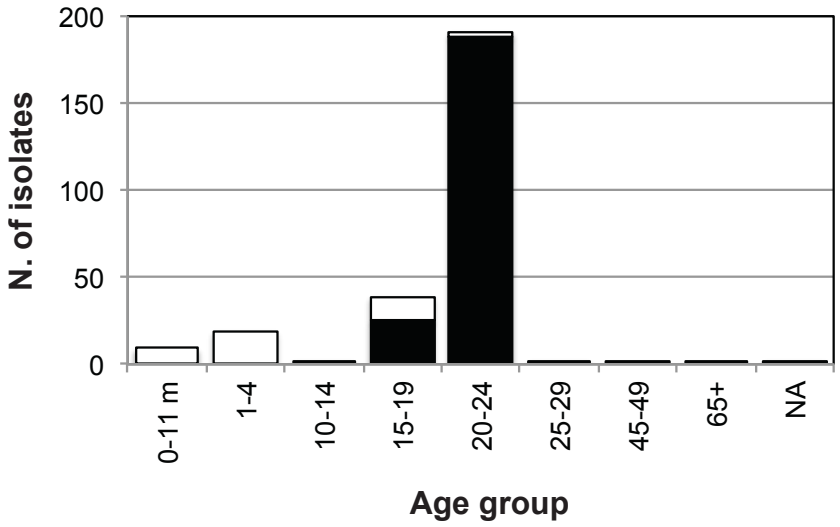


Figure 2.1 Epidemiology of the isolates: age group distribution, black bars represent isolates from healthy carriers, white bars represent isolates from disease cases, isolates lacking of individual age information are indicated by (NA).

Age group	carriage	disease	Total	Odds ratio	(95% CI)	FET* p-value
0-11 m	0	9	9	0.00	0.00 - 0.11	3E-07
1-4 y	0	19	19	0.00	0.00 - 0.04	2E-15
10-14 y	0	2	2	0.00	0.00 - 1.31	0.039
15-19 y	27	14	41	0.37	0.17 - 0.85	0.015
20-24 y	190	3	193	121	35 - 652	2E-31
25-29 y	0	2	2	0.00	0.00 - 1.31	0.039
45-49 y	0	1	1	0.00	0.00 - 9.7	0.2
>65 y	0	1	1	0.00	0.00 - 9.7	0.2
NA	0	2	2	0.00	0.00 - 1.31	0.039
Total	213	53	266	Mantel-Haenszel OR =1	0.73 - 1.37	Test for heterogeneity : X^2 = 63.5 (p-value 0)

Table 2.1 Age group distribution of the isolates, odds ratios and Fisher's exact test (FET) statistics

Another important example is the distribution of isolates by serogroup and clonal complex as shown in Figure 2.2 and Table 2.2 respectively. Most serogroup B and C isolates were collected from disease cases with significant odds statistics while all non-groupable strains were associated with carriage, reflecting the importance of the capsule for causing invasive disease (see the complete odds ratio statistics in Table 2.3)). A relevant portion of non-groupable isolates (72/88) were tested by sequencing the *siaD* gene, to determine genetically the type of capsule originally associated to the isolate. In 21 cases the isolate was of serogroup B, 6 isolates of serogroup C and for the remaining 45 isolates was not possible to determine the type by the sequencing of *siaD*. Clonal complexes ST-41/44, cc92, and cc106 were only associated with carriage, whereas cc11, cc18 and cc32 were selectively isolated from disease. Fifty out 53 strains belonging to the cc11 are serogroup C and they have been reported as an important outbreak in Czech Republic during 1993 [71]. Carriage associated strains belong to a wider range of clonal complexes and they appear more genetically variable than disease strains.

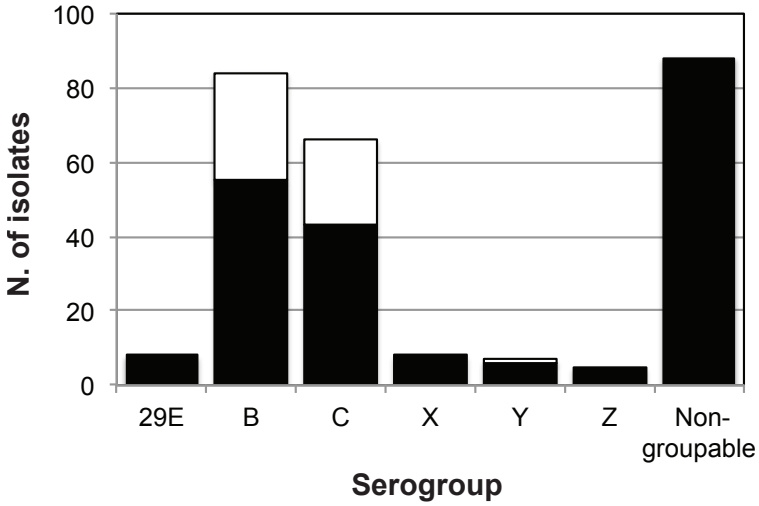


Figure 2.2 Serogroup distribution of meningococcal strains isolated in Czech Republic during the year 1993. Black bars represent isolates from healthy carriers, white bars represent isolates from disease cases.

Clonal complex	Carriage	disease	Total	Odds ratio	(95% CI)	FET* p-value
ST-11 complex/ET-37 complex	32	20	52	0.29	0.14 - 0.61	0.0004
ST-41/44 complex/Lineage 3	31	1	32	8.82	1. 4 - 367	0.008
ST-92 complex	21	0	21	Inf	1.38 - Inf	0.01
ST-106 complex	19	0	19	Inf	1.22 - Inf	0.02
ST-116 complex	13	0	13	Inf	0.77 - Inf	0.08
ST-18 complex	5	8	13	0.14	0.033 - 0.50	0.0009
ST-53 complex	11	0	11	Inf	0.63 - Inf	0.13
ST-549 complex	5	0	5	Inf	0.23 - Inf	0.59
ST-103 complex	4	0	4	Inf	0.16 - Inf	0.59
ST-231 complex	3	1	4	0.74	0.058 - 40	1
ST-32 complex/ET-5 complex	1	3	4	0.08	0.0015 - 1.0	0.026
Other and NA	68	20	88	0.77	0.40 - 1.5	0.42
Total	213	53	266	Mantel-Haenszel OR =1	0.72 - 1.39	Test for heterogeneity : X^2 = 34.3 (p-value 0)

Table 2.2 Clonal complex distribution of the isolates, odds ratios and Fisher's exact test (FET) statistics.

Serogroup	carriage	disease	Total	Odds ratio	(95% CI)	FET* p-value
29E	8	0	8	Inf	0.43 - Inf	0.36
B	55	29	84	0.29	0.15 - 0.56	0.0001
C	43	23	66	0.33	0.17 - 0.66	0.001
X	8	0	8	Inf	0.42 - Inf	0.36
Y	6	1	7	1.50	0.18 - 70.6	1
Z	5	0	5	Inf	0.23 - Inf	0.59
Non-groupable	88	0	88	Inf	9.37 - Inf	4E-11
Total	213	53	266	Mantel-Haenszel OR =1	0.71 - 1.41	Test for heterogeneity : X^2 = 27.7 (p-value 0)

Table 2.3 Serogroup distribution, odds ratios and Fisher's exact test (FET) statistics.

2.4.2 Variability of fHbp, NadA, and NHBA

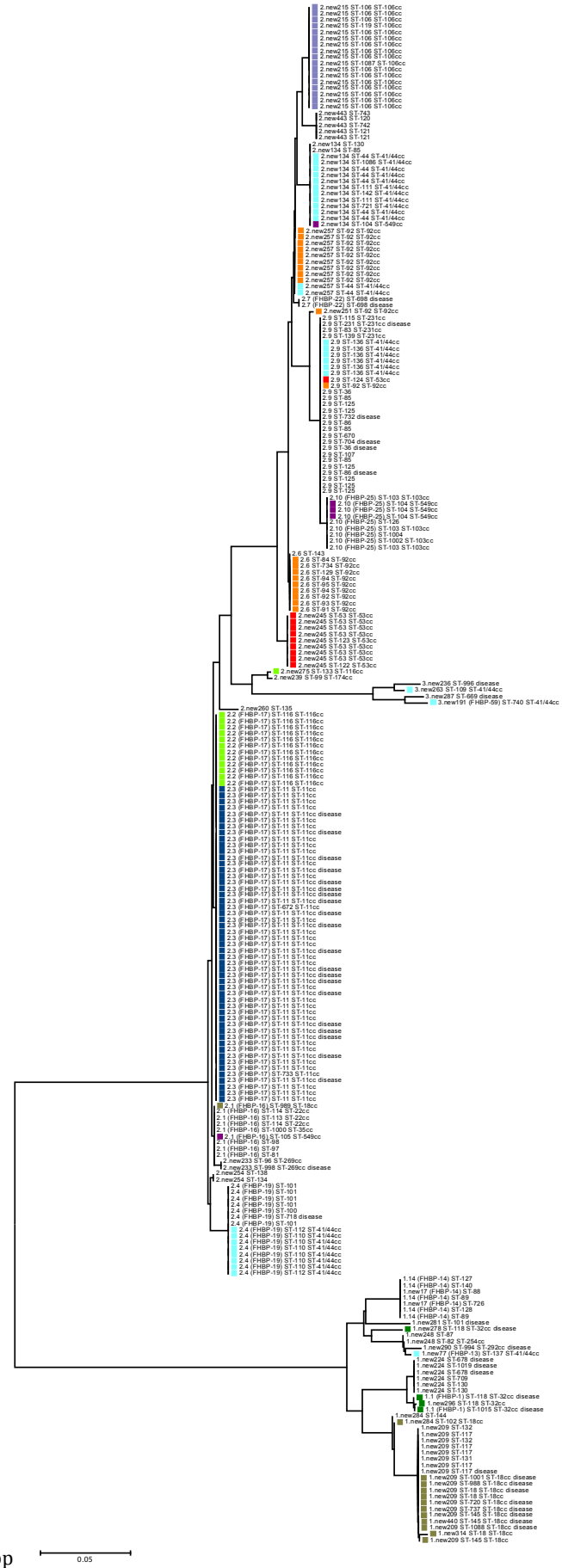
The comparison of the molecular variability of fHbp, NadA, and NHBA were consistent with the results shown above. A higher degree of variability was observed in the carrier strain panel, as compared to the clinical isolates, particularly for fHbp. NadA, was more rarely present in carrier strains, as compared to clinical isolates. Some newly described STs were found to harbour NadA; however no new NadA variants were discovered.

2.4.3 fHbp

2.4.3.1 Distributions of variants and significant associations with cc

fHbp was present in all isolates from Czech Republic, and showed the described structure in three molecular variants (fHbp-1, fHbp-2 and fHbp-3 [3, 39]) as represented by the molecular phylogeny reconstruction shown in Figure 2.3. In total, 259/266 genes were sequenced, obtaining 32 unique fHbp subvariants. Of these fHbp protein sequences, 10 belonged to the variant 1 (45 isolates), 19 to variant 2 (211 isolates) and 3 to variant 3 (3 isolates). The most frequent subvariants, present in more than 10 isolates, were fHbp-2.17, 2.24, 1.37, 2.286, 2.19, 2.34, 2.101 and 2.21. They represented the 70% (187/266) of the strains. fHbp-2.17 and 1.37 predominated in disease; fHbp-2.24 was evenly distributed. Other subvariants appeared to be exclusively (fHbp-2.286, 2.34, 2.101 and 2.21) or predominantly (fHbp-2.19) present in carriage strains only, suggesting to be typical of carriage. Some of them, fHbp-2.17, 2.286 and 2.34, were also particularly rare in other collections of pathogenic isolates [4, 5]. Each fHbp subvariant seemed to be related to a specific clonal complex (for example fHbp-2.17 to cc41/44, fHbp-2.286 to cc106,

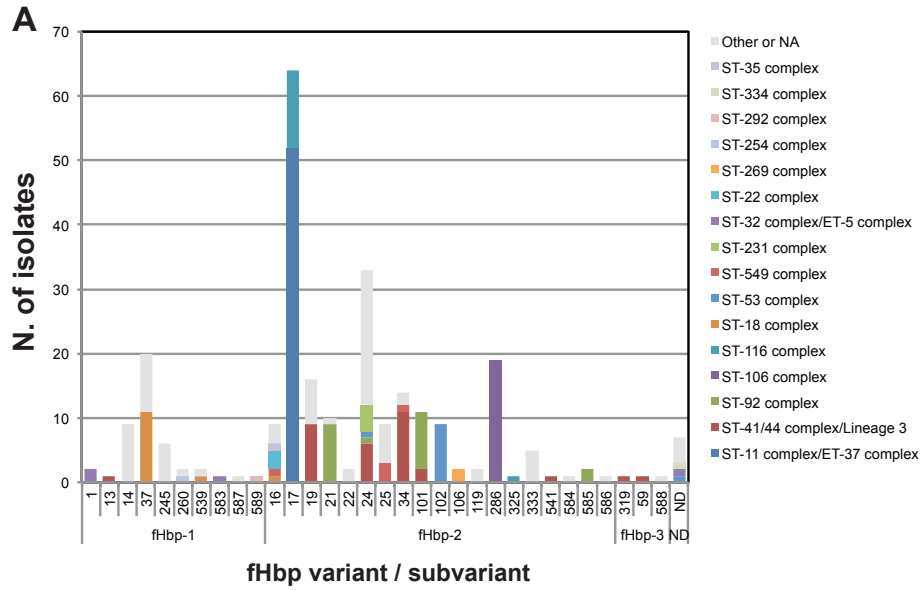
fHbp-1.37 to cc18). In most clonal complexes, one of the three variants of fHbp predominated (A) (index of association V_{fhbp} = 0.739). This relationship could be responsible of relevant associations with phenotypes or epidemiological traits of the clonal complexes, *e.g.* higher capability of inducing disease. In the Czech sample considered, fHbp-2 strains predominated in almost all clonal complexes; whereas cc18 and cc32 showed a significant proportion of fHbp-1 strains. Conversely, the isolates from disease were mainly from cc11, cc18 and cc32. A relevant association of fHbp-2 with carriage was observed, whereas fHbp-1 was found more frequently associated with disease (Table 2.4). This relationship could be related to the particular characteristics of the strain sample, which was geographically and temporally limited. Though, is worth of note is that similar observations were included in a carriage study performed by Harrison *et. al,* in the United States [72].

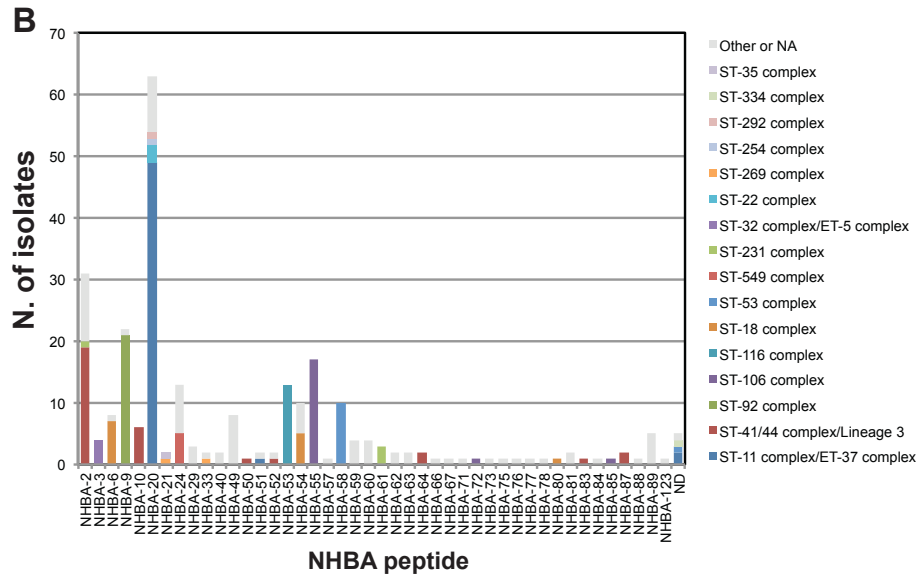


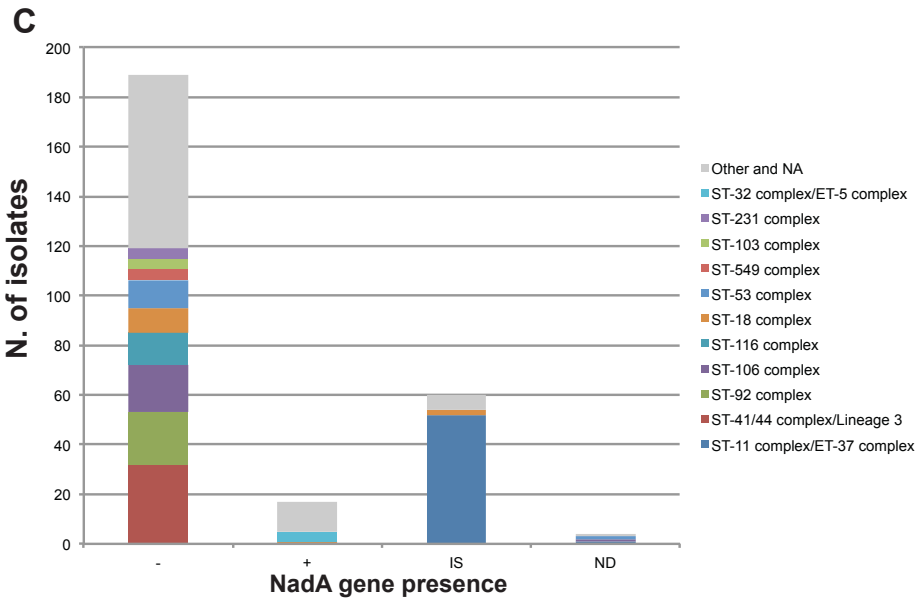
fHbp

(see next page)

Figure 2.3 Molecular phylogeny of fHbp. Molecular evolutionary relationships were inferred using the Maximum Likelihood method. The evolutionary distances were computed using the JTT matrix-based method and the units are number of amino acid substitutions per site. The 9 most represented clonal complexes were coloured (cc11, cc41/44, cc92, cc106, cc116, cc18, cc53, cc549, cc32). All protein subvariants-peptides are strongly associated with clonal complexes.







(see next page)

Figure 2.4 Vaccine antigens distribution among clonal complexes. Each protein subvariant / peptide was represented for both fHbp (A) and NHBA (B). The presence or absence of the gene was considered for NadA (C), a special case is the gene interrupted by IS1301, typical of cc11.

fHbp variant	carriage	disease	Total	Odds ratio	(95% CI)	FET* p-value
1	29	17	46	0.33	0.16 - 0.72	0.003
2	179	32	211	3.43	1.67 – 7.01	0.0004
3	1	3	4	0.08	0.00 - 1.00	0.025
ND	4	1	5	0.99	0.09 – 49.9	1
Total	213	53	266	Mantel-Haenszel OR =1	0.66 - 1.52	Test for heterogeneity: X^2 = 29.5 (p-value 0)

Table 2.4 Odds ratios representing the association of fHbp variants (1, 2 or 3) to carriage or disease.

2.4.3.2 Global distribution of fHbp variants within cc41/44

The expansion of clones originating from a common ancestor can be traced by observing the strain similarity of MLST allelic profiles and the positioning of clones in a minimum spanning tree diagram performed for example with the goeBURST algorithm [70]. The clonal complex ST-41/44 is known to be structured in two main sub-complexes (ST-41, ST-44), and with the help of the goeBURST cc41/44 diagram, isolates were assigned in three sub-groups (Figure 2.5): ST-41, ST-44 and Singlets. To perform this assignment, data of 2492 cc41/44 isolates retrieved from MLST database were analyzed. The disease outcome of the 2492 isolates reported from MLST database was peculiarly distributed (see Figure 2.6): 59% were from invasive cases, 31% from healthy carriers and 12% from other of uncharacterized cases. The percentages appeared different when we looked at sub-complexes. In fact, ST-41 sub-complex was mainly associated with invasive disease events (72% of ST-41 strains sub-complex are from disease), whereas ST-44 sub-complex included a relevant proportion of strains associated with carriage (45% of strains of this sub-complex are from carriers). In particular the 59% of carriage strains, belonged to the ST-44 sub complex.

To have a more complete picture of the fHbp variant distribution within cc41/44, and to assess any correlation with pathogenic or carriage strains, we expanded the results obtained in the Czech strain panel, with the results obtained from other epidemiological studies where fHbp was sequenced. In total, 255 cc41/44 isolates were analysed. We compared the clonal complex distribution and disease outcome of the 255 isolates panel against the data of 2492 cc41/44 isolates collected from the entire MLST database.

In the 255 isolates panel, the ST-41 sub-complex (192 isolates) included a predominance (97%) of strains associated with disease. Conversely, the ST-44 sub complex (51 isolates) included a relevant fraction (53%) of carriage associated strains. And the 79% of all carriage isolates, within the 255, belong to the ST-44 sub complex. In particular looking at the 32 Czech carriage clones, 31 are distributed into the sub-complex expanded from clone ST-44. The only strain causing disease belongs to the sub-complex ST-41.

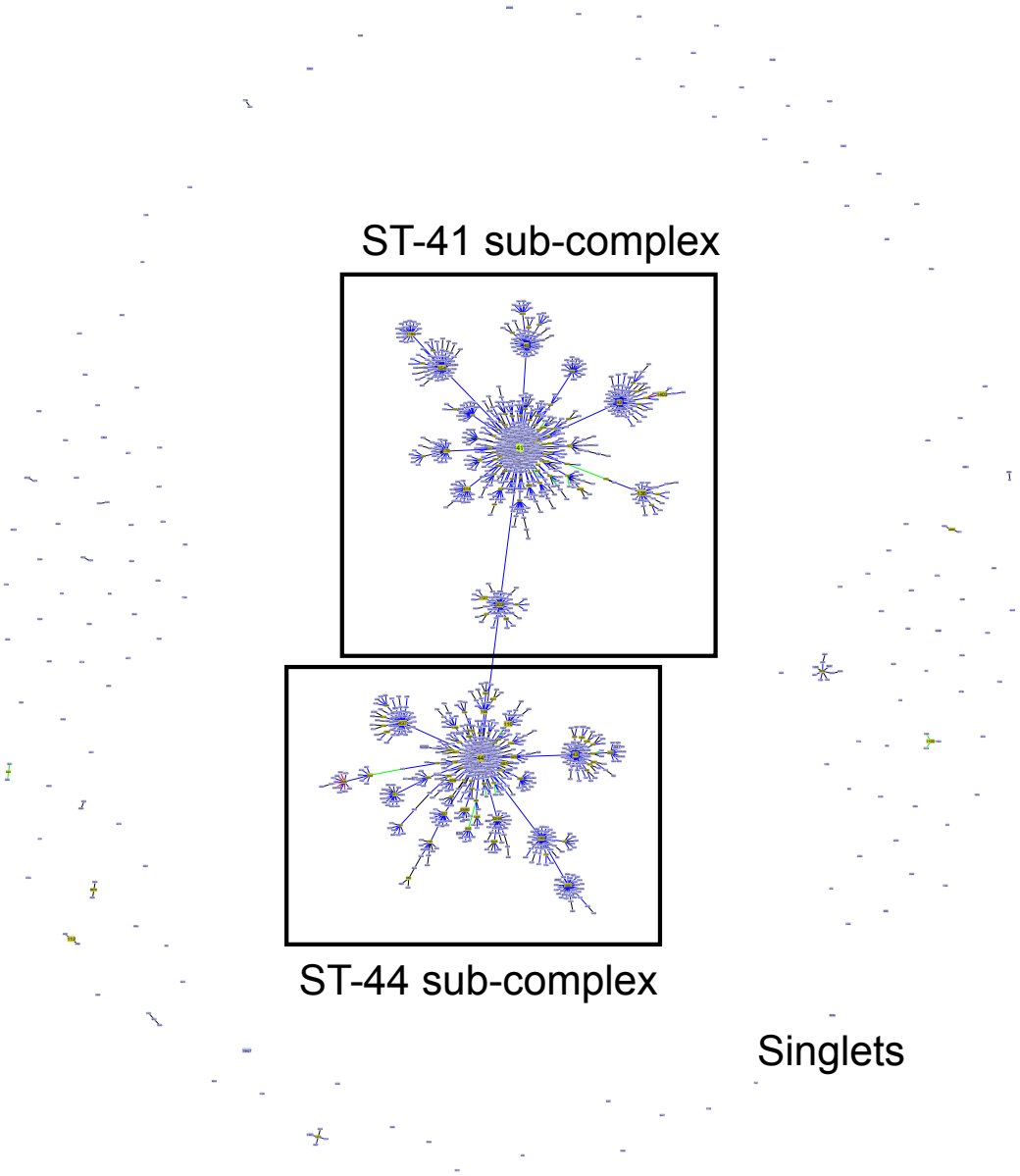


Figure 2.5 Clonal complex ST-41/44 minimum spanning tree reconstruction done with Phyloviz.

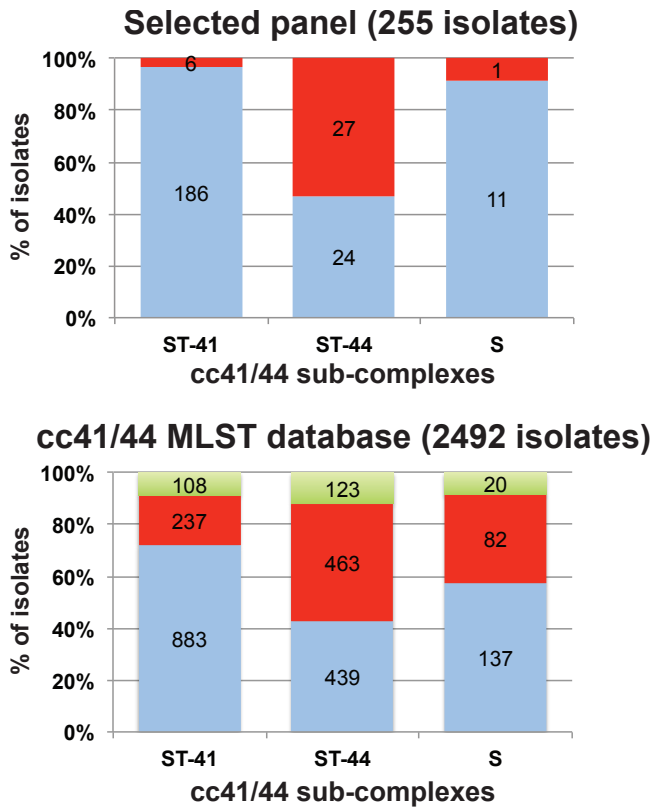


Figure 2.6 cc41/44 isolates and carriage/disease outcome distribution within the 255 isolates panel (a) and the 2492 isolates from MLST database (b). Blu bars represent isolates collected from disease cases, red represent isolates collected from healthy carriers and green undefined isolates.

Within the 255 cc41/44 isolates, only 27 protein subvariants were observed. The overall sequence variability of fHbp is shown in Figure 2.7 by a ML phylogenetic tree reconstruction. fHbp main variants 1, 2 and 3 relative distributions were compared with the origin of the sample (disease/carriage) and with the belonging to the sub-complexes ST-41, or ST-44. FHbp variant 2 strains mostly correspond to isolates from carriage, whereas fHbp variant 1 always correspond to isolates from disease. Strains with fHbp variant 3 are not significantly skewed into the two subgroups. In parallel, fHbp-2 strains mostly belong to the group expanded from ST-44 clone, while strains carrying fHbp-1 to the group expanded from the ST-41 clone. Also, strains with fHbp-3 do not significantly cluster into either subgroup. The V association statistics were V=0.67 and 0.53 when variants are compared with carriage/disease or cc41/44 sub-groups respectively and odds ratios are shown in the tables.

The distribution of fHbp in Czech strains belonging to other clonal complexes is more homogeneous. For example, all cc11 (50/52 isolates are ST-11) isolates harboured fHbp-2.17. These isolates were responsible of an outbreak in Czech republic, the proportion

carriage/disease were 32/20. Other carriage associated clonal complexes were cc92 (21 isolates, 8 different ST). Predominant cc92 ST was the one central to the clonal complex, namely ST-92 (13 isolates). Cc92 strains had predominantly fHbp-2.101 and fHbp-2.21. Else, cc106 again was more homogenous, 19/19 cc106 isolates (16/19 ST-106) harboured fHbp-2.286. 12/13 cc116 isolates (12/13 ST-116) harboured fHbp-2.17. The disease associated clonal complex ST-18 was more variable in terms of sequence types (9 STs out of 13 isolates). Conversely, 11/13 isolates harboured fHbp-1.37, indicating a quite stable fHbp genotype in this clonal complex.

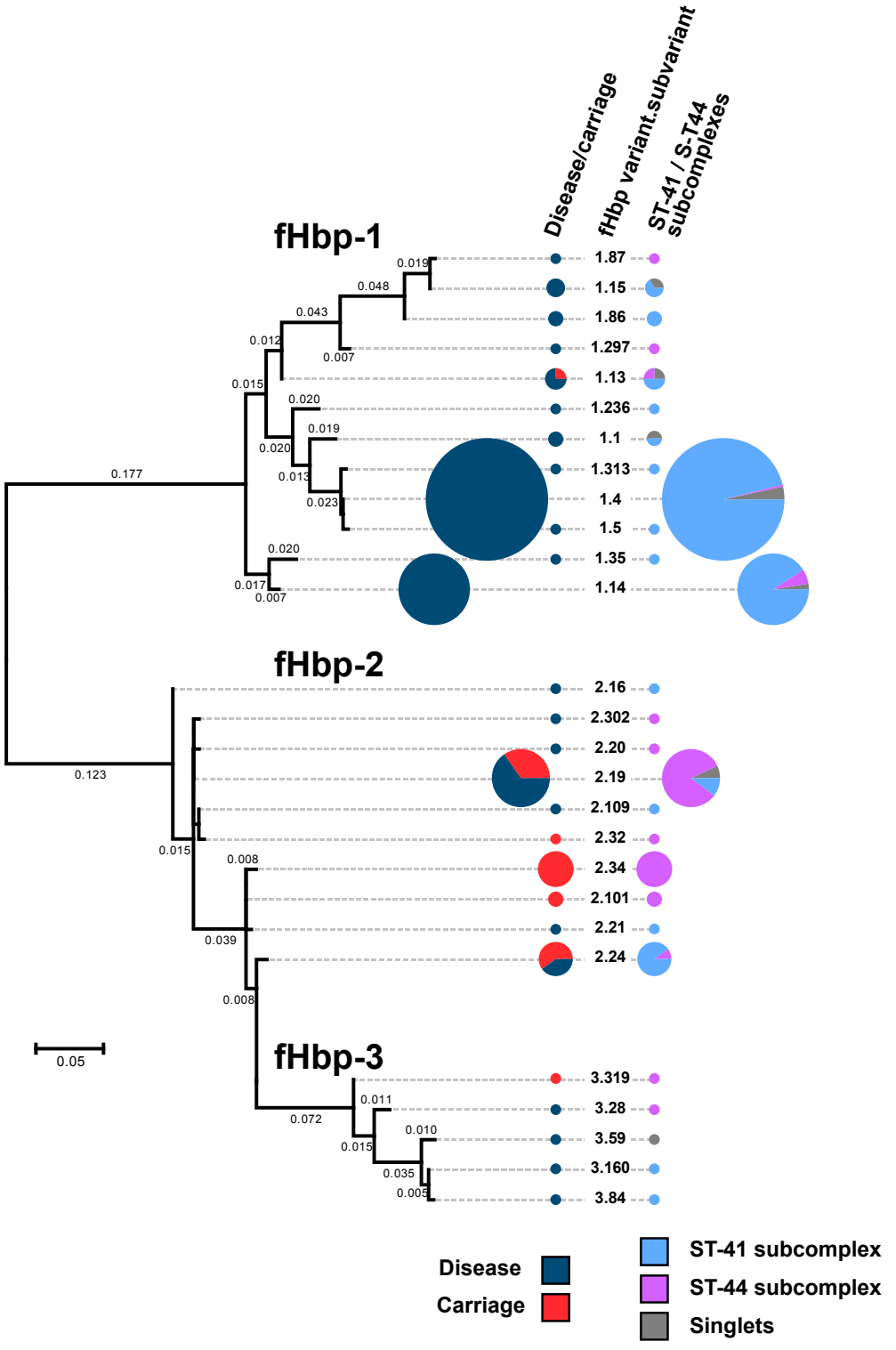
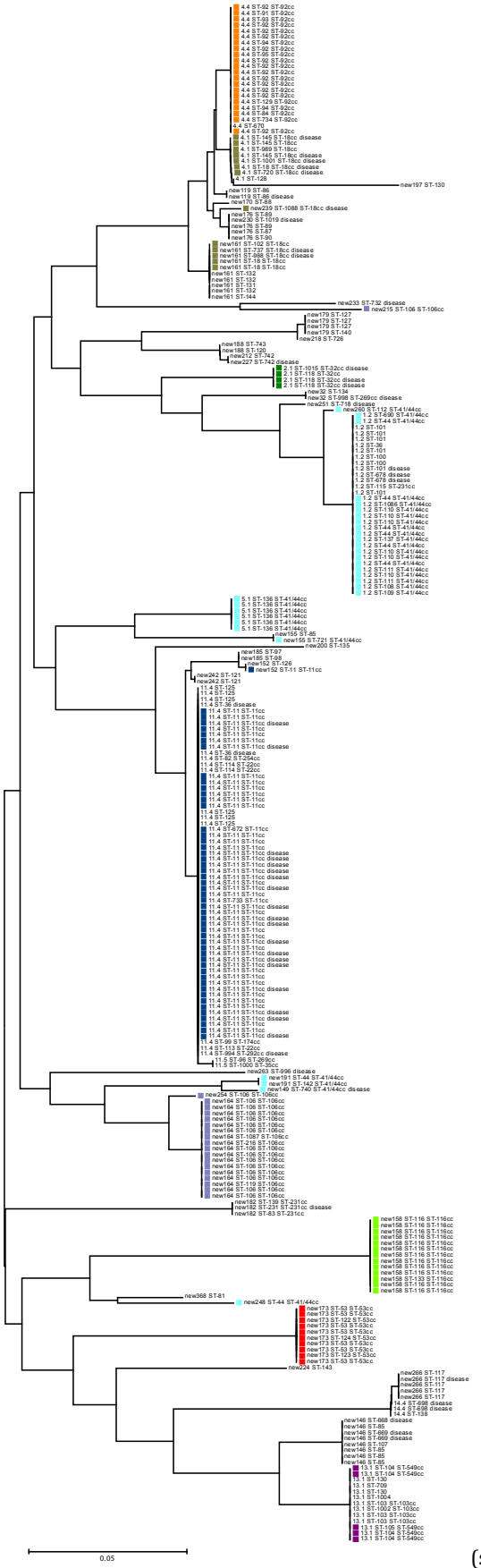


Figure 2.7 fHbp Molecular variability represented by a ML phylogenetic tree of the 27 protein forms observed in the 255 cc41/44 strain panel. Each tree leaf is linked, by a dotted line, to two distinct pie charts representing the disease/carriage and the cc41/44 sub complexes distributions. Chart areas are proportional to the number of isolates.

2.4.4 NHBA

2.4.4.1 Distributions of variants and significant associations with clonal complexes and carriage

The neisserial heparin binding antigen was present in all strains tested (261/266) showing 44 different peptides. 8 NHBA peptides were present in a minimum of 10 strains, NHBA-20, 2, 9, 55, 24, 53, 54 and 58). Altogether, they represent the 69% of the isolates (179/261). The number of protein forms was higher than fHbp. Conversely, variability within clonal complexes is lower, and the clonal distribution is comparable (index of association V_{NHBA} = 0.732). In fact the phylogenetic reconstruction shown in Figure 2.8 and the bar plot distribution shown in B clearly identify clades and specific protein forms mostly dedicated to a single clonal complex and related to carriage isolates only. Some examples of association were: NHBA-2 and cc41/44; NHBA-9 and cc92, NHBA-55 and cc106 both with a significant proportion of carriage isolates; NHBA-53 and cc116, NHBA-58 and cc53. NHBA-20 was predominantly present in cc11, though it was also shared with other complexes and it shows a significant proportion of disease isolates. NHBA-54 was typical of cc18 strains, NHBA-24 was equally distributed in cc103 and cc549 strains, all strains isolated from carriage. The summary of the most frequent NHBA peptides and how it is distributed with respect to healthy carriers or disease cases is shown in Table 2.5.



(see next page)

Figure 2.8 Molecular phylogeny of NHBA. Molecular evolutionary relationships were inferred using the Maximum Likelihood method. The evolutionary distances were computed using the JTT matrix-based method and the units are number of amino acid substitutions per site. The 9 most represented clonal complexes were coloured (cc11, cc41/44, cc92, cc106, cc116, cc18, cc53, cc549, cc32). All protein subvariants-peptides are strongly associated with clonal complexes

NHBA peptide sequence	carriage	disease	Total	Odds ratio	(95% CI)	FET* p-value
20	41	22	63	0.33	0.17 - 0.66	0.001
2	28	3	31	2.46	0.72 - 13.17	0.16
9	22	0	22	Inf	1.43 - Inf	0.01
55	17	0	17	Inf	1.05 - Inf	0.03
24	13	0	13	Inf	0.76 - Inf	0.08
53	13	0	13	Inf	0.76 - Inf	0.08
54	8	2	10	0.98	0.19 - 9.71	1
58	10	0	10	Inf	0.55 - Inf	0.22

Table 2.5 Odds ratios representing the association of most frequent NHBA peptide sequences with carriage or disease.

2.4.5 NadA

2.4.5.1 Presence / absence and variants

The *nadA* gene was present in 79 out of 266 Czech isolates (30%). With respect to data reported on our previous studies, in addition to serogroup B and C, also serogroup X isolates and some newly assigned STs were found to harbour *nadA* gene. Actually, the gene was present in 23 out of 79 positive isolates (29%) as an integer gene. *nadA* positive strains belonged to cc32 (4 isolates), cc18 (2 isolates) and to 12 different STs did not belonging to already assigned clonal complexes (17 isolates) (C). Sequence analyses revealed that in 56 (71%) isolates the gene was interrupted by the insertion sequence *IS*1301. 52 out 56 isolates showing an interrupted *nadA* gene belonged to the cc11, mostly to the serogroup C outbreak The association of *nadA* gene interrupted by *IS*1301 and cc11 strains, in particular belonging to the ET-15 clone, was described by Elias *et al.* [73]. The clonal complex assignment is a good descriptor of the NadA presence due to a relevant association (V_{nadA} = 0.71). Five NadA variants were found, both in invasive than in carrier isolates (C): NadA-1 and NadA-3, which we described as mainly associated with invasive isolates [21], NadA-4, mainly associated with carriage [22], NadA-5, which was described

to be associated with cc213 and a NadA-4/NadA-5 related variant first described by Lucidarme and colleagues [13].

2.4.5.2 *Differences between carriage and disease*

In the Czech strain panel, the presence of the *nadA* gene was differently distributed when we consider strains isolated from healthy carriers and disease isolates. *nadA* was rarely present in carrier strains (37%), as compared with clinical isolates (63%) (Table 2.6). In particular, *nadA* negative strains had a relevant proportion of strains from carriage (OR 3.45, p-value 0.0001). Instead, *nadA* positive isolates were associated with disease (OR = 0.32, p-value 0.05). These observations are in agreement with the role of NadA in the adhesion and pathogenicity of *Nm*. NadA-1 and NadA-3 variants were more frequent in disease isolates; on the contrary, NadA-4, NadA-5 and NadA4\NadA-5 variants were more frequent in carrier isolates and showed a higher degree of sequence variability.

NadA presence	carriage	disease	Total	Odds Ratio	(95% CI)	FET* p-value
-	164	25	189	3.45	1.77 - 6.76	0.0001
+	10	7	17	0.32	0.10 - 1.04	0.05
IS	39	21	60	0.34	0.17 - 0.68	0.001
IS and -	203	46	249	2.20	0.71 - 6.23	0.15
Total	217	53	266	Mantel-Haenszel OR =1	0.76 - 1.54	Test for heterogeneity: X^2 = 33.7 (p-value 0)

Table 2.6 Odds ratios representing the association of *nada* gene presence (+), absence (-) or interrupted locus (IS) to carriage or disease.

2.4.6 Variability of PorA and FetA variable regions

Interestingly, the distribution of the variable regions of the surface proteins PorA and FetA, was similar to the previously described surface proteins. In fact, despite the fact that variable epitopes of the proteins are significantly more spread, showing a higher number of variants (20 for PorA_VR1, 42 for PorA_VR2 and 38 for FetA_VR), they resulted comparably associated to the clonal complexes (index of association are V_{PorA_VR1} = 0.503, V_{PorA_VR2} = 0.675 and V_{fetA_VR} = 0.711). Due to the high number of peptides observed and consequent small respective frequencies, only few variants have a significant OR statistics with respect to carriage or disease. The most relevant PorA subtype associated to carriage is PorA P1.5-1 (typical of cc106 and cc92, and of other complexes, as well). Subtype P1.5,2 with FetA3-6, which were typical of cc11 isolates, and P1.7,16 with FetA3-3, which were

all typical of cc32 isolates, showed relevant odds ratio statistics of association with disease cases.

2.5 Discussion and conclusions

The genetic variability of a panel of meningococcal isolates collected in Czech Republic during the year 1993 was studied with a particular focus on the differences between isolates collected from healthy carriers and isolates collected from people affected by disease. The distribution and prevalence of specific surface proteins, fHbp, NadA and NHBA, main components of the 4CMenB (Bexsero®) vaccine against *Nm*, were studied in function of different genetic characters of the meningococcal isolates. Clonal complex grouping was the major descriptor of the antigens variability because of the already described association between molecular variants and clonal complexes. For all the surface proteins evaluated, the distribution of variants was strongly related to the genetic characteristics of the strains. Also, odds ratios deviations demonstrated relevant relationships of protein diversity with carriage, or with disease. fHbp-2 was strongly associated with strains derived from healthy carriers, with fHbp-2.17 and fHbp-2.24 predominating, whereas fHbp-1 and fHbp-3 were more strongly associated with strains derived from disease cases. The higher frequency of fHbp-1 in invasive isolates was in agreement also with data obtained in recent studies in Europe [5] and United States [74]. The *nadA* gene was only present in a subset of isolates. It was prevalent and more conserved in disease isolates, whereas it was less frequent and more variable in carrier isolates. Finally, NHBA protein was always present in the tested isolates, it was strongly associated to the clonal complexes and some specific protein forms were related to carriage only. NHBA-20 was the most frequent both in carriage than in disease isolates in the present strain panel. On the contrary, in more recent (2007-2008) invasive strain collections in Europe, NHBA-2 was predominant [5].

Molecular phylogeny was reconstructed for fHbp and NHBA proteins, showing that clonal complex grouping strongly correlates to specific tree clades. Rarely the fHbp subvariants and the NHBA peptides, which are typical of a clonal complex, are found in other complexes. This would indicate a limited rate of recombination between complexes after their diversification. Interesting differences in the antigenic repertoires of carriage and pathogenic strains belonging to the same clonal complex were found. NHBA protein showed a high level of clonality, in particular a high association index with clonal complex designation. The analysis of the fHbp structuring within cc41/44 was investigated in a

collection of 255 carriage and disease isolates, by comparison of protein sequences. The analysis demonstrated that fHbp variants and subvariants are non-randomly distributed in ST-41/44 sub-complexes. fHbp-1 subvariants are mostly associated with cc41/44 isolates collected from disease cases, which are preferentially included into the ST-41 sub-complex. Instead, fHbp-2 subvariants are mostly associated with carriage strains. The majority of strains from healthy carriers belong to ST-44 sub-complex.

This study was conducted using a well-recognized strain panel for studies on meningococcal carriage. It was optimal for a comparison between diversity on carriage vs diversity in disease strains, because pathogenic strains occurred in the same area, and during the same epidemiological year of isolation, were included in the same sample. These findings are important for vaccine development, in particular once evaluating the vaccine effectiveness and protection against both disease-causing and carriage strains. In fact, the distribution of the 4CMenB vaccine antigens in the meningococcal isolates from healthy individuals, suggests that vaccination could have an impact on carriage and suggests that the determination of the vaccine coverage on carriage would be of complex estimation as in disease-causing isolates of serogroup B [5]. However, as recently shown for the 4CMenB vaccine [75], the ability of vaccine-induced antibodies to mediate killing depends on their ability to react with the antigen variant/peptide present in the vaccine and on the amount of antigen expressed by a given strain that can vary in response of yet uncharacterized environmental factors, as recently demonstrated for NadA [76]. Therefore, only typing methods taking into account this aspects would strain coverage and further studies will be needed to evaluate whether this vaccine will have any impact on meningococcal carriage.

Chapter 3. An analysis of the sequence variability of meningococcal fHbp, NadA and NHBA over a 50-year period in the Netherlands.

3.1 Abstract

Studies of meningococcal evolution and genetic population structure, including the long-term stability of non-random associations between variants of surface proteins, are essential to evaluate the effect of a vaccine. In this work we analysed the sequence variability of factor H-binding protein (fHbp), Neisserial Heparin-Binding Antigen (NHBA) and Neisseria adhesin A (NadA), in a panel of invasive isolates collected in the Netherlands over a period of 50 years. To our knowledge, this strain collection covers the longest time period of any collection available worldwide. Long-term persistence of several antigen sub/variants and of non-overlapping antigen sub/variant combinations was observed. Our data suggest that certain antigen sub/variants including those used in 4CMenB are conserved over time and promoted by selection.

3.2 Introduction

N. meningitidis bacterial population is structured in highly diverse lineages that show patterns of an extensive genetic exchange [77]. Studies of genetic variation, by MultiLocus Enzyme Electrophoresis (MLEE) [78] and subsequently by MultiLocus Sequence Typing (MLST) [58], showed that meningococci are structured in lineages defined as 'clonal complexes' [79, 80], some of which are clearly associated to the capability to cause invasive disease and for this reason were defined as 'hypervirulent' clonal complexes [81]. Strains belonging to the hypervirulent clonal complexes are over-represented in collections of pathogenic isolates. Clonal complexes are relatively genetically stable over time, despite high rates of recombination [58].

The persistence of clonal complex structure despite high levels of recombination can be explained by evolutionary models that invoke positive selection [65, 66], which has important implications for the design of protein-based vaccines against meningococci of

serogroup B (MenB). In fact, outer membrane vesicle (OMV) vaccines that rely on the immunogenic properties of PorA [82-85], have been used: though, those vaccines only provide protection for homologous strains carrying the same PorA [86]. To overcome antigenic variability, vaccines based on multiple outer membrane proteins have been proposed to provide protection against a broad range of meningococcal isolates. Novel antigens were identified by Reverse Vaccinology [16, 86, 87] and combined into a multicomponent vaccine against MenB, 4CMenB (Bexsero®). 4CMenB includes OMVs from the New Zealand MeNZB® vaccine and three major protein antigens: factor H-binding protein (fHbp), Neisserial Heparin-Binding Antigen (NHBA) and *Neisseria* adhesin A (NadA).

Statistical association studies indicate that the repertoire of fHbp, NHBA and NadA is structured among hyperinvasive lineages. Isolates from the same clonal complexes have similar profiles for each antigen, even when derived from disparate geographical locations and time periods [39].

We investigated the prevalence and sequence variation of fHbp, NHBA and NadA in a panel of 165 pathogenic meningococcal isolates randomly selected from those collected in the Netherlands over a period of 50 years.

3.3 Methods

3.3.1 Bacterial isolates

One hundred sixty-five meningococcal isolates collected from clinical cases (from blood or cerebrospinal fluid) in the Netherlands were randomly selected at the Netherlands Reference Laboratory for Bacterial Meningitis (NRLBM). Approximately 30 isolates were included from each decade as follows: every 2nd, 5th, 10th, 20th, 20th and 10th isolate were chosen of the years 1960, 1970, 1980, 1990, 2000 and 2008-2009, respectively.

Upon receipt of bacterial isolates in the NRLBM, a monoculture of the causative isolate was frozen and stored at –80°C. All isolates were passaged fewer than 5 times. All isolates were characterized by serogroup, and sequencing of PorA, FetA, fHbp, NHBA, NadA and MLST [67, 80, 88].

Discrimination between ST44 and ST41 subcomplexes of clonal complex41/44 was assessed by PCRs targeting MoxR like AAA ATPase and *nmeSI*, respectively [89, 90].

3.3.2 Sequence analysis and measure of the associations between different loci

DNA sequences were assembled and analysed using Sequencher version 4.10.1 sequence analysis software (Gene Codes Corporation, Ann Arbor, MI USA), BioEdit (developed by Tom Hall, *Ibis Biosciences*), Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0 [91] and Jalview [92].

Molecular typing of fHbp, NHBA and NadA sub/variants was based on amino acid sequence determination. fHbp sub/variants were designated as variants.oxford.peptide (http://pubmlst.org/neisseria/fHbp/). NHBA sub/variants were designated as peptides. NadA sub/variants were designated as variant.peptide.

The structure of MLST clonal complexes was analysed with PHYLOViZ 1.0 based on the goeBURST algorithm [70].

The Cramer's V coefficient [93] was used to measure association statistics with clonal complexes. The Standardized Index of Association I_A^S [94] was used to test the stability of associations between different loci. The non-overlapping structure of antigen variant combinations was measured by f* metrics [66]. The three statistical parameters V, I_A^S and f^* are based on the frequency of the alleles and vary between 0 (random distribution) and 1 (perfect association or non overlapping distribution in the case of f^*).

3.4 Results

3.4.1 Prevalence and Distribution of Sequence Types (STs) and MLST Clonal Complexes Over Time

Ninety-five different STs and 22 clonal complexes were found. In detail, ST-8 (9%), ST-11 (8%), ST-41 (8%), ST-1 (5%) were the most commonly represented STs. ST-1, ST-41 and ST-11 were the only STs detected in more than two consecutive decades based on the sample studied. In contrast, 82 (86%) STs were observed only once, 7 (7%) twice and 3 (3%) in two or three not consecutive decades. Twenty-seven percent of the isolates belonged to cc41/44, 13% to cc8, 9% to cc11, 8% to cc32 and 7% to cc269. The remaining 36% of isolates included 17 different clonal complexes. Ten clonal complexes were observed only once, 3 were present twice, whereas 9 persisted for more than two decades (Figure 3.1).

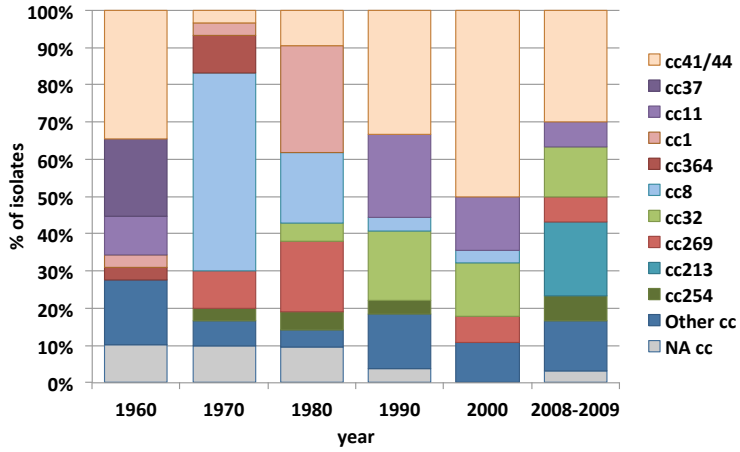


Figure 3.1 Distribution of the MLST clonal complexes from 1960 to 2008-2009 in the Netherlands. 'Other cc' groups together 12 less frequent clonal complexes (cc5, cc22, cc35, cc103, cc162, cc1157, cc18, cc60, cc167, cc231, cc461 and cc334), having a global frequency inferior to 2% of the entire dataset. 'NA cc' represents STs without a clonal complex assignment.

3.4.2 Diversity and Distribution of fHbp

The *fHbp* gene was present in all isolates. Forty-nine different nucleotide sequences were identified. Forty-five out of the 47 different amino acid sequences were in frame. 98% of amino acid sequences belonged to the three variants: fHbp-1 (52%), fHbp-2 (39%) and fHbp-3 (7%). Concerning the remaining 2%, in one case the fHbp variant was a natural chimera between fHbp-1 and fHbp-3, which was recently described [12]; one sequence presented a frame-shift, causing a premature end of the encoded peptide; moreover, in one case the gene sequence was not obtained.

fHbp-2 predominated in 1960-1980, whereas fHbp-1 became more prevalent thereafter (Figure 3.2). fHbp-1.14 (16%), 2.16 (13%), 2.24 (8%), 1.4 (8%), 1.1 (8%; included in 4CMenB), 1.13 (5%), 2.22 (4%), 2.19 (4%) and 3.45 (2%) were the most frequent sub-variants (Figure 3.3A).

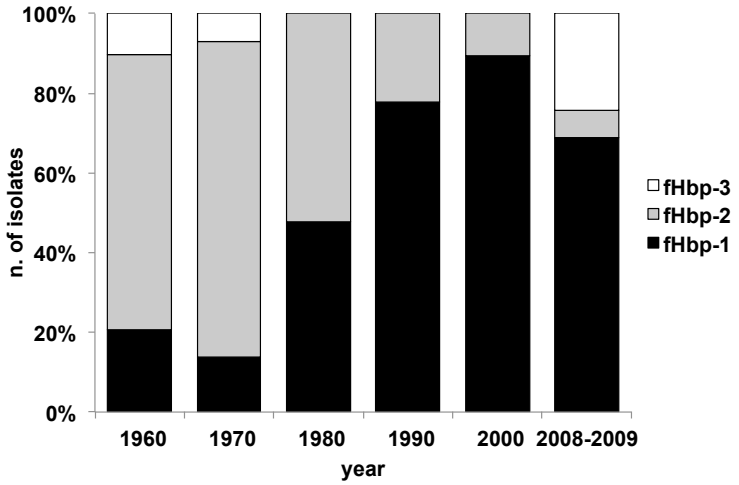


Figure 3.2 Distribution of the three fHbp variants Over Time.

3.4.3 Diversity and Distribution of NHBA

The *nhba* gene was found in all isolates considered. In one case the gene presented a frame-shift mutation, probably causing a lack of protein expression. Forty-eight different nucleotide sequences, corresponding to 43 different amino acid sequences, were identified. NHBA-2 (24%), 20 (22%), 3 (6%), 29 (6%), 6 (5%), 12 (4%), 21 (4%), 18 (3%) and 24 (3%) were the most frequently occurring (Figure 3.3B). NHBA-2 (included in 4CMenB) was predominant in 1960 (34%), 1990 (33%), 2000 (43%) and 2008-2009 (20%), whereas NHBA-20 was predominant in 1970 (53%) and NHBA-29 in 1980 (28%).

3.4.4 Diversity and Distribution of NadA

Of 165 isolates, 60 (36%) had the *nadA* gene. Eight isolates had the *nadA* gene interrupted by the insertion sequence IS*1301.* Therefore, they most likely did not express the protein [21]. The gene of two isolates presented a frame-shift mutation. Therefore, 50/165 (30%) isolates were considered positive for NadA. The gene was absent in cc41/44 and cc269 with one exception in both clonal complexes, strain 600091 and 701844, respectively. Variants NadA-1, NadA-2, NadA-3 and NadA-5, which are generally present in pathogenic strains, accounted for 28%, 16%, 42% and 8% of the positive isolates, respectively. NadA-4, the variant described in carrier strains, was found in 6% of the positive isolates. NadA-3 and NadA-4 were most frequent in 1960-1970, NadA-1 and NadA-2 were prevalent in 1990-2000, NadA-1 and NadA-5 in 2008-2009. Eleven sub/variants were identified, the most frequent being NadA-3.8, the NadA sub-variant in 4CMenB, present in 20 out of the 50 isolates with the *nadA* gene (40%), which was predominant in 1960-'70-'80. NadA-1.1 (28%) has been prevalent after 1990 (Figure 3.3C).

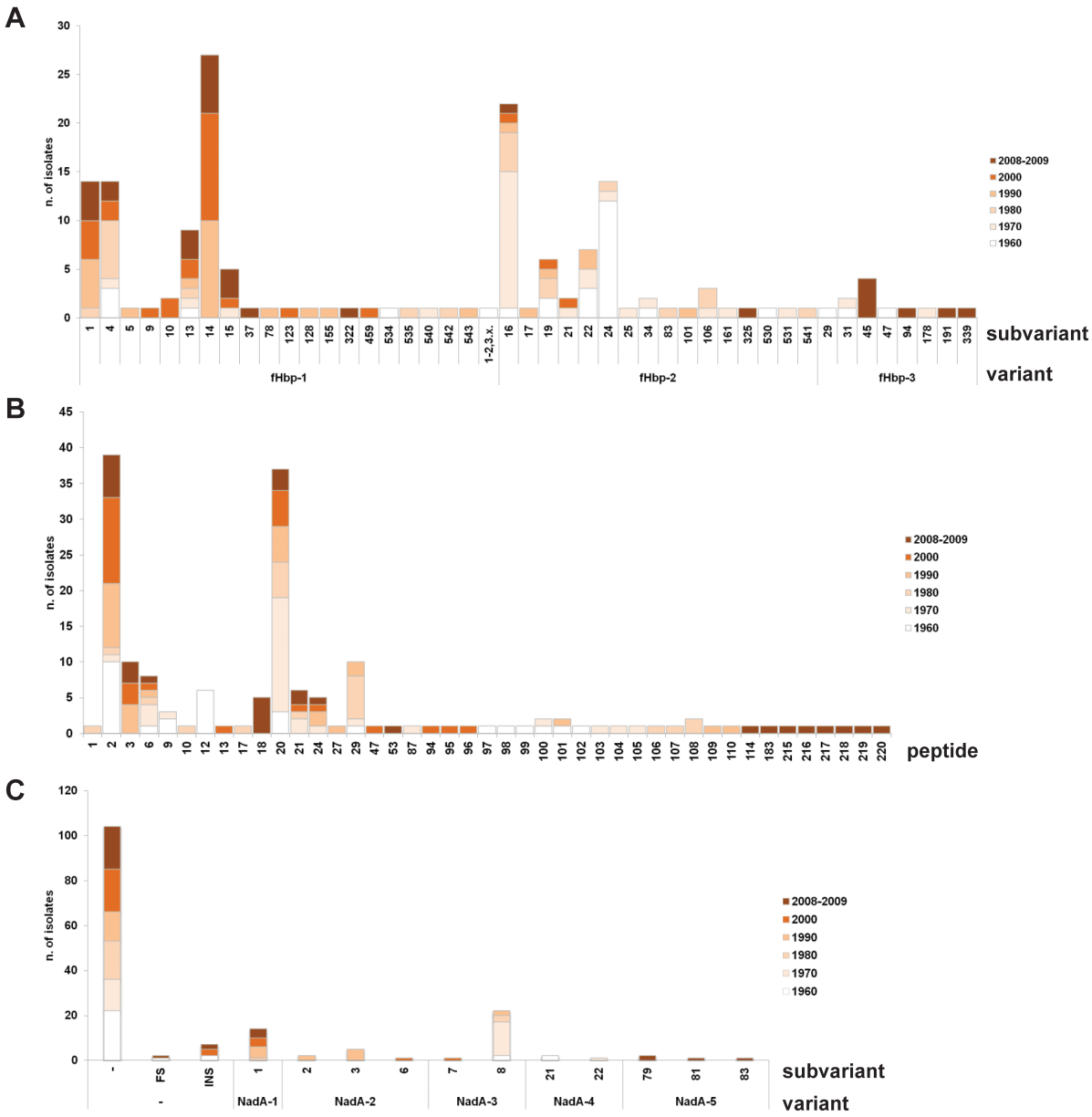


Figure 3.3 Distribution of protein variants of fHbp (A), NHBA (B) and NadA (C) from 1960 to 2008-2009.

3.4.5 Association of fHbp, NadA, and NHBA diversity with MLST Clonal Complexes

We observed clear associations between fHbp, NadA, and NHBA unique sequences with clonal complexes. These results, also obtained in other strain panels [39], were based on a standardized measure of association, the Cramer's V coefficient (Table 3.1). The value of this coefficient of correlation is high, in particular for fHbp and NHBA (V_{NHBA} = 0.769, V_{fHbp} = 0.704 and V_{NadA} = 0.582).

	Measure of association with clonal complexes (Cramer's V§)	Measure of the association between two different loci (Standardized Index of Association§, I_A^S)			Measure of non-overlapping structure of combinations (f* metrics§)		
Gene	Housekeeping genes	*fHbp*	*nhba*	*nadA*	*fHbp*	*nhba*	*nadA*
fHbp	0.705		0.371	0.667		0.284	0.601
nhba	0.769	0.371		0.512	0.284		0.462
nadA	0.582	0.667	0.512		0.601	0.462	

§The three statistical parameters V, I_A^S and f^* are based on the frequency of the alleles and vary between 0 (random distribution) and 1 (perfect association or non overlapping distribution in the case of f^*).

Table 3.1 - Associations statistics between different loci.

Several sub/variants were associated with specific clonal complexes. For example, fHbp-1.1 was always present in cc32 isolates. Twenty-five out of 27 fHbp-1.14 were present in cc41/44 (92%). fHbp-2.16 was almost always represented in cc8 isolates (91%). fHbp-1.15 and fHbp-3.45 were found mostly in cc269 (80%) and cc213 (75%), respectively. However, eleven different fHbp sub/variants were found in cc41/44, the most frequent being fHbp-1.14 (55%). fHbp-2.16 was almost always represented in cc8 isolates. Seven and 9 different sub/variants were represented within cc11 and cc269, respectively.

Similar association with clonal complexes was observed for NHBA peptides. NHBA-3 was always present (100%) in cc32, NHBA-12 in cc37, NHBA-18 in cc213 and NHBA-21 in cc269. NHBA-2 was predominant in cc41/44 (37 out of 39 isolates, 95%). NHBA-29 was found mostly in cc1 (80%). NHBA-20 was the most frequent peptide in cc8 (57%), even if it was also present in cc11 (32%).

Also NadA variants were associated with clonal complexes. NadA-1 was solely present in cc32 strains. NadA-2 (7 out of 8 isolates, 88%) was mostly present in cc11 strains. NadA-3 (81%) was mostly present in cc8 strains. NadA-5 was solely present in cc213 strains.

3.4.6 Evolution of cc41/44

The only clonal complex present at all observed time points in the sample was cc41/44. It showed a shift over time between its two central STs [80] (Figure 3.4). The ST-44 sub-complex was predominant in 1960-1970, and ST-41 was prevalent after 1980.

Still, due to association between (any) protein diversity and MLST, a trend in the ratio of fHbp variants was observed to follow changes in the central genotypes (Figure 3.4A). From 1960 to 1980, when the ST-44 sub-complex was predominant, fHbp-2 was most

represented, and sub/variants fHbp-2.24 and 2.19 were predominant. After 1980, as sub-complex ST-41 became predominant, fHbp-1 became the most frequent variant and sub/variant fHbp-1.14 predominated after 1990 (Figure 3.3A).

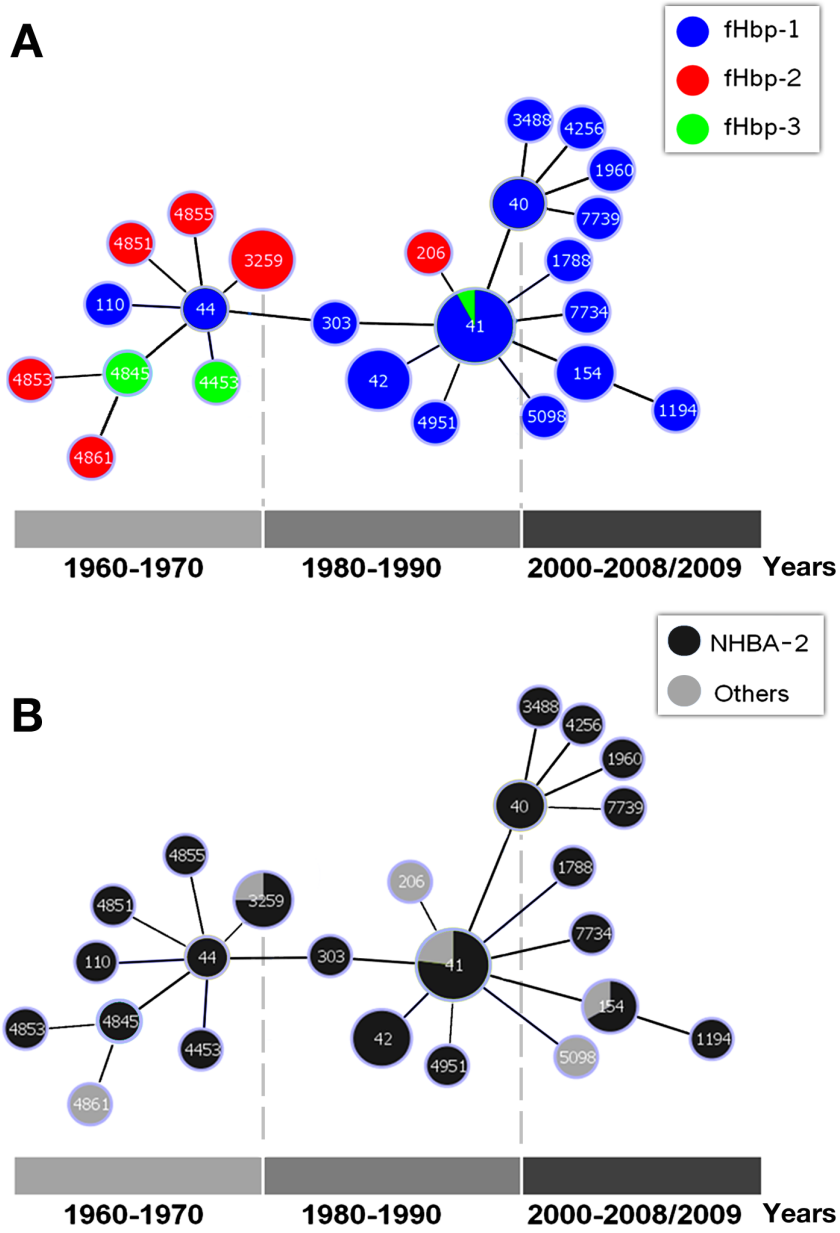


Figure 3.4 The evolution of cc41/44 over 50 years.

Similar associations with each of the two sub-complexes of cc41/44 were also observed for PorA and FetA variants. PorA P1.7-2, 4 and FetA F1-5 were associated with the ST-41 sub-complex isolates collected in 1990 and afterward, while this combination was not observed among the ST-44 sub-complex isolates. The latter was more heterogeneous with PorA P1.18, 25-7 prevalent.

In contrast, NHBA-2 was the most frequently occurring peptide in both cc41/44 sub-complexes across the 50-year period examined (Figure 3.4B). As for all clonal complexes, the association with ST-41 and ST-44 sub-complexes, within cc41/44, was evaluated using Cramer's V coefficient. NHBA-2 was predominant but evenly distributed between the two sub-complexes and consequently NHBA did not show a relevant enrichment of specific peptides in one of the sub-complexes (V_{NHBA} = 0.298). In contrast, the overall distribution of the fHbp sub/variants in cc41/44 was not uniform and showed a relevant association with sub-complexes (V_{fHbp} = 0.683).

3.4.7 Long-term Persistence of Antigen Sub/variants

A number of antigen sub/variants appeared in the dataset only once, indicating a low frequency in the population. This might suggest that they were less fit and excluded by selection over time, while other sub/variants persisted over decades. Of the 45 fHbp sub/variants identified in the current sample, 30 were observed once, and 10 were seen over a period of at least 20 years. Among the 43 NHBA peptides observed, 30 occurred once, and 7 have been encountered over at least 20 years. Eleven NadA sub/variants were identified, of which 5 were observed once and 2 were present over at least 20 years. The proteins included in 4CMenB (fHbp-1.1 and NadA-3.8) were observed over thirty years, while NHBA-2 was observed over fifty years (Figure 3.5).

In general, sub/variants that were found persisting over decades in this present study were found also in other strain panels from different time periods [18, 39, 57]. Moreover, several sub/variants have been associated over time also with different clonal complexes. For instance fHbp-1.13 was identified in cc254 strains collected during 1970, 1980, 1990 and 2008-2009, and in cc60 strains in 2000. fHbp-1.14 was identified in cc41/44 isolates in 1990, 2000 and 2008-2009, and in cc254 isolates collected in 2008-2009. fHbp-2.16 was found in cc8 isolates collected between 1970 and 2000, and in cc22 isolates from 2008-2009.NHBA-20 was identified in cc11 isolates in 1960, 1990, 2000 and 2008-2009, while in 1970 and 1980 it was prevalent in cc8.

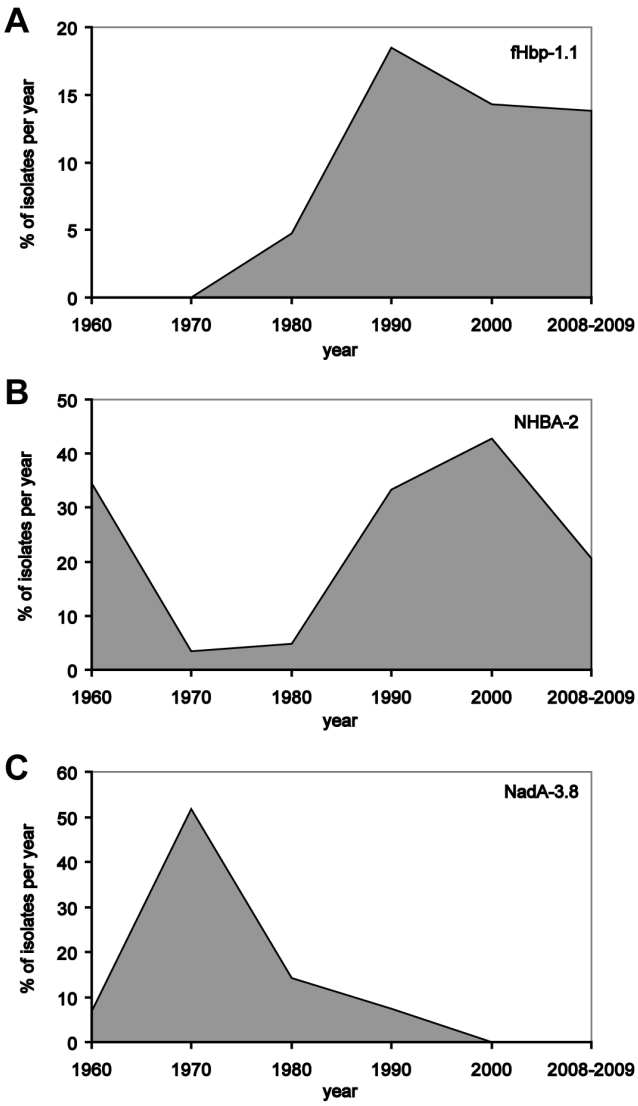


Figure 3.5 Persistence of the fHbp (a), NHBA (b) and NadA (c) sub/variants included in 4CMenB vaccine formulation.

3.4.8 Long-term Persistence of Antigen Sub/variant Combinations

Of 83 combinations of fHbp, NadA, and NHBA sub/variants**,** most were found only once, yet certain combinations were stable and persisted over time. Examples include: fHbp-1.1:NadA.1.1, which persisted for 30 years; and fHbp-2.16:NHBA-20, which persisted for 40 years (Figure 3.6). The relative presence of several combinations changed over time. For example, fHbp-2.16: NHBA-20 was the most frequent in 1970, fHbp-1.4:NHBA-29 was more commonly found in 1980, and fHbp-1.14:NHBA-2 was predominant after 1990.

In addition to association of antigens sub/variants with clonal complexes, several antigen sub/variant combinations were maintained in the same clonal complex. The fHbp-1.1:NadA-1.1 combination, first observed in 1980, persisted in all isolates of cc32 for thirty years. The fHbp-1.14:NHBA-2 combination was found in cc41\44 isolates in 1990 and

persisted for 20 years. The fHbp-1.4:NHBA-29 combination was found in cc1 isolates from 1960 to 1980. The fHbp-2.16:NHBA-20:NadA-3.8 combination was observed in cc8 isolates in 1970 and in 1980.

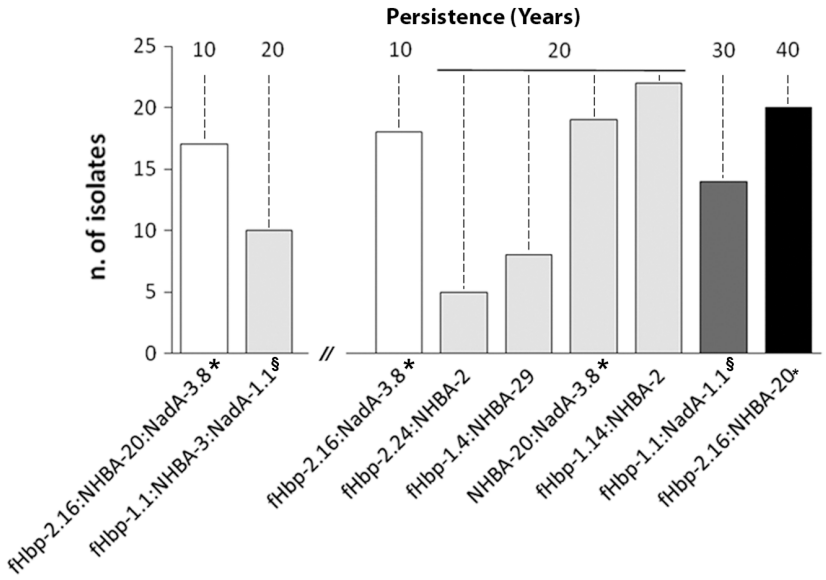


Figure 3.6 Longevity of the most frequent combinations of two and three antigen sub/variants.

The Standardized Index of Association I_A^S [94] was used to test the stability of associations between pairs of loci, i.e. the presence of linkage disequilibrium (Table 3.1). The I_A^S of the *fHbp* with respect to *nhba* and *nadA* was 0.371 and 0.667 respectively, the index between *nhba* and *nadA* was 0.512. These three values showed a particular faculty of *nadA* to be associated with *fHbp* and *nhba*. For example, these rates were comparable with the same index calculated for the association between *porA* VR1 and VR2 (I_A^S = 0.566), two very close loci belonging to the same gene. The linkage disequilibrium between the *fHbp*, *nhba* and *nadA* loci showed also a non-overlapping structure (measured by f*metrics, Table 3.1). The non-overlapping structure, in particular between *fHbp* and *nadA* (f*metrics = 0.601) can be interpreted as the result of the pressure of selection exerted by the immune system in maintaining antigenic combinations in the neisserial population [66].

3.5 Discussion and Conclusions

The sequence variability of fHbp, NHBA and NadA have been examined in several strain panels worldwide and investigations are still ongoing [39][18, 57]. We evaluated prevalence and sequence variations of fHbp, NHBA and NadA in a panel of invasive

meningococcal isolates collected in the Netherlands from 1960 to 2009, which, is to our knowledge, the widest time frame of any collection worldwide.

Among the meningococci isolated in a 50 year period many sub/variants of each antigen fHbp, NHBA and NadA were identified. Although most sub/variants were observed only once, or persistent over a period of a few years, a significant number, such as those used in 4CMenB, were observed to persist over time. Of note, the antigen sub/variants we found to be stable and conserved over at least twenty years have been identified as the most frequent in other panels of invasive isolates [1, 18, 39, 56, 95-97]. Short-lived sub/variants were likely less fit or occurred too infrequently to be observed, which may suggest that they were excluded by immune selection.

Associations between antigen sub/variants and clonal complexes measured by the Cramer's V coefficient, a standardized measure of association, were also identified and were consistent with previous studies in other strain panels [39, 57].

The three protein sub/variants chosen for inclusion in 4CMenB (fHbp-1.1, NHBA-2 and NadA-3.8) were the most common worldwide and were expected to provide cross-reactivity [17, 56, 98, 99]. An interesting and unexpected finding of this study is their long-term persistence in the current strain panel, which was thirty years for fHbp-1.1 and NadA-3.8, and fifty years for NHBA-2.

Long term stability and persistence was observed also for several combinations of protein sub/variants. Antigen combinations showed a non-overlapping structure. As reported in literature, if immune selection was absent or neutral, then all antigen variants and sub/variants would occur for similar time spans. On the contrary, it has been suggested that selection may cause strong linkage disequilibrium (or non-random association of alleles at the different loci) between some locus pairs [48, 65, 96]. As observed in this study, the stability and persistence over decades of discrete fHbp, NHBA and NadA sub/variants and non-overlapping antigen sub/variant combinations suggest they were maintained by natural selection. In the case of antigen sub/variants, the persistence was noticed also in different genetic environments such as different clonal complexes. Persistent antigen combinations in association with certain clonal complexes may indicate that acquisition of new alleles encoding antigen variants may impair fitness of that clonal complex [90].

Our observation of the persistence of the antigen sub/variants in 4CMenB may also indicate that the vaccine will be able to provide protection against populations of meningococci over time, as the antigens they target have tended to persist.

We used the Standardized Index of Association I_A^S to quantify the extent of allele association and long-term stability and f* metrics to evidence a non-overlapping structure of antigen combinations. Buckee and colleagues first used this parameter to evaluate the long-term stability of FetA and PorA combinations on carried meningococci [66]. Given the novelty of the Buckee study, our results contribute not only to an understanding of invasive meningococcal strains but also to the development of approaches for evaluating long-term stability in bacterial populations over time.

In the cc41/44 complex, the predominance of antigen sub/variants and STs shifted over time. This clonal complex is of additional public health interest because cc41/44 strains are almost equally often isolated from healthy carriers and from cases of invasive disease [53, 89, 97]. Recently, a cc41/44 outbreak in the city of Aachen, Germany and 3 neighboring counties (Greater Aachen) has been described [89]. Strains of cc41/44 are also among the most important causes of serogroup B disease in the USA. The central STs of cc41/44, that is to say ST-41 and ST-44, are putative 'ancestral genotypes' [80]. In the Netherlands strain collection, the ST-44 sub-complex was predominant during the 1960s, whereas the ST-41 sub-complex became predominant from 1980 onwards, a circumstance also observed in Belgium [97] and New Zealand [53] from 1990 onwards. As previously published, cc41/44 isolates collected during 1980 and later harbored the restriction modification system *nmeSI* [90]. In contrast, 80% of the isolates of the ST44 sub-complex appeared to have two different genes at the genomic position of the *nmeSI* system encoding a MoxR like AAA ATPase and a protein with a Von Willenbrand domain, respectively (data not shown). In eBURST analyses, fHbp variant distribution changed with the relative predominance of the ST-44 and ST-41 sub-complexes. In 1960-1970, when the ST-44 sub-complex was predominant, fHbp-2 and fHbp-3 were more common in the strain collection. After 1980 ST-41 and ST-41 associated fHbp-1 became most prevalent. Although it increased in prevalence over time, fHbp-1 was the only variant that was present in both sub-complexes at all time periods. Two additional surface proteins, PorA and FetA showed a similar trend to fHbp. NHBA-2, the most frequent sub/variant in cc41/44, was also observed in both sub-complexes at all time periods. Given the intrinsic potential variability of NHBA, the maintenance of the same sub/variant in both sub-complexes over fifty years, despite changes in the genotype and in the predominance of the other protein variants, could be a consequence of selective pressure or fitness constraints.

We examined the broadest collection of pathogenic meningococcal isolates over the longest time span available globally. Significantly, we confirmed that the sequence

conservation of specific fHbp, NadA, and NHBA sub/variants observed across strains and geographic regions in recent years has also been present over the last several decades. Thus, the selection of fHbp, NadA, and NHBA as antigens in 4CMenB is supported by current and past molecular epidemiology. The hypothesis that the stability of certain sub/variants and combinations of fHbp, NadA, and NHBA likely results from natural selection also supports earlier interpretations that these proteins contribute to meningococcal survival and pathogenesis or fitness [17, 95, 100-102]. Further, our results may help to support the long-term validity of fHbp, NHBA and NadA characterization and additional typing systems for meningococci currently being implemented.

Further studies are needed in order to verify whether the observations in this strain panel are generalizable. To our knowledge, no similar panels, composed of invasive meningococcal isolates collected over a so long time period, exist in other countries. Further work will also be needed to articulate the results of the present study with ongoing efforts to evaluate the clinical effects and evaluation of the strain coverage of 4CMenB [103]. The Meningococcal Antigen Typing System (MATS) has been recently described as a qualitative and quantitative assay to predict 4CMenB vaccine coverage. This assay measures for each strain the expression level and the cross-reactivity of each vaccine antigen [5]. It will be important in future to apply MATS to old and new isolates to evaluate the temporal dynamics of changes in epidemiology and of potential antigenic shift for the vaccine-target antigens in normal condition and following vaccine introduction.

A limitation of the present study is that it comprises genetic data only. The study of temporal patterns of genetic associations among vaccine protein variants and MLST clonal complexes is important, even if as an initial step. Also, protein expression was evaluated for genes that have an insertion or a frame shift mutation, only. As protein expression would yield even greater differences in bactericidal titer, the evaluation of MATS results in this panel would be very interesting.

Another limitation is that only 165 invasive isolates randomly selected were tested over a period of 50 years, and how representative are the selected isolates could appear to be actually doubtful. However, the number of isolates for each clonal complex indeed reflects the relative incidence of clonal complexes and represents a substantial spectrum of different serogroup B meningococci in the Netherlands over the last decades. In addition, carriage isolates were not included. Limiting selection to pathogenic isolates may have resulted in the over-representation of hypervirulent lineages, instead of a more even balance of all meningococcal clones [104]. Moreover, we cannot comment on the two-

variant fHbp vaccine currently in clinical trials because only one of the two variants, fHbp-3.45, was identified in this study. Surprisingly, only four isolates at one time point harbored this fHbp variant, whereas fHbp-1.55, was not found in the current strain panel.

Data obtained in this study highlight the importance of monitoring over time the evolutionary pattern of surface proteins included as vaccine antigens. The stability of certain sub/variants was of course observed in a pre-vaccination era, therefore in the absence of a strong immune selection against the three antigen sub/variants. It will be interesting to monitor the long-term persistence even after the introduction of the vaccine.

Stability and longevity suggest that several fHbp, NHBA and NadA sub/variants are maintained by selection despite the fact that recombination continuously generates new sub/variants. In particular, the long-term persistence of the three antigen sub/variants included in the vaccine, fHbp-1.1, NHBA-2 and NadA-3.8, may be indicative for long term broad coverage of 4CMenB.

Appendix – Supplementary Material

2/12/14

					fHbp			NHBA			NadA		
Species	# of strains	Strain	Accession number	N. of contigs	fHbp % identity vs 1.1	fHbp % identity vs 1.1 - species average	fHbp notes after manual check	NHBA % identity vs peptide 2	NHBA % identity vs peptide 2 - species average	NHBA notes after manual check	NadA % identity vs NadA-3.8	NadA % identity vs NadA-3.8 - species average	NadA notes after manual check
Neisseria_bacilliformis	1	Neisseria_bacilliformis_ATCC_BAA_1200	AFAY00000000	69	28.3	28.3	weak homology						
Neisseria_cinerea	5	Neisseria_cinerea_ATCC_14685	ACDP00000000, ATCC14685 §	35, 35	96.6	93.9	Variant 1	27.5	27.5	weak homology in different locus, in the locus the gene is replaced by an intergenic region	66.03	54.5	~200aa at the border of a contig
		Neisseria_cinerea_CCUG25879	CCUG25879 §	249	93.3		Variant 1			the locus is interrupted by a contig edge, but the upstream locus contains the intergenic region instead of the gene			
		Neisseria_cinerea_CCUG27178A	CCUG27178A §	274	94.9		Variant 1			in the locus the gene is replaced by an intergenic region			
		Neisseria_cinerea_CCUG346T	CCUG346T §	299	94.9		Variant 1			the locus is interrupted by a contig edge, but the upstream locus contains the intergenic region instead of the gene	43		
		Neisseria_cinerea_CCUG5746	CCUG5746 §	250	80.8		Variant 1			the locus is higly fragmented in multiple contigs			
Neisseria_elongata_glycolytica	1	Neisseria_elongata_glycolytica_ATCC_29315	ADBF00000000	153	31.8	31.8	weak homology						
Neisseria_flavescens	2	Neisseria_flavescens_NRL30031_H210	ACEN00000000	117	42	41.6	weak homology	95.5	61.2	contig edge	25.68	25.1	
		Neisseria_flavescens_SK114	ACQV00000000	31	41.2		weak homology	26.8		weak homology	24.42		
Neisseria_gonorrhoeae	17	Neisseria_gonorrhoeae_1291	ABZF00000000	175	63	63.3	Variant 3	82.4	81.2				
		Neisseria_gonorrhoeae_35_02	ABZG00000000	156	63		Variant 3	82.2					
		Neisseria_gonorrhoeae_DG18	ABZH00000000	153	63		Variant 3	82.8					
		Neisseria_gonorrhoeae_DG2	ACIG00000000	134	64.9		frameshifted, Variant 3	82.4					
		Neisseria_gonorrhoeae_F62	ADAA00000000	134	62.6		Variant 3	82.8					
		Neisseria_gonorrhoeae_FA_1090	AE004969	1	63		Variant 3	82.8					
		Neisseria_gonorrhoeae_FA19	ABZI00000000	168	63		Variant 3	82.2					
		Neisseria_gonorrhoeae_FA6140	ABZK00000000	161	62.6		Variant 3	82.8					
		Neisseria_gonorrhoeae_MS11	ABZK00000000	196	64.9		frameshifted, Variant 3	82.5					
		Neisseria_gonorrhoeae_NCCP11945	CP001050	1	63		Variant 3	82.4					
		Neisseria_gonorrhoeae_PID1	ABZM00000000	145	64.5		frameshifted, Variant 3	62.1		400bp deletion, frameshifted			
		Neisseria_gonorrhoeae_PID18	ABZL00000000	175	64.9		frameshifted, Variant 3	82.4					
		Neisseria_gonorrhoeae_PID24 .1	ABZN00000000	168	62.6		Variant 3	87.4		contig edge			
		Neisseria_gonorrhoeae_PID332	ABZO00000000	168	62.6		Variant 3	82.4					
		Neisseria_gonorrhoeae_SK_92_679	ABZP00000000	197	62.6		Variant 3	82.2					
		Neisseria_gonorrhoeae_SK_93_1035	ABZQ00000000	156	62.6		Variant 3	76.5		frameshifted			
		Neisseria_gonorrhoeae_TCDC_NG08107	CP002440	1	62.6		Variant 3	82.2					
Neisseria_lactamica	8	**Neisseria_lactamica_020_06**	FN995097, 020-06 §	1, 1		27.9	in the locus OPA like proten NLA_18150	86.4	83.8			56.0	
		Neisseria_lactamica_ATCC_23970	ACEQ00000000, 23970 §	101, 101	24.6		weak homology in different locus, in the locus OPA like proten homologous to NLA_18150	84.9			53.44		different locus, YadA-like C-term region (120aa) common to a family of surface-exposed bacterial proteins
		Neisseria_lactamica_NS19	AEPI00000000	132	31.2		weak homology in different locus, in the locus OPA like proten homologous to NLA_18150, upstream not conserved, downstream conserved	80.5			47.3		~120aa at C-term
		Neisseria_lactamica_Y92_1009	CACL00000000, Y92-1009 §	44, 44			in the locus OPA like proten NLA_18150	84.0			67.14		different locus, YadA-like C-term region (120aa) common to a family of surface-exposed bacterial proteins
		Neisseria_lactamica_014-24	014-24 §	439			2.8 kb insertion in upstream, in the locus OPA like proten homologous to NLA_18150, downstream not conserved	89.1					
		Neisseria_lactamica_030-24	030-24 §	389			in the locus OPA like proten NLA_18150	83.3		contig edge			
		Neisseria_lactamica_0900251	0900251 §	431			in the locus OPA like proten NLA_18150, upstream not conserved, downstream conserved	81.1					
		Neisseria_lactamica_8206	8206 §	574			in the locus OPA like proten NLA_18150	81.2		contig edge			
Neisseria_macacae	1	Neisseria_macacae_ATCC_33926	AFQE00000000	200	42.6	42.6	weak homology, different locus				39.9	39.9	weak homology
Neisseria_meningitidis	27	**Neisseria_meningitidis_053442 (s. C)**	CP003381	1	95.6	83.7	Variant 1	87.3	82.6			89.2	
		Neisseria_meningitidis_8013 (s. C)	FM999788	1	74.7		Variant 2	85.6			99.51		
		Neisseria_meningitidis_961_5945 (s. B)	AEQK00000000	166	74		Variant 2	84.0			94.75		
		Neisseria_meningitidis_alpha14 (cnl)	AM889136	1	71.8		Variant 2	82.2					
		Neisseria_meningitidis_alpha710 (s. B)	CP001561	1	70.3		Variant 2	87.1					
		Neisseria_meningitidis_ATCC_13091 (s. B)	AEFF00000000	147	74.7		Variant 2	87.7					
		Neisseria_meningitidis_CU385 (s. B)	AEQJ00000000	47	99.6		Variant 1	90.2			95.58		
		Neisseria_meningitidis_ES14902 (s. B)	AEQI00000000	47	93.1		Variant 1	84.0			ND		frameshifted
		Neisseria_meningitidis_FAM18 (s. C)	AM421808	1	71.1		Variant 2	82.6			98.99		
		Neisseria_meningitidis_G2136 (s. B)	CP002419	1	74		Variant 2	84.0			100		
		Neisseria_meningitidis_H44_76 (s. B)	AEQZ00000000, CP002420	46, 1	99.6		Variant 1	91.4					
		Neisseria_meningitidis_K1207 (s. C)	ADWM00000000	222	93.4		Variant 1	72.3			ND		contig interruption
		Neisseria_meningitidis_M01_240013 (s. B)	AEQL00000000	58	73.6		Variant 2	83.3					
		Neisseria_meningitidis_M01_240149 (s. B)	CP002421	1	95.6		Variant 1	100.0					
		Neisseria_meningitidis_M01_240355 (s. B)	CP002422	1	67		Variant 3	82.8			25.31		possible frameshift
		Neisseria_meningitidis_M04_240196 (s. B)	CP002423	1	88.5		Variant 1	83.8					
		Neisseria_meningitidis_M07/9 (s. B)	AEQH00000000	30	73.5		Variant 2	100.0					
		Neisseria_meningitidis_M13399 (s. B)	AEQG00000000	44	64.1		Variant 3	84.7					
		Neisseria_meningitidis_M6190 (s. B)	AEQF00000000	46	93.8		Variant 1	84.0			ND		IS4 insertion
		Neisseria_meningitidis_MC58 (s. B)	AE002098	1	100		Variant 1	91.4			95.58		
		Neisseria_meningitidis_N1568 (s. X)	AEQD00000000	112	92.1		frameshifted but low quality sequence	88.4					
		Neisseria_meningitidis_NS44 (s. Y)	AEPI00000000	365	69.7		Variant 2	76.4					
		Neisseria_meningitidis_NZ_05_33 (s. B)	CP002424	1	91.2		Variant 1	100.0					
		Neisseria_meningitidis_OX99_30304 (s. B)	AEQO00000000	128	73.6		Variant 2	82.6					
		Neisseria_meningitidis_S0108 (s. C)	ADWN00000000	226	93.4		Variant 1	72.3			ND		contig interruption
		Neisseria_meningitidis_WUE_2594 (s. A)	FR774048	1	95.6		Variant 1	82.4			100		
		Neisseria_meningitidis_Z2491 (s. A)	AL157959	1	95.6		Variant 1	82.6					
Neisseria_mucosa	2	Neisseria_mucosa_ATCC_25996	ACJD00000000	67	22.8	32.0	weak homology, different locus	28.0	24.7	weak homology	42.72	42.7	different locus, YadA-like C-term region (120aa) common to a family of surface-exposed bacterial proteins
		Neisseria_mucosa_C102	ACRG00000000	23	41.2		weak homology, different locus	21.4		weak homology			
Neisseria_oral_taxon	1	Neisseria_oral_taxon_014_F0314	ADEA00000000	48	29.2	29.2	weak homology, different locus				20.72	20.7	
Neisseria_polysaccharea	7	Neisseria_polysaccharea_ATCC_43768	ADBE00000000, ATCC43768 §	139, 139	75.4	72.5	frameshifted, Variant 3	87.5	84.3				
		Neisseria_polysaccharea_NS342	AEPH00000000	288	63.2		frameshifted, Variant 3	85.6					
		Neisseria_polysaccharea_15883	15883 §	7	72		Variant 3	82.2					
		Neisseria_polysaccharea_CCUG18031	CCUG18031 §	309	75.4		frameshifted, Variant 3	84.2					
		Neisseria_polysaccharea_CCUG24845	CCUG24845 §	292	75.4		frameshifted, Variant 3	83.9					
		Neisseria_polysaccharea_CCUG24846	CCUG24846 §	323	71.1		contig edge, Variant 3	83.2		contig edge			
		Neisseria_polysaccharea_CCUG27182	CCUG27182 §	503	75.2		frameshifted, Variant 3	83.6		contig edge			
Neisseria_shayeganii	1	Neisseria_shayeganii_871	AGAY00000000	112							28.3	28.3	weak homology
Neisseria_sicca	3	Neisseria_sicca_4320	AEPD00000000	169	40.8	41.9	weak homology	25.8	25.8	weak homology	37.3	37.3	weak homology
		Neisseria_sicca_ATCC_29256	ACKO00000000	61	43.8		weak homology	20.2		weak homology	42.47	42.5	different locus, YadA-like C-term region (120aa) common to a family of surface-exposed bacterial proteins
		Neisseria_sicca_D51	AEPG00000000	455	41.1		weak homology	28.9		weak homology	36.8	36.8	weak homology
Neisseria_subflava	1	Neisseria_subflava_NJ9703	ACEO00000000	63	41.6	41.6	weak homology	20.3	20.3	weak homology			
Neisseria_wadsworthii	1	Neisseria_wadsworthii_9715	AGAZ00000000	97									
Neisseria_weaveri	2	Neisseria_weaveri_ATCC_51223	AFWB00000000	40	27.9	28.9	weak homology						
		Neisseria_weaveri_LMG_5135	AFWQ00000000	46	30		weak homology						

Fasta cut-offs
weak homology (> e-30)
not found > e-5

% of identity are vs Vaccine variants

Table S1 The complete table of the isolates. The serogroup typing was indicated in parentheses for *N. meningitidis* strains.

References

1. Giuliani MM, Adu-Bobie J, Comanducci M, Arico B, Savino S, Santini L, Brunelli B, Bambini S, Biolchi A, Capecchi B *et al*: **A universal vaccine for serogroup B meningococcus**. *Proc Natl Acad Sci U S A* 2006, **103**(29):10834-10839.
2. Parkhill J, Achtman M, James KD, Bentley SD, Churcher C, Klee SR, Morelli G, Basham D, Brown D, Chillingworth T *et al*: **Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491**. *Nature* 2000, **404**(6777):502-506.
3. Masignani V, Comanducci M, Giuliani MM, Bambini S, Adu-Bobie J, Arico B, Brunelli B, Pieri A, Santini L, Savino S *et al*: **Vaccination against Neisseria meningitidis using three variants of the lipoprotein GNA1870**. *J Exp Med* 2003, **197**(6):789-799.
4. Wang X, Cohn A, Comanducci M, Andrew L, Zhao X, MacNeil JR, Schmink S, Muzzi A, Bambini S, Rappuoli R *et al*: **Prevalence and genetic diversity of candidate vaccine antigens among invasive Neisseria meningitidis isolates in the United States**. *Vaccine* 2011, **29**(29-30):4739-4744.
5. Vogel U, Taha MK, Vazquez JA, Findlow J, Claus H, Stefanelli P, Caugant DA, Kriz P, Abad R, Bambini S *et al*: **Predicted strain coverage of a meningococcal multicomponent vaccine (4CMenB) in Europe: a qualitative and quantitative assessment**. *The Lancet infectious diseases* 2013.
6. Smith NH, Holmes EC, Donovan GM, Carpenter GA, Spratt BG: **Networks and groups within the genus Neisseria: analysis of argF, recA, rho, and 16S rRNA sequences from human Neisseria species**. *Mol Biol Evol* 1999, **16**(6):773-783.
7. Bennett JS, Jolley KA, Earle SG, Corton C, Bentley SD, Parkhill J, Maiden MC: **A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus Neisseria**. *Microbiology* 2012, **158**(Pt 6):1570-1580.
8. Marri PR, Paniscus M, Weyand NJ, Rendon MA, Calton CM, Hernandez DR, Higashi DL, Sodergren E, Weinstock GM, Rounsley SD *et al*: **Genome sequencing reveals widespread virulence gene exchange among human Neisseria species**. *PLoS One* 2010, **5**(7):e11835.
9. Vazquez JA, de la Fuente L, Berron S, O'Rourke M, Smith NH, Zhou J, Spratt BG: **Ecological separation and genetic isolation of Neisseria gonorrhoeae and Neisseria meningitidis**. *Curr Biol* 1993, **3**(9):567-572.
10. Madico G, Welsch JA, Lewis LA, McNaughton A, Perlman DH, Costello CE, Ngampasutadol J, Vogel U, Granoff DM, Ram S: **The meningococcal vaccine candidate GNA1870 binds the complement regulatory protein factor H and enhances serum resistance**. *J Immunol* 2006, **177**(1):501-510.
11. Seib KL, Serruto D, Oriente F, Delany I, Adu-Bobie J, Veggi D, Arico B, Rappuoli R, Pizza M: **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* 2009, **77**(1):292-299.
12. Seib KL, Brunelli B, Brogioni B, Palumbo E, Bambini S, Muzzi A, DiMarcello F, Marchi S, van der Ende A, Arico B *et al*: **Characterization of diverse subvariants of the meningococcal factor H (fH) binding protein for their ability to bind fH, to mediate serum resistance, and to induce bactericidal antibodies**. *Infect Immun* 2011, **79**(2):970-981.
13. Lucidarme J, Comanducci M, Findlow J, Gray SJ, Kaczmarski EB, Guiver M, Vallely PJ, Oster P, Pizza M, Bambini S *et al*: **Characterization of fHbp, nhba (gna2132), nadA, porA, and sequence type in group B meningococcal case isolates collected in England and Wales during January 2008 and potential coverage of an investigational group B meningococcal vaccine**. *Clin Vaccine Immunol* 2010, **17**(6):919-929.
14. Hadad R, Jacobsson S, Pizza M, Rappuoli R, Fredlund H, Olcén P, Unemo M: **Novel meningococcal 4CMenB vaccine antigens – prevalence and polymorphisms of the encoding genes in Neisseria gonorrhoeae**. *APMIS* 2012:n/a-n/a.

15. Fletcher LD, Bernfield L, Barniak V, Farley JE, Howell A, Knauf M, Ooi P, Smith RP, Weise P, Wetherell M *et al*: **Vaccine potential of the Neisseria meningitidis 2086 lipoprotein**. *Infect Immun* 2004, **72**(4):2088-2100.
16. Pizza M, Scarlato V, Masignani V, Giuliani MM, Arico B, Comanducci M, Jennings GT, Baldi L, Bartolini E, Capecchi B *et al*: **Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing**. *Science* 2000, **287**(5459):1816-1820.
17. Serruto D, Spadafina T, Ciucchi L, Lewis LA, Ram S, Tontini M, Santini L, Biolchi A, Seib KL, Giuliani MM *et al*: **Neisseria meningitidis GNA2132, a heparin-binding protein that induces protective immunity in humans**. *Proc Natl Acad Sci U S A* 2010, **107**(8):3770-3775.
18. Jacobsson S, Hedberg ST, Molling P, Unemo M, Comanducci M, Rappuoli R, Olcen P: **Prevalence and sequence variations of the genes encoding the five antigens included in the novel 5CVMB vaccine covering group B meningococcal disease**. *Vaccine* 2009, **27**(10):1579-1584.
19. Cotter SE, Surana NK, St Geme JW, 3rd: **Trimeric autotransporters: a distinct subfamily of autotransporter proteins**. *Trends Microbiol* 2005, **13**(5):199-205.
20. Lyskowski A, Leo JC, Goldman A: **Structure and biology of trimeric autotransporter adhesins**. *Adv Exp Med Biol* 2011, **715**:143-158.
21. Comanducci M, Bambini S, Brunelli B, Adu-Bobie J, Arico B, Capecchi B, Giuliani MM, Masignani V, Santini L, Savino S *et al*: **NadA, a novel vaccine candidate of Neisseria meningitidis**. *J Exp Med* 2002, **195**(11):1445-1454.
22. Comanducci M, Bambini S, Caugant DA, Mora M, Brunelli B, Capecchi B, Ciucchi L, Rappuoli R, Pizza M: **NadA diversity and carriage in Neisseria meningitidis**. *Infect Immun* 2004, **72**(7):4217-4223.
23. Harris SL, Zhu D, Murphy E, McNeil LK, Wang X, Mayer LW, Harrison LH, Jansen KU, Anderson AS: **Preclinical evidence for the potential of a bivalent fHBP vaccine to prevent Neisseria meningitidis Serogroup C Disease**. *Human vaccines* 2011, **7 Suppl**:68-74.
24. Hong E, Giuliani MM, Deghmane AE, Comanducci M, Brunelli B, Dull P, Pizza M, Taha MK: **Could the multicomponent meningococcal serogroup B vaccine (4CMenB) control Neisseria meningitidis capsular group X outbreaks in Africa?** *Vaccine* 2013, **31**(7):1113-1116.
25. Beernink PT, Caugant DA, Welsch JA, Koeberling O, Granoff DM: **Meningococcal factor H-binding protein variants expressed by epidemic capsular group A, W-135, and X strains from Africa**. *The Journal of infectious diseases* 2009, **199**(9):1360-1368.
26. Pajon R, Fergus AM, Koeberling O, Caugant DA, Granoff DM: **Meningococcal factor H binding proteins in epidemic strains from Africa: implications for vaccine development**. *PLoS neglected tropical diseases* 2011, **5**(9):e1302.
27. Tamura K: **Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C-content biases**. *Mol Biol Evol* 1992, **9**(4):678-687.
28. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods**. *Mol Biol Evol* 2011, **28**(10):2731-2739.
29. Huson DH, Bryant D: **Application of phylogenetic networks in evolutionary studies**. *Mol Biol Evol* 2006, **23**(2):254-267.
30. Kloepper TH, Huson DH: **Drawing explicit phylogenetic networks and their integration into SplitsTree**. *BMC Evol Biol* 2008, **8**:22.
31. McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P: **The fine-scale structure of recombination rate variation in the human genome**. *Science* 2004, **304**(5670):581-584.
32. Watterson GA: **On the number of segregating sites in genetical models without recombination**. *Theor Popul Biol* 1975, **7**(2):256-276.
33. Martin DP, Lemey P, Lott M, Moulton V, Posada D, Lefeuvre P: **RDP3: a flexible and fast computer program for analyzing recombination**. *Bioinformatics* 2010, **26**(19):2462-2463.
34. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: **Versatile and open software for comparing large genomes**. *Genome biology* 2004, **5**(2):R12.

35. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees**. *Mol Biol Evol* 1987, **4**(4):406-425.
36. Tamura K, Nei M, Kumar S: **Prospects for inferring very large phylogenies by using the neighbor-joining method**. *Proc Natl Acad Sci U S A* 2004, **101**(30):11030-11035.
37. Otto TD, Dillon GP, Degrave WS, Berriman M: **RATT: Rapid Annotation Transfer Tool**. *Nucleic Acids Res* 2011, **39**(9):e57.
38. Pearson WR, Wood T, Zhang Z, Miller W: **Comparison of DNA sequences with protein sequences**. *Genomics* 1997, **46**(1):24-36.
39. Bambini S, Muzzi A, Olcen P, Rappuoli R, Pizza M, Comanducci M: **Distribution and genetic variability of three vaccine components in a panel of strains representative of the diversity of serogroup B meningococcus**. *Vaccine* 2009, **27**(21):2794-2803.
40. Darling AE, Mau B, Perna NT: **progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement**. *PLoS One* 2010, **5**(6):e11147.
41. Feil EJ, Maiden MC, Achtman M, Spratt BG: **The relative contributions of recombination and mutation to the divergence of clones of Neisseria meningitidis**. *Mol Biol Evol* 1999, **16**(11):1496-1502.
42. Feil EJ, Holmes EC, Bessen DE, Chan MS, Day NP, Enright MC, Goldstein R, Hood DW, Kalia A, Moore CE *et al*: **Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences**. *Proc Natl Acad Sci U S A* 2001, **98**(1):182-187.
43. Hanage WP, Fraser C, Spratt BG: **Fuzzy species among recombinogenic bacteria**. *BMC Biol* 2005, **3**:6.
44. Lucidarme J, Tan L, Exley RM, Findlow J, Borrow R, Tang CM: **Characterization of Neisseria meningitidis isolates that do not express the virulence factor and vaccine antigen factor H binding protein**. *Clinical and vaccine immunology : CVI* 2011, **18**(6):1002-1014.
45. Beernink PT, Granoff DM: **The modular architecture of meningococcal factor H-binding protein**. *Microbiology* 2009, **155**(Pt 9):2873-2883.
46. Bennett JS, Thompson EA, Kriz P, Jolley KA, Maiden MC: **A common gene pool for the Neisseria FetA antigen**. *Int J Med Microbiol* 2009, **299**(2):133-139.
47. Martin DP: **Recombination detection and analysis using RDP3**. *Methods Mol Biol* 2009, **537**:185-205.
48. Caugant DA, Maiden MC: **Meningococcal carriage and disease--population biology and evolution**. *Vaccine* 2009, **27 Suppl 2**:B64-70.
49. Sim RJ, Harrison MM, Moxon ER, Tang CM: **Underestimation of meningococci in tonsillar tissue by nasopharyngeal swabbing**. *Lancet* 2000, **356**(9242):1653-1654.
50. Finne J, Finne U, Deagostini-Bazin H, Goridis C: **Occurrence of alpha 2-8 linked polysialosyl units in a neural cell adhesion molecule**. *Biochem Biophys Res Commun* 1983, **112**(2):482-487.
51. Sierra GV, Campa HC, Varcacel NM, Garcia IL, Izquierdo PL, Sotolongo PF, Casanueva GV, Rico CO, Rodriguez CR, Terry MH: **Vaccine against group B Neisseria meningitidis: protection trial and mass vaccination results in Cuba**. *NIPH Ann* 1991, **14**(2):195-207; discussion 208-110.
52. Bjune G, Hoiby EA, Gronnesby JK, Arnesen O, Fredriksen JH, Halstensen A, Holten E, Lindbak AK, Nokleby H, Rosenqvist E *et al*: **Effect of outer membrane vesicle vaccine against group B meningococcal disease in Norway**. *Lancet* 1991, **338**(8775):1093-1096.
53. Martin DR, Walker SJ, Baker MG, Lennon DR: **New Zealand epidemic of meningococcal disease identified by a strain with phenotype B:4:P1.4**. *The Journal of infectious diseases* 1998, **177**(2):497-500.
54. Jackson C, Lennon DR, Sotutu VT, Yan J, Stewart JM, Reid S, Crengle S, Oster P, Ypma E, Aaberge I *et al*: **Phase II meningococcal B vesicle vaccine trial in New Zealand infants**. *Arch Dis Child* 2009, **94**(10):745-751.
55. Metruccio MM, Fantappie L, Serruto D, Muzzi A, Roncarati D, Donati C, Scarlato V, Delany I: **The Hfq-dependent small noncoding RNA NrrF directly mediates Fur-dependent positive regulation of succinate dehydrogenase in Neisseria meningitidis**. *J Bacteriol* 2009, **191**(4):1330-1342.

56. Welsch JA, Moe GR, Rossi R, Adu-Bobie J, Rappuoli R, Granoff DM: **Antibody to genome-derived neisserial antigen 2132, a Neisseria meningitidis candidate vaccine, confers protection against bacteremia in the absence of complement-mediated bactericidal activity**. *The Journal of infectious diseases* 2003, **188**(11):1730-1740.
57. Lucidarme J, Comanducci M, Findlow J, Gray SJ, Kaczmarski EB, Guiver M, Kugelberg E, Vallely PJ, Oster P, Pizza M *et al*: **Characterization of fHbp, nhba (gna2132), nadA, porA, sequence type (ST), and genomic presence of IS1301 in group B meningococcal ST269 clonal complex isolates from England and Wales**. *J Clin Microbiol* 2009, **47**(11):3577-3585.
58. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA *et al*: **Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms**. *Proc Natl Acad Sci U S A* 1998, **95**(6):3140-3145.
59. Thompson EA, Feavers IM, Maiden MC: **Antigenic diversity of meningococcal enterobactin receptor FetA, a vaccine component**. *Microbiology* 2003, **149**(Pt 7):1849-1858.
60. Russell JE, Jolley KA, Feavers IM, Maiden MC, Suker J: **PorA variable regions of Neisseria meningitidis**. *Emerg Infect Dis* 2004, **10**(4):674-678.
61. Brehony C, Wilson DJ, Maiden MC: **Variation of the factor H-binding protein of Neisseria meningitidis**. *Microbiology* 2009, **155**(Pt 12):4155-4169.
62. Jolley KA, Kalmusova J, Feil EJ, Gupta S, Musilek M, Kriz P, Maiden MC: **Carried meningococci in the Czech Republic: a diverse recombining population**. *J Clin Microbiol* 2000, **38**(12):4492-4498.
63. Yazdankhah SP, Kriz P, Tzanakaki G, Kremastinou J, Kalmusova J, Musilek M, Alvestad T, Jolley KA, Wilson DJ, McCarthy ND *et al*: **Distribution of serogroups and genotypes among disease-associated and carried isolates of Neisseria meningitidis from the Czech Republic, Greece, and Norway**. *J Clin Microbiol* 2004, **42**(11):5146-5153.
64. Callaghan MJ, Buckee C, McCarthy ND, Ibarz Pavon AB, Jolley KA, Faust S, Gray SJ, Kaczmarski EB, Levin M, Kroll JS *et al*: **Opa protein repertoires of disease-causing and carried meningococci**. *J Clin Microbiol* 2008, **46**(9):3033-3041.
65. Buckee CO, Jolley KA, Recker M, Penman B, Kriz P, Gupta S, Maiden MC: **Role of selection in the emergence of lineages and the evolution of virulence in Neisseria meningitidis**. *Proc Natl Acad Sci U S A* 2008, **105**(39):15082-15087.
66. Buckee CO, Gupta S, Kriz P, Maiden MC, Jolley KA: **Long-term evolution of antigen repertoires among carried meningococci**. *Proceedings Biological sciences / The Royal Society* 2010, **277**(1688):1635-1641.
67. Jolley KA, Chan MS, Maiden MC: **mlstdbNet - distributed multi-locus sequence typing (MLST) databases**. *BMC Bioinformatics* 2004, **5**:86.
68. R Development Core Team: **R: A Language and Environment for Statistical Computing**. In. Vienna, Austria; 2011.
69. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput**. *Nucleic acids research* 2004, **32**(5):1792-1797.
70. Francisco AP, Bugalho M, Ramirez M, Carrico JA: **Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach**. *BMC Bioinformatics* 2009, **10**:152.
71. Kriz P, Giorgini D, Musilek M, Larribe M, Taha MK: **Microevolution through DNA exchange among strains of Neisseria meningitidis isolated during an outbreak in the Czech Republic**. *Research in microbiology* 1999, **150**(4):273-280.
72. Marsh JW, Shutt KA, Pajon R, Tulenko MM, Liu S, Hollick RA, Kiehlbauch JA, Clark TA, Stephens DS, Arnold KE *et al*: **Diversity of factor H-binding protein in Neisseria meningitidis carriage isolates**. *Vaccine* 2011, **29**(35):6049-6058.
73. Elias J, Vogel U: **IS1301 fingerprint analysis of Neisseria meningitidis strains belonging to the ET-15 clone**. *J Clin Microbiol* 2007, **45**(1):159-167.
74. Anderson AS, Hao L, Jiang Q, Harris SL, Jones TR, Perez JL, York L, Eiden J, Jansen KU: **Potential impact of the bivalent rLP2806 vaccine on Neisseria meningitidis carriage and invasive serogroup B disease**. *Human vaccines & immunotherapeutics* 2012, **9**(3).

75. Giuliani MM, Biolchi A, Serruto D, Ferlicca F, Vienken K, Oster P, Rappuoli R, Pizza M, Donnelly J: **Measuring antigen-specific bactericidal responses to a multicomponent vaccine against serogroup B meningococcus**. *Vaccine* 2010, **28**(31):5023-5030.
76. Fagnocchi L, Biolchi A, Ferlicca F, Boccadifuoco G, Brunelli B, Brier S, Norais N, Chiarot E, Bensi G, Kroll JS *et al*: **Transcriptional regulation of the nadA gene in Neisseria meningitidis impacts the prediction of coverage of a multicomponent meningococcal serogroup B vaccine**. *Infect Immun* 2013, **81**(2):560-569.
77. Jolley KA, Wilson DJ, Kriz P, McVean G, Maiden MC: **The influence of mutation, recombination, population history, and selection on patterns of genetic diversity in Neisseria meningitidis**. *Mol Biol Evol* 2005, **22**(3):562-569.
78. Caugant DA, Mocca LF, Frasch CE, Froholm LO, Zollinger WD, Selander RK: **Genetic structure of Neisseria meningitidis populations in relation to serogroup, serotype, and outer membrane protein pattern**. *J Bacteriol* 1987, **169**(6):2781-2792.
79. Maiden MCJ: **Multilocus sequence typing of bacteria**. In: *Annual Review of Microbiology.* vol. 60; 2006: 561-588.
80. Brehony C, Jolley KA, Maiden MCJ: **Multilocus sequence typing for global surveillance of meningococcal disease**. *FEMS microbiology reviews* 2007, **31**(1):15-26.
81. Caugant DA, Tzanakaki G, Kriz P: **Lessons from meningococcal carriage studies**. *FEMS microbiology reviews* 2007, **31**(1):52-63.
82. Jodar L, Feavers IM, Salisbury D, Granoff DM: **Development of vaccines against meningococcal disease**. *Lancet* 2002, **359**(9316):1499-1508.
83. Vanderley P, Poolman JT: **CONSTRUCTION OF A MULTIVALENT MENINGOCOCCAL VACCINE STRAIN BASED ON THE CLASS-1 OUTER-MEMBRANE PROTEIN**. *Infection and Immunity* 1992, **60**(8):3156-3161.
84. Peeters C, Rumke HC, Sundermann LC, vanderVoort EMR, Meulenbelt J, Schuller M, Kuipers AJ, vanderLey P, Poolman JT: **Phase I clinical trial with a hexavalent PorA containing meningococcal outer membrane vesicle vaccine**. *Vaccine* 1996, **14**(10):1009-1015.
85. Cartwright K, Morris R, Rumke H, Fox A, Borrow R, Begg N, Richmond P, Poolman J: **Immunogenicity and reactogenicity in UK infants of a novel meningococcal vesicle vaccine containing multiple class 1 (PorA) outer membrane proteins**. *Vaccine* 1999, **17**(20-21):2612-2619.
86. Rappuoli R: **Reverse vaccinology, a genome-based approach to vaccine development**. *Vaccine* 2001, **19**(17-19):2688-2691.
87. Tettelin H, Saunders NJ, Heidelberg J, Jeffries AC, Nelson KE, Eisen JA, Ketchum KA, Hood DW, Peden JF, Dodson RJ *et al*: **Complete genome sequence of Neisseria meningitidis serogroup B strain MC58**. *Science* 2000, **287**(5459):1809-1815.
88. Heckenberg SGB, de Gans J, Brouwer MC, Weisfelt M, Piet JR, Spanjaard L, van der Ende A, van de Beek D: **Clinical features, outcome, and meningococcal genotype in 258 adults with meningococcal meningitis - A prospective cohort study**. *Medicine* 2008, **87**(4):185-192.
89. Elias J, Schouls LM, van de Pol I, Keijzers WC, Martin DR, Glennie A, Oster P, Frosch M, Vogel U, van der Ende A: **Vaccine preventability of meningococcal clone, Greater Aachen Region, Germany**. *Emerg Infect Dis* 2010, **16**(3):465-472.
90. Bart A, Dankert J, van der Ende A: **Representational difference analysis of Neisseria meningitidis identifies sequences that are specific for the hyper-virulent lineage III clone**. *Fems Microbiology Letters* 2000, **188**(2):111-114.
91. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0**. *Mol Biol Evol* 2007, **24**(8):1596-1599.
92. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ: **Jalview Version 2-a multiple sequence alignment editor and analysis workbench**. *Bioinformatics* 2009, **25**(9):1189-1191.
93. Cramér H: **Mathematical Methods of Statistics**. In: *Mathematical Methods of Statistics.* Edited by Cramér H: Princeton: Princeton University Press; 1946.
94. Haubold B, Hudson RR: **LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage Analysis**. *Bioinformatics* 2000, **16**(9):847-848.
95. Capecchi B, Adu-Bobie J, Di Marcello F, Ciucchi L, Masignani V, Taddei A, Rappuoli R, Pizza M, Arico B: **Neisseria meningitidis NadA is a new invasin which promotes bacterial**

adhesion to and penetration into human epithelial cells. *Mol Microbiol* 2005, **55**(3):687-698.
96. Gupta S, Maiden MCJ, Feavers IM, Nee S, May RM, Anderson RM: **The maintenance of strain structure in populations of recombining infectious agents**. *Nature Medicine* 1996, **2**(4):437-442.
97. Van Looveren M, Vandamme P, Hauchecorne M, Wijdooghe M, Carion F, Caugant DA, Goossens H: **Molecular epidemiology of recent Belgian isolates of Neisseria meningitidis serogroup B**. *J Clin Microbiol* 1998, **36**(10):2828-2834.
98. Plested JS, Welsch JA, Granoff DM: **Ex Vivo Model of Meningococcal Bacteremia Using Human Blood for Measuring Vaccine-Induced Serum Passive Protective Activity**. *Clinical and Vaccine Immunology* 2009, **16**(6):785-791.
99. Welsch JA, Ram S, Koeberling O, Granoff DM: **Complement-dependent synergistic bactericidal activity of antibodies against factor H-binding protein, a sparsely distributed meningococcal vaccine antigen**. *Journal of Infectious Diseases* 2008, **197**(7):1053-1061.
100. Seib KL, Oriente F, Adu-Bobie J, Montanari P, Ferlicca F, Giuliani MM, Rappuoli R, Pizza M, Delany I: **Influence of serogroup B meningococcal vaccine antigens on growth and survival of the meningococcus in vitro and in ex vivo and in vivo models of infection**. *Vaccine* 2010, **28**(12):2416-2427.
101. Dunphy KY, Beernink PT, Brogioni B, Granoff DM: **Effect of Factor H-Binding Protein Sequence Variation on Factor H Binding and Survival of Neisseria meningitidis in Human Blood**. *Infection and Immunity* 2011, **79**(1):353-359.
102. Echenique-Rivera H, Muzzi A, Del Tordello E, Seib KL, Francois P, Rappuoli R, Pizza M, Serruto D: **Transcriptome analysis of Neisseria meningitidis in human whole blood and mutagenesis studies identify virulence factors involved in blood survival**. *PLoS pathogens* 2011, **7**(5):e1002027.
103. Donnelly J, Medini D, Boccadifuoco G, Biolchi A, Ward J, Frasch C, Moxon ER, Stella M, Comanducci M, Bambini S *et al*: **Qualitative and quantitative assessment of meningococcal antigens to evaluate the potential strain coverage of protein-based vaccines**. *Proc Natl Acad Sci U S A* 2010, **107**(45):19490-19495.
104. Didelot X, Maiden MCJ: **Impact of recombination on bacterial evolution**. *Trends in Microbiology* 2010, **18**(7):315-322.