MAJOR ARTICLE

# Sequence Diversity of the Factor H Binding Protein Vaccine Candidate in Epidemiologically Relevant Strains of Serogroup B *Neisseria meningitidis*

**Ellen Murphy,[1] Lubomira Andrew,[1] Kwok-Leung Lee,[1] Deborah A. Dilts,[1] Lorna Nunez,[1] Pamela S. Fink,[1] Karita Ambrose,[1] Ray Borrow,[4] Jamie Findlow,[4] Muhamed-Kheir Taha,[5] Ala-Eddine Deghmane,[5] Paula Kriz,[6] Martin Musilek,[6] Jitka Kalmusova,[6] Dominique A. Caugant,[7] Torill Alvestad,[7] Leonard W. Mayer,[2] Claudio T. Sacchi,[2,8] Xin Wang,[2] Diana Martin,[9] Anne von Gottberg,[10] Mignon du Plessis,[10] Keith P. Klugman,[3,10] Annaliesa S. Anderson,[1] Kathrin U. Jansen,[1] Gary W. Zlotnick,[1] and Susan K. Hoiseth[1]**

[1]Wyeth Vaccines Research, Pearl River, New York; [2]Centers for Disease Control and Prevention, and [3]Rollins School of Public Health, Emory University, Atlanta, Georgia, [4]Health Protection Agency, Manchester Royal Infirmary, Manchester, United Kingdom; [5]Institut Pasteur, Invasive Bacterial Infections Unit, Paris, France; [6]National Institute of Public Health, Prague, Czech Republic; [7]Norwegian Institute of Public Health, Oslo, Norway; [8]Instituto Adolfo Lutz, Sao Paulo, Brazil; [9]Institute of Environmental Science and Research, Porirua, New Zealand; and [10]Respiratory and Meningeal Pathogens Research Unit, National Institute for Communicable Diseases, Johannesburg, South Africa

***Background.*** Recombinant forms of *Neisseria meningitidis* human factor H binding protein (fHBP) are undergoing clinical trials in candidate vaccines against invasive meningococcal serogroup B disease. We report an extensive survey and phylogenetic analysis of the diversity of *fhbp* genes and predicted protein sequences in invasive clinical isolates obtained in the period 2000–2006.

***Methods.*** Nucleotide sequences of *fhbp* genes were obtained from 1837 invasive *N. meningitidis* serogroup B (MnB) strains from the United States, Europe, New Zealand, and South Africa. Multilocus sequence typing (MLST) analysis was performed on a subset of the strains.

***Results.*** Every strain contained the *fhbp* gene. All sequences fell into 1 of 2 subfamilies (A or B), with 60%–75% amino acid identity between subfamilies and at least 83% identity within each subfamily. One fHBP sequence may have arisen via inter-subfamily recombination. Subfamily B sequences were found in 70% of the isolates, and subfamily A sequences were found in 30%. Multiple fHBP variants were detected in each of the common MLST clonal complexes. All major MLST complexes include strains in both subfamily A and subfamily B.

***Conclusions.*** The diversity of strains observed underscores the importance of studying the distribution of the vaccine antigen itself rather than relying on common epidemiological surrogates such as MLST.

Invasive disease caused by *Neisseria meningitidis* is a rapidly progressing, disseminated infection with a case fatality rate of ~10% [1], and 10%–20% of survivors experience serious permanent sequelae (eg, neurological impairment, digit, limb, or hearing loss). Vaccines for meningococcal serogroups A, C, Y, and W135, which were initially based on the serogroup-definin polysaccharides and more recently on conjugates of the

Received 18 November 2008; accepted 11 March 2009; electronically published 17 June 2009.

Presented in part: International Pathogenic Neisseria Conference, Rotterdam, The Netherlands, 7–12 September 2008 (abstracts O45 and P137).

Financial support: Wyeth Vaccines Research, Institut Pasteur (M-K.T and A-E.D.), and the Internal Grant Agency, Ministry of Health of the Czech Republic (Grant 1A8688-3 to P. K.).

Reprints or correspondence: Susan K. Hoiseth, Ph.D., Wyeth Vaccines Research, 401 N. Middletown Rd., Pearl River, NY 10965 (hoiseths@wyeth.com).

**The Journal of Infectious Diseases 2009; 200:379–89**

0022-1899/2009/20003-0009$15.00
DOI: 10.1086/600141

Potential Conflicts of Interest: E.M., L.A., K.L.L., D.A.D., L.N., P.S.F., K.A., A.S.A., K.U.J., G.W.Z., and S.K.H. are employees of Wyeth. R.B., J.F., M.K.T., A.-E.D., P.K., M.M., J.K., D.A.C., T.A., L.W.M., C.T.S., X.W., D.M., A.v.G., M.d.P., and K.P.K. received funding from Wyeth for provision of the strain collections, associated strain information, and some of the sequence data reported here. R.B. has received assistance to attend scientific meetings from Wyeth, Sanofi Pasteur, and Baxter Bioscience; has served as an ad hoc consultant for Wyeth, GlaxoSmithKline, Novartis, Sanofi Pasteur, and Baxter Bioscience; and has performed other contract research funded by Wyeth, Novartis Vaccines, Baxter Bioscience, GlaxoSmithKline, Sanofi Pasteur, Alexion Pharmaceuticals, Emergent Europe, and Merck. J.F. has received assistance to attend scientific meetings from Wyeth and Novartis Vaccines. L.W.M. and X.W. have received funding (under Cooperative Research and Development Agreements) from Novartis Vaccines and Wyeth. K.P.K. has consulted for Wyeth and Novartis and has received other research funding from Wyeth. D.A.C. has served as an ad hoc consultant for Wyeth, Novartis, and Sanofi Pasteur. P.K. has served as an ad hoc consultant for Wyeth. A.V.G. has consulted for Norvartis. D.M. has received contract funding from Novartis. M.-K.T has also received funding from Chiron and Sanofi Pasteur for contract research and has consulted for Wyeth, GlaxoSmithKline, Sanofi Pasteur, and Novartis.

polysaccharides with protein carriers, have been available for several years. In countries in which these vaccines, particularly meningococcal serogroup C conjugate vaccines, have been adopted, disease caused by strains of these serogroups has significantl decreased [2]. However, a polysaccharide-based vaccine is not feasible for serogroup B (MnB) strains, because of the similarity of the MnB capsular polysaccharide, a homopolymer of ($\alpha$2→8) sialic acid, to the polysaccharide component of human neural cell adhesion molecule [3, 4]. Therefore, the focus of attention for MnB vaccines has shifted to surface-exposed protein antigens that are capable of eliciting protective immunity.

Factor H binding protein (fHBP), an ~28 kD lipoprotein, was firs identifie as a protective antigen through biochemical fractionation of a soluble outer membrane preparation that had been shown to elicit a broadly cross-reactive, PorA-independent bactericidal response against a variety of heterogeneous serogroup B strains [5–7]. This lipoprotein, called LP2086 by Fletcher et al. [7], was also identifie by Masignani et al. [8], who referred to it as GNA1870. It has recently been shown to bind complement factor H [9], with specificit for factor H from humans or higher primates [10]. Accordingly, the protein has been renamed fHBP. Recruitment of factor H to the surface of the bacterium, thereby inhibiting the alternative complement pathway, may be an important survival mechanism for pathogenic *Neisseria* [11]. Individuals carrying a genetic polymorphism in a presumed regulatory region for factor H have been reported to have elevated serum levels of factor H and an increased risk for meningococcal disease [12], and strains in which *fhbp* has been deleted are more susceptible to complement-mediated killing [13]. Thus, in addition to its ability to elicit a bactericidal response in humans [14], fHBP plays an important role in virulence for this organism.

Early analysis of fHBP sequences from 63 strains that represented a variety of *N. meningitidis* lineages identif ed 39 unique protein variants (differing by at least 1 amino acid) that we classifie into 2 distinct groups, subfamilies A and B, with >83% amino acid identity within a subfamily but only 60%–75% amino acid identity between the subfamilies [7]. Bactericidal antibody responses were found to be mainly specifi to each fHBP subfamily [7]. In a similar analysis, Masignani et al. [8] observed 24 unique variants among 78 strains and classifie fHBP into 3 groups of variants: group 1 (corresponding to subfamily B) and groups 2 and 3 (together corresponding to subfamily A).

Members of the fHBP family are currently being evaluated clinically in 2 separate MnB candidate vaccines, 1 of which contains both a subfamily A and subfamily B protein, and the other of which contains only a subfamily B protein (in combination with other antigens). Measuring vaccine performance and demonstrating the sensitivity of disease-causing strains to vaccine-induced antibody is more complex for protein-based vaccines than for capsule-based vaccines. To lay the groundwork to address these issues, we have undertaken a comprehensive survey to examine the distribution and diversity of fHBP in an epidemiologically relevant, systematic collection of recent MnB clinical isolates.

## MATERIALS AND METHODS

***Strains.*** A set of 1263 invasive MnB strains was obtained from the public health laboratories of the United States, Norway, France, the Czech Republic, and from the Health Protection Agency in Manchester (HPA), which covers England, Wales, and Northern Ireland (table 1). For the United States, strains were from the Active Bacterial Core Surveillance sites [16], which collectively cover ~13% of the US population. Available isolates for the period 2000–2005 (>80% of reported cases for those years) were included in the study. For the European countries, strains were collected in a systematic way by order of date received at the reference laboratory. Every seventh (Czech Republic) or every eighth (HPA, France, and Norway) isolate from 2001–2006 was included. The isolates from Europe that were included in this collection thus represent ~13% of invasive MnB isolates from the respective reference laboratories for the period covered by this study. The starting collections from these reference laboratories are estimated to cover 80%–85% of all invasive MnB isolates in France [17] and 50%–70% of those in the Czech Republic. For the HPA collection and Norway, coverage is estimated to be 95%. A total of 9 strains from all sources were not viable upon receipt at Wyeth and were excluded from the collection.

We also sequenced *fhbp* from 574 additional invasive MnB strains from various collections and years, for a total of 1837 strains (table 1). These additional strains were mainly from the United States, HPA, Norway, Czech Republic, New Zealand, and South Africa. They were included in our analysis of sequence diversity but were not included in analyses of variant frequency unless otherwise noted (eg, isolates from South Africa, for which all available isolates for 2005 were evaluated).

***Polymerase chain reaction (PCR) and sequencing.*** The strategy for sequencing *fhbp* employed an initial PCR reaction with primers designed to recognize conserved regions upstream of and partially overlapping the leader peptide of *fhbp* and within 45 nucleotides downstream of the termination codon (5′-CTATTCTGCGTATGACTAGGAG-3′ and 5′-GTCCGAACGGTAAATTATCGTG-3′, respectively). PCR templates were prepared by boiling several *N. meningitidis* colonies from chocolate agar plates in 100 $\mu$L of distilled $H_2O$ for 5 min, then diluting 1:4 in distilled $H_2O$. PCR reactions (50 $\mu$L total volume) contained 4 $\mu$L DNA, Premix Taq-EX TAKARA enzyme cocktail (TAKARA Bio USA), and primers (0.4 $\mu$mol/L each). The amplificatio steps were 95°C for 5 min, 33 cycles of 95°C

**Table 1. *Neisseria meningitidis* Serogroup B Strains**

| Collection, country | Source, location | Years collected | Year | | | | | | | Total |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | 2000 | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | |
| Systematically collected strain collection | | | | | | | | | | |
| United States | Centers for Disease Control and Prevention, Atlanta, Georgia | 2000–2005 | 98 | 53 | 53 | 88 | 79 | 61 | 0 | 432 |
| England, Wales, Northern Ireland | Health Protection Agency, Manchester | 2001–2006 | 0 | 116 | 94 | 96 | 77 | 81 | 72 | 536 |
| France | *Neisseria* Unit, Institut Pasteur, Paris | 2001–2006 | 0 | 36 | 38 | 43 | 40 | 44 | 43 | 244 |
| Norway | Norwegian Institute of Public Health, Oslo | 2001–2006 | 0 | 5 | 4 | 4 | 3 | 3 | 4 | 23 |
| Czech Republic | National Institute of Public Health, Prague | 2001–2006 | 0 | 5 | 5 | 4 | 4 | 6 | 4 | 28 |
| Total | … | … | 98 | 215 | 194 | 235 | 203 | 195 | 123 | 1263 |
| Additional strains[a] | | | | | | | | | | |
| United States | Centers for Disease Prevention and Control, Atlanta, Georgia | 1996–1999 (mainly); 2006 | … | … | … | … | … | … | … | 110 |
| England, Wales, Northern Island | Health Protection Agency, Manchester | 1997–2000 (mainly) | … | … | … | … | … | … | … | 53 |
| Czech Republic | National Institute of Public Health, Prague | 1984–2007[b] | … | … | … | … | … | … | … | 177 |
| Norway | Norwegian Institute of Public Health, Oslo | 1996–2006[c] | … | … | … | … | … | … | … | 146 |
| New Zealand | Institute of Environmental Science and Research, Porirua | 1992–2006 | … | … | … | … | … | … | … | 26 |
| South Africa | National Institute for Communicable Diseases, Johannesburg | 2005[d] | … | … | … | … | … | … | … | 54 |
| Other | Chile (1 strain); Netherlands (5 strains); France (2 strains) | … | … | … | … | … | … | … | … | 8 |
| Total | … | … | … | … | … | … | … | … | … | 574 |

[a] Includes 57 strains sequenced by Fletcher et al. [7].
[b] Includes strains from 2001–2006 not included in the 1263 strain set and strains from 1984–2000 and 2007.
[c] Includes additional strains from 2001–2006 not included in the 1263 strain set
[d] These strains represent all available serogroup B strains collected in 2005 [15].

for 50 s, 59°C for 50 s, and 72°C for 50 s, followed by an extension step at 72°C for 7 min. Amplifie DNA was purif ed using AMPure magnetic beads (Agencourt) and resuspended in 80 μL 10 mmol/L Tris-acetate buffer (pH 8.0).

Sequencing primers were slightly internal to the initial PCR primers and were hybridized to conserved regions in the N-terminus of *fhbp* or to subfamily-specifi C-terminal regions. Conserved N-terminal and internal primers were 5′-TATGACT-AGGAGCAAACCTG-3′ and 5′-AGCTCATTACCTTGGAGAG-CGGA-3′, respectively. Subfamily-specifi C-terminal and internal primers were 5′-TACTGTTTGCCGGCGATG-3′ and 5′-GAATGCTTTGCCGTGATACTCGGCT-3′, respectively, for subfamily A and were 5′-TTCGGACGGCATTTTCACAATGG-3′ and 5′-GGCGATTTCAAATGTTCGATTT-3′, respectively, for subfamily B. All primer reactions were run for all strains; only those corresponding to the correct subfamily were successful. Sequencing reactions contained 4 μL PCR product, 2 μL 5× buffer (ABI PRISM BigDye Terminator V.3.1 Cycle Sequencing Ready Reaction Kit), 4 μL ABI BDT_V3 polymerase, and 1 μL primer (3.2 pmol/μL) in a reaction volume of 20 μL.

Plates were handled on a Rapid Plate 96 channel Qiagen Bio Robot and with a Multi-probe II 4 tip system. Reactions were heated to 96°C for 30 s and then cycled at 96°C for 10 s, 50°C for 5 s, and 60°C for 4 min for 25 cycles. Electrophoresis was performed on an ABI3730 DNA sequencer (Applied Biosystems). Sequences were examined to verify double-stranded coverage with Sequencher 4.0, and sequences were subsequently trimmed to match the length of the mature protein (lipid attachment cysteine to termination codon). *fhbp* variant numbers were assigned sequentially to new sequences with use of a nomenclature that indicated subfamily, protein, and nucleotide variant (eg, A22_003 indicates the third nucleotide sequence variant of subfamily A protein variant number 22). This nomenclature replaces that used in our other publications [7]. Initially, all strains were sequenced twice from independent PCR reactions. After sequencing ~100 strains with no discrepancies between duplicates, only those strains containing new *fhbp* sequences were confi med by repeat PCR and sequencing. Sequences were deposited in GenBank with accession numbers FJ184079-FJ184274. Multilocus sequence typing (MLST) was performed according to the protocols at the *Neisseria* MLST Web site (http://pubmlst.org/neisseria/) [18].

***Bioinformatic analyses.*** Protein sequences were aligned with ClustalW 1.83 [19]. Distances were calculated by the neighbor-joining method [20] in ClustalW, with correction for multiple substitutions [21]. Consensus sequences were calculated using the EMBOSS utility "con" [22]. Each variant had equal weight in calculation of the consensus, with no adjustment for variant frequency. Trees were displayed using MEGA software, version 4.0 [23]. Independent subfamily A and subfamily B alignments were generated by separating the alignment of all sequences, without realignment or gap removal, to preserve a common numbering system. Networks were generated using Splitstree, version 4.0 [24], with default parameters.

For construction of minimum spanning trees, the aligned unique protein sequences were reduced to the set of 77 (subfamily A) or 101 (subfamily B) variable positions, and the remaining data were converted to a numerical matrix. Generation of minimum spanning trees and construction of clusters of fHBP sequences were performed with Bionumerics, version 5.1 (Applied Maths), with use of a categorical coeffi ient. Linkage priorities were assigned as follows: (1) maximum number of single-locus variants, (2) maximum number of single- and double-locus variants, and (3) maximum number of sequences belonging to a single type, with hypothetical types being permitted.

## RESULTS

***Summary of* fhbp *sequencing.*** *fhbp* was detected in all 1837 strains examined. All strains encoded a full-length protein except for 1 (0030/01), which carried a premature stop codon at nucleotide 366. There were 218 unique nucleotide sequences encoding 173 unique protein variants of mature fHBP. Two variants carried a substitution in the termination codon that resulted in a 3 amino acid extension at the C-terminus (B82 and B112). Phylogenetic analysis indicated that all genes fell into 1 of 2 groups (figu e 1*A*), which had previously been named subfamilies A and B [7]. There were 74 unique fHBP subfamily A and 99 unique subfamily B protein sequences, of which 41 and 67, respectively, were represented by just 1 strain. Pairwise identities within a subfamily ranged from 83% to 99%; between subfamilies, pairwise identities ranged from 60% to 75%. A consensus alignment is shown in figu e 2. Evaluation of individual residue variation revealed that 111 amino acids (44% of all residues) were invariant among all proteins of both subfamilies (coded red in figu e 2). Within subfamily B, 166 residues were invariant; subfamily A is somewhat less diverse, with 192 positions completely conserved. At 32 positions, the residues were subfamily-defining that is, the residues were identical within and characteristic for either subfamily A or B (coded yellow in fig re 2).

Despite the clear existence of 2 subfamilies, 1 variant, A62 (found in 2 genetically and geographically unrelated strains, 0167/03 [Czech Republic, ST-5932] and 21626 [France, ST-2688]), was an apparent result of recombination between subfamilies. The N-terminus of A62_001 was 100% identical to B09 from nucleotides 1–556, and its C-terminus was 100% identical to A22 from nucleotides 507–768. The 2 A62 genes differed by 2 nucleotides. Most of the diversity between the 2 families is found in the C-terminal domain, and therefore, the recombinant variant was assigned to subfamily A. Such inter-subfamily recombination events evidently occur very rarely, be-

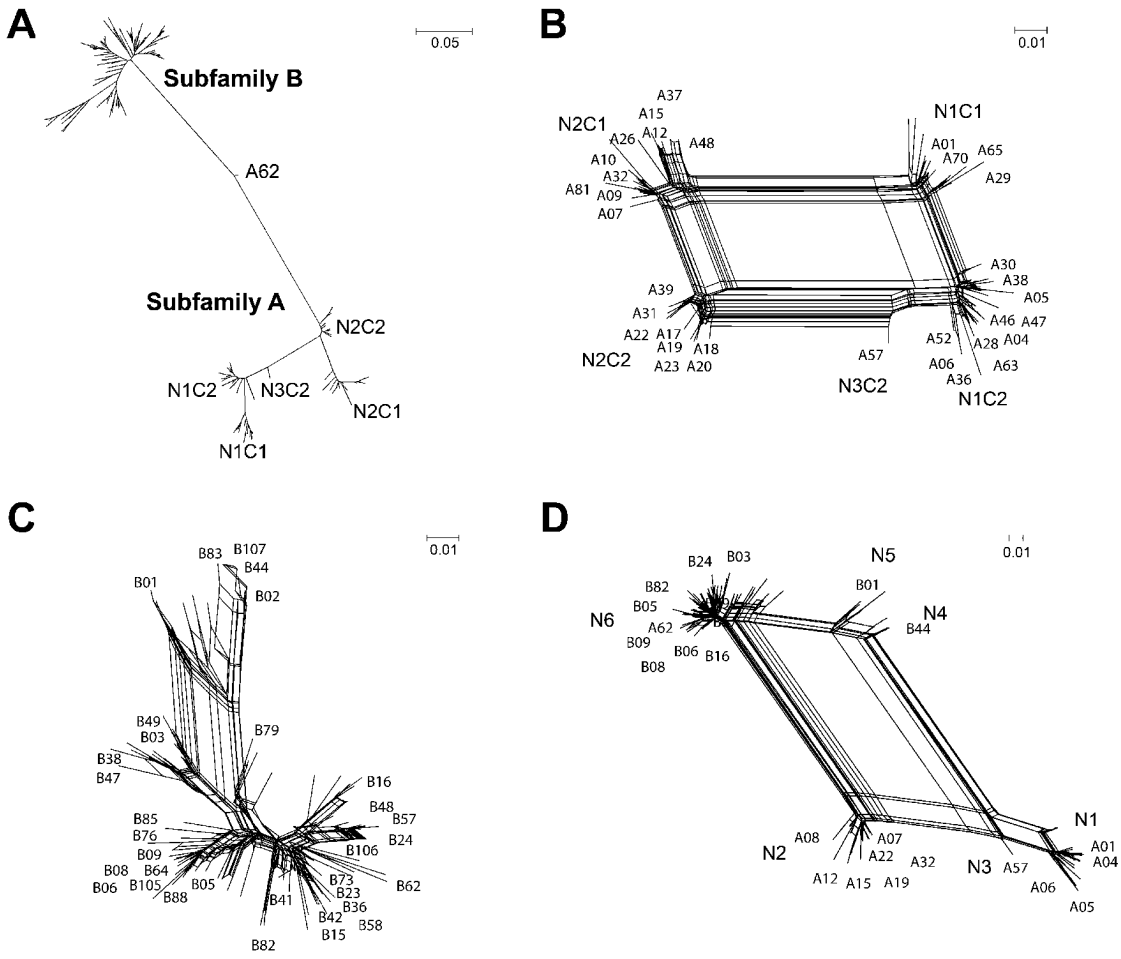


**Figure 1.** *A,* Neighbor-joining tree generated based on the ClustalW alignment of 172 unique fHBP protein sequences. The truncated B110 sequence was excluded from the alignment. The tree was bootstrapped 500 times and drawn using MEGA software. *B,* SplitsTree analysis of subfamily A fHBP variants. *C,* SplitsTree analysis of subfamily B variants. The trees in panels *B* and *C* are based on full-length alignments of subfamily A and subfamily B variants, respectively. *D,* SplitsTree analysis of the N-terminal domain (residues 1–139 [26]) of all subfamily A and B variants. Representative variants are labeled. Subdomain assignments are described in the Results. In the nomenclature of Masignani et al. [8], variant 2 is equivalent to subfamily A variants carrying the N2 domain (N2C1 and N2C2); variant 3 is equivalent to subfamily A variants carrying the N1 domain (N1C1 and N1C2); and variant 1 corresponds to subfamily B variants.

cause only these 2 recombinants were identifie among >1800 strains.

***Network analysis of fHBP sequences.*** Phylogenetic network analysis provides a better description of sequence relationships than standard tree representations when, as in the case of *N. meningitidis,* evolution most likely proceeds via horizontal transfer and recombination rather than linear, branched speciation events [25]. Network analysis using SplitsTree, version 4.0, subdivided subfamily A into 4 groups with multiple and equally likely paths among them (figu e 1*B*). Visual examination of the aligned sequences in each group further revealed the existence of 2 N-terminal domain types (N1 and N2) and 2 C-terminal domain types (C1 and C2), which have apparently recombined in all 4 possible combinations. The 2 groups with the N1 domain, N1C1 and N1C2, are together equivalent to variant 3 in Masignani et al. [8], whereas the N2-containing variants, N2C1 and N2C2, together correspond to variant 2 in Masignani et al. [8]. The average pairwise distance of all N1-containing variants versus those with the N2 domain was 0.15266, whereas the distance between variants carrying C1 (ie, N1C1 and N2C1) versus those with C2 (N1C2 and N2C2) was 0.11529, indicating that, within subfamily A, more diversity is contributed by the N-terminal domains than by the C-terminal domains. However, the genetic distance between either of the N1- or N2-containing subfamily A groups and subfamily B (0. 47303 and 0.39271, respectively) was 2.5–3 times greater than the distance between the 2 subgroups of subfamily A, reaffi ming the essentially bifurcated nature of the fHBP phylogenetic tree (fi ure 1*A*).

N1 and N2 were distinguished by 20 "signature" residues

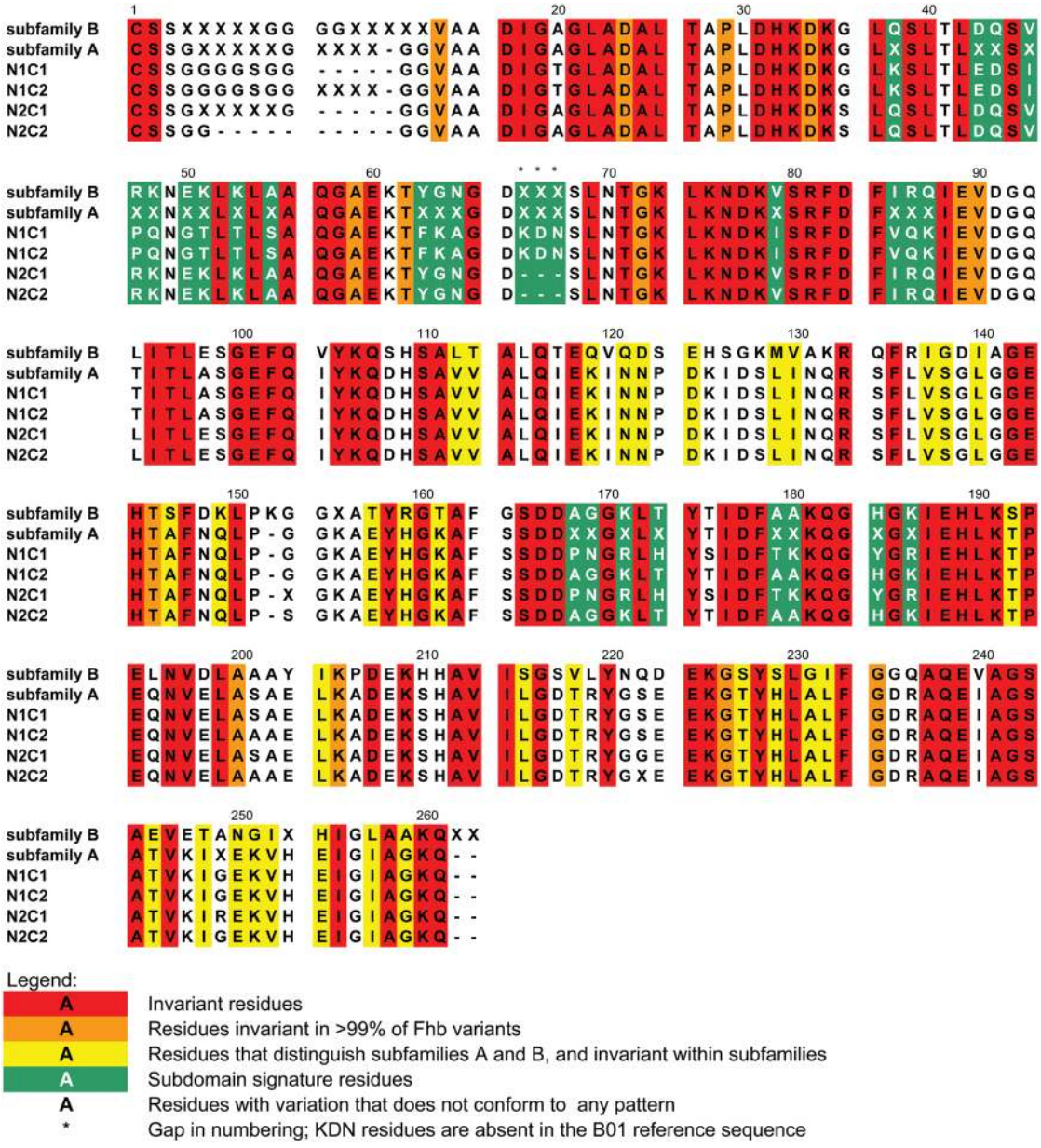


**Figure 2.** fHBP consensus alignment and evaluation of residue conservation. Alignment of the consensus sequences of 97 subfamily B variants (truncated variant B110 excluded), 73 subfamily A variants (hybrid A62 excluded), and each of the 4 groups of subfamily A variants N1C1 ($n = 14$), N1C2 ($n = 23$), N2C1 ($n = 21$), and N2C2 ($n = 14$). A57, which carried an N3 domain, was included only in the complete subfamily A alignment. Residues are numbered according to the full-length structure of the B01 variant [26]; asterisks (*) indicate the position of the KDN residues in some variants, which are absent in the reference sequence B01. Residues are color-coded according to the type of variation exhibited at each position. Residues conserved between both subfamilies are colored in red and are 100% identical among all variants; orange indicates residues that differ in just 1 variant. Subfamily-defining residues (ie, residues that differ between subfamily A and subfamily B but are invariant within each subfamily) are shown in yellow. Subdomain signature residues (shown in green) are those that define the 4 subgroups of subfamily A (figure 1*B*); these are noted as "X" in the consensus line for all subfamily A variants. Variable residues (black on white) are those that varied in more than a single variant without conforming to any pattern. Many of these are conserved in the majority of sequences, and thus display identical consensus residues across all 6 groups. "X" indicates that there is no consensus residue. Only 1 position in the alignment (180) contained as many as 5 different residues, all other positions were limited to a maximum of 4.

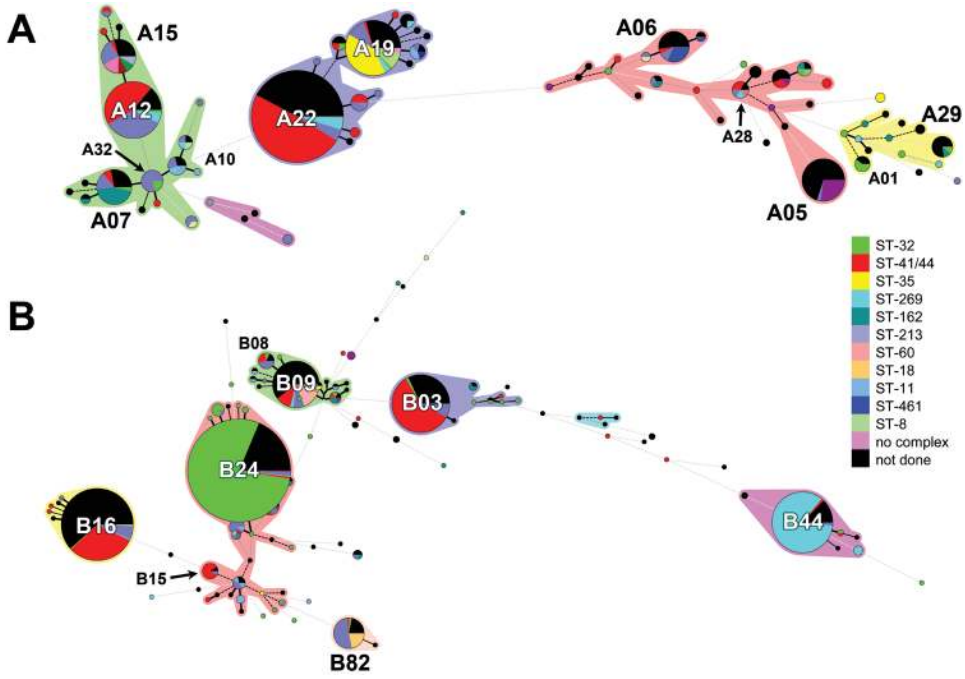


**Figure 3.** Minimum spanning trees generated based on variable amino acid positions in the protein alignment of fHBP variants in 1317 strains (systematically collected strains and strains from South Africa). Trees were generated with Bionumerics with use of a categorical coefficient and are based on 77 variable positions in subfamily A *(A)* and 101 variable positions in subfamily B *(B)*. The size of each circle is proportional to the number of strains carrying that variant. Each color in the circles represents a different MLST clonal complex. Groups based on fHBP protein sequence (connected by colored background) were formed where the neighboring distance was ≤5 amino acid differences, with formation of hypothetical types allowed. Connecting lines are proportional to distance.

(marked in green in figu e 2), including the KDN insertion after residue 67, that were 100% diagnostic for either N1 or N2. A third form of the N terminal domain, N3 (found in only 1 strain [variant A57]), had the signature residues of N1 except for the 3 most C-terminal signature residues, where it carried the N2 signature Ile85, Arg86 and Gln87. In the C-terminal domain, 8 residues were 100% diagnostic for either C1 or C2 in subfamily A (fi ure 2).

Network analysis of subfamily B did not yield obvious subgroups similar to that of subfamily A (figu e 1*C*). However, consideration of just the N-terminal domains (residues 1–139 [26]) of all variants of both subfamilies (figu e 1*D*) indicated a close relationship between subfamily A and B variants. The majority of subfamily B variants (84% of subfamily B strains in the systematically collected set) possessed an N-terminal domain (N6) with all signature residues of the A subfamily N2 domain. Variation at the nonsignature residues resulted in the independent branching of N2 and N6. This analysis also revealed 7 subfamily B variants with N-terminal domains that, like N3, appeared to be a result of recombination between an N1- and N2-like domain between residues 79 and 85 (N4, exemplifie by B44), as well as 4 subfamily B variants that could have arisen by recombination of an N1- and N2-like domain between residues 55 and 63 (N5).

Minimum spanning trees constructed from the variable positions in the subfamily A and subfamily B protein alignments are shown in figu e 3. Complexes were allowed to form from neighboring sequences containing up to 5 amino acid differences. In both trees, it is evident that a few "founder" sequences and closely related variants (with 1–2 amino acid differences) make up the majority of the fHBP population. In subfamily B, 31 variants (39 strains) were not included within any complex; 26 of these occurred in just 1 strain, and 5 others occurred in ≤4 strains. In subfamily A, there were just 9 variants (11 strains) that were not included within any complex.

***Geographic distribution and frequency of fHBP variants.*** Overall, 71% of the systematically collected strains and 70% of the larger strain set were subfamily B variants, the remainder being subfamily A variants. Within each individual country, the percentage of strains carrying subfamily A variants ranged from 23% to 35%, and the percentage carrying subfamily B variants ranged from 65% to 77% (table 2). The distribution of strains in the United States is affected by an ongoing epidemic in Oregon that is not representative of disease elsewhere in the country. If Oregon is excluded from the analysis of strains from the United States, the subfamily distribution for the other US sites is 45% subfamily A and 55% subfamily B. In South Africa (where all available isolates from 2005 were evaluated), sub-

**Table 2. Frequency of the 10 Most Common fHBP Variants in the Systematically Collected Strain Set, by Geographic Location**

| Country | No. of strains | Percentage in subfamily | | Common variants, percentage of strains | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | A | B | B24 | B16 | B44 | B03 | B09 | A22 | A19 | A12 | A05 | A07 |
| Czech Republic | 28 | 25.0 | 75.0 | 17.9 | 3.6 | 10.7 | 7.1 | 0 | 3.6 | 0 | 10.7 | 3.6 | 0 |
| France | 244 | 32.4 | 67.6 | 19.7 | 6.6 | 5.7 | 22.5 | 1.6 | 12.7 | 3.7 | 2.5 | 1.2 | 1.2 |
| Norway | 23 | 34.8 | 65.2 | 34.8 | 8.7 | 4.3 | 17.4 | 0 | 30.4 | 0 | 0 | 4.3 | 0 |
| England, Wales, Northern Ireland | 536 | 23.3 | 76.7 | 3.9 | 23.5 | 22.0 | 6.7 | 9.9 | 7.8 | 1.7 | 0.4 | 4.3 | 0.9 |
| United States | | | | | | | | | | | | | |
| Including Oregon[a] | 432 | 34.5 | 65.5 | 42.6 | 5.1 | 0.2 | 3.2 | 3.9 | 10.4 | 3.5 | 6.3 | 0.2 | 3.0 |
| Not including Oregon[a] | 291 | 45.0 | 55.0 | 29.2 | 6.5 | 0.3 | 4.8 | 4.5 | 12.7 | 4.5 | 8.6 | 0.3 | 4.5 |
| Total | 1263 | 29.1 | 70.9 | 21.1 | 13.2 | 10.8 | 8.8 | 5.8 | 10.0 | 2.6 | 3.0 | 2.3 | 1.7 |
| South Africa | 54 | 57.4 | 42.6 | 11.1 | 14.8 | 1.9 | 3.7 | 3.7 | 13.0 | 13.0 | 0 | 0 | 0 |

[a] Statistics for the United States are impacted by the epidemic in Oregon [27].

family A strains were isolated more frequently than were subfamily B strains (31 [57.4%] of 54 strains were subfamily A), including a high frequency of a variant (A32) that was rare elsewhere.

Within the systematically collected strain pool, 143 fHBP variants were identifie among the 1263 strains; 92 variants were found just once, and 20 variants were found in 2 strains each. Only 18 variants were found in ⩾10 isolates; of these, 10 variants (5 each of subfamily A and subfamily B) accounted for 79% of the systematically collected strain set (table 2). Frequencies for all variants are given in table 3.

Although the 10 most common variants were almost always among the most common in each country, their rank order differed (table 2). In the United States, B24 accounted for 42.6% of the isolates from 2000–2005. B24 is associated with the ST-32 clonal complex, which has been responsible for long-term hyperendemic disease in parts of the United States (particularly in Oregon) [27] and many parts of Europe [17, 28]. Strains carrying the B24 variant were common in Norway (35%) and France (20%) but were less frequent in the HPA set (4%), where the common subfamily B variants were B16 (23.5%) and B44 (22.0%). B44 was isolated less frequently in France, Norway, and the Czech Republic and was isolated only once in the United States. No major changes in variant frequency within countries were observed over the 6 years surveyed (f gure 4).

***Relationship of MnB genetic lineage and fHBP sequence type.*** Data for common MLST clonal complexes from the systematically collected strains from the United States, France, Norway, and the Czech Republic are shown in figu e 5. Most clonal complexes contained both subfamily A and subfamily B variants. For example, the ST-41/44 clonal complex is highly diverse and contains a wide diversity of fHBP variants, including 107 subfamily A strains (13 different variants) and 129 subfamily B strains (17 variants). Two subfamily A (A12 and A22) and 2 subfamily B variants (B03 and B16) accounted for 39% and 42%, respectively, of ST-41/44 strains. The major exception was the ST-32 clonal complex, where 234 (87%) of 270 strains carried the B24 variant. However, despite the high frequency of B24 within this lineage, 27 other fHBP variants representing both subfamilies (10 subfamily A and 17 subfamily B) occurred in the remaining 36 strains. Analysis of PorA subtyping data for 500 strains revealed a similar diversity of fHBP variants within most PorA types (data not shown).

## DISCUSSION

Our molecular subtyping analysis confi ms that there are 2 phylogenetic groups of fHBP sequences, subfamilies A and B. Sequence variation is distributed throughout the length of the proteins and is interspersed with highly conserved residues; there are no hypervariable regions characteristic of other surface antigens, such as PorA. The pattern of natural sequence variation of fHBP suggests that evolution is operating under different constraints within each subfamily. Variation in both the N- and C-terminal domains of subfamily A is relatively limited, whereas subfamily B shows more evidence of recombination in the N-terminal domain and more allowable mutation overall.

The division of fHBP variants into 2 sequence families parallels the functional immune reactivity of these proteins. The 111 residues common to and conserved between both subfamilies map mainly to the interior core structure of fHBP, whereas 32 subfamily-specifi residues lie mainly on one surface of the structure [26]. These residues are likely to be largely responsible for the subfamily-specifi immune response. Because monovalent fHBP vaccines elicit bactericidal antibodies that are

**Table 3. Domain Assignments and Frequency for All *fhbp* Variants**

This table is available in its entirety in the online version of the *Journal of Infectious Diseases*

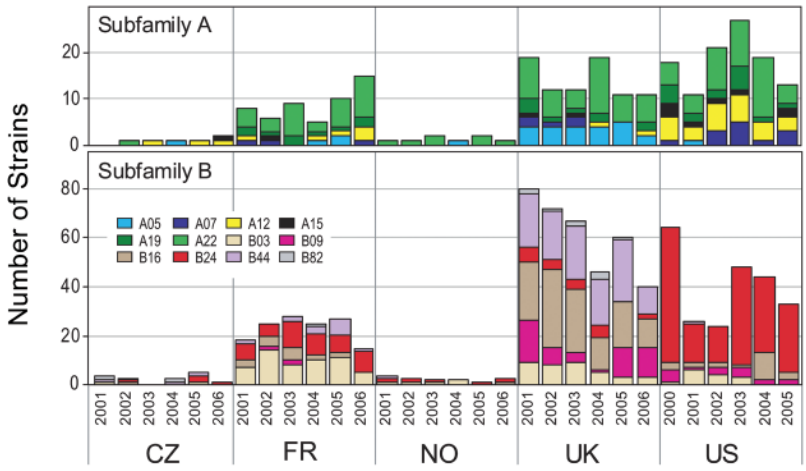


**Figure 4.** fHBP variants by country and by year in the systematically collected strain set. The 6 most common variants for each subfamily are included. CZ, Czech Republic; FR, France; NO, Norway; UK, United Kingdom; US, United States.

largely subfamily specifi [7, 29, 30], a bivalent vaccine is necessary to provide coverage against both subfamilies. A bivalent vaccine given to mice, rabbits, monkeys, or humans is capable of generating serum bactericidal antibody responses against a variety of heterologous strains [5–7, 14, 31].

In assessing the likelihood that a surface protein–based vaccine will recognize a majority of clinical isolates, it is imperative to monitor the variation and expression of each individual antigen rather than relying on epidemiological surrogates. In a few instances, a particular fHBP variant may predict MLST (eg, 97% of B24 strains belong to the ST-32 clonal complex), but the opposite is not necessarily true (figu e 5). Earlier, less comprehensive studies [32, 33] showed that neither PorA nor MLST is predictive of fHBP variant. Similarly, this study showed that even those complexes, such as ST-32, that contain a dominant variant also contain examples of many other fHBP variants of both subfamilies.

A number of *N. meningitidis* surface protein–based vaccines have been considered for development, but the hurdles have been considerable. These include a failure to generate bactericidal antibodies in humans (Neisserial surface protein A [NspA]) [34], a high degree of sequence divergence in critical epitopes (PorA and Tbp) [35–37], and an absence of the gene in many isolates (*N. meningitidis* Adhesin A [NadA]) [32, 38]. However, fHBP is ubiquitous in *N. meningitidis,* is expressed on the bacterial surface, generates bactericidal antibodies in humans, and sequence divergence within each of the 2 subfamilies does not limit the development of a broad, subfamily-specifi bactericidal response [5–7, 14, 30, 39]. All of these attributes make a bivalent recombinant fHBP vaccine a promising candidate for prevention of MnB disease.

The strains described in this study have also been evaluated for surface expression of fHBP and serogroup B capsule [40]. They show a range of expression for both, with no correlation between amount of capsule and level of surface expression of fHBP. The best predictor for killing by anti-fHBP antibodies appears to be the level of fHBP surface expression [30]. All 1837 strains evaluated in this study contained the fHBP gene, consistent with the important role that fHBP is believed to play in survival of the organism in vivo. An interesting observation was the detection of 3 strains with truncated or hybrid variants of fHBP; additional studies are necessary to determine whether these proteins are functional. Ongoing studies are aimed at

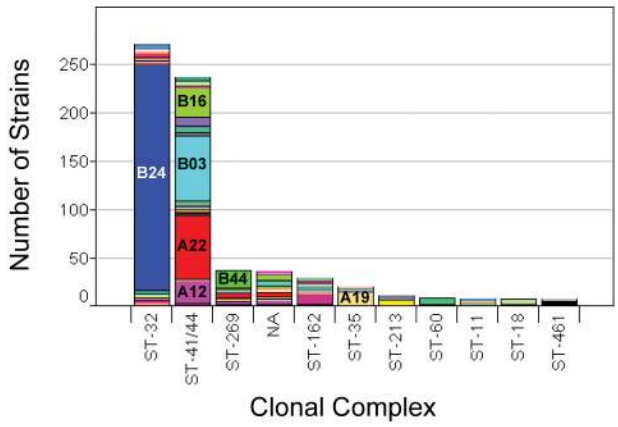


**Figure 5.** Distribution of fHBP variants within clonal complexes. Multilocus sequence type was determined for 694 of the 727 strains from the United States, France, Norway, and Czech Republic from the systematic collection. Data for 672 strains belonging to clonal complexes containing at least 6 strains (461 subfamily B and 211 subfamily A) are included in the figure; 22 strains in less common complexes are not shown. NA indicates sequence types (STs) that were not assignable to a clonal complex. Each color represents a different fHBP variant; the most common variants are labeled.

further refinin the correlates of fHBP surface expression and protection in relation to the selection of appropriate assay strains for the evaluation of Phase III sera. The epidemiologically and genetically characterized collection of invasive MnB strains described here forms the framework for this ongoing work.

## Acknowledgments

We thank the Active Bacterial Core Surveillance sites for their efforts in collecting the isolates from the United States, the Group for Enteric, Respiratory, and Meningeal Disease Surveillance in South Africa (GERMS-SA) for their efforts in collecting the isolates from South Africa, and all the other laboratories responsible for isolating and documenting the primary case isolates. The technical assistance of Scott Sigethy and Diane Pawlyk (Wyeth); Eva Hong, Corinne Ruckly, Dario Giorgini, and Jean-Michel Alonso (Institut Pasteur); Vlasta Pavlikova, Renata Pospisilova, Monika Heroldova and Romana Richterova (NIPH Prague); and Linda de Gouveia and Kedibone Mothibeli (South Africa) is greatly appreciated.

## References

1. Rosenstein NE, Perkins BA, Stephens DS, Popovic T, Hughes JM. Meningococcal disease. N Engl J Med **2001**; 344:1378–88.
2. Trotter CL, Andrews NJ, Kaczmarski EB, Miller E, Ramsay ME. Effectiveness of meningococcal serogroup C conjugate vaccine 4 years after introduction. Lancet **2004**; 364:365–7.
3. Finne J, Bitter-Suermann D, Goridis C, Finne U. An IgG monoclonal antibody to group B meningococci cross-reacts with developmentally regulated polysialic acid units of glycoproteins in neural and extraneural tissues. J Immunol **1987**; 138:4402–7.
4. Finne J, Leinonen M, Makela PH. Antigenic similarities between brain components and bacteria causing meningitis. Implications for vaccine development and pathogenesis. Lancet **1983**; 2:355–7.
5. Farley J, Fletcher L, Bernfiel L, et al. Characterization, cloning and expression of different subfamilies of the ORF 2086 gene from *Neisseria meningitidis*. In: Program and abstracts of the 13th International Pathogenic Neisseria Conference (Oslo, Norway). **2002**.
6. Bernfiel L, Fletcher L, Howell A, et al. Identificatio of a novel vaccine candidate for group B *Neisseria meningitidis*. In: Program and abstracts of the 13th International Pathogenic Neisseria Conference (Oslo, Norway). **2002**.
7. Fletcher LD, Bernf eld L, Barniak V, et al. Vaccine potential of the *Neisseria meningitidis* 2086 lipoprotein. Infect Immun **2004**; 72:2088–100.
8. Masignani V, Comanducci M, Giuliani MM, et al. Vaccination against *Neisseria meningitidis* using three variants of the lipoprotein GNA1870. J Exp Med **2003**; 197:789–99.
9. Madico G, Welsch JA, Lewis LA, et al. The meningococcal vaccine candidate GNA1870 binds the complement regulatory protein factor H and enhances serum resistance. J Immunol **2006**; 177:501–10.
10. Granoff DM, Welsch JA, Ram S. Binding of complement Factor H (fH) to *Neisseria meningitidis* is specifi for human fH and inhibits complement activation by rat and rabbit sera. Infect Immun **2009**; 77: 764–9.
11. Schneider MC, Exley RM, Chan H, et al. Functional signif cance of factor H binding to *Neisseria meningitidis*. J Immunol **2006**; 176:7566–75.
12. Haralambous E, Dolly SO, Hibberd ML, et al. Factor H, a regulator of complement activity, is a major determinant of meningococcal disease susceptibility in UK Caucasian patients. Scand J Infect Dis **2006**;38: 764–71.
13. 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. J Infect Dis **2008**; 197:1053–61.
14. Nissen MD, Marshall HS, Richmond P, et al. A randomized, placebo-controlled, double-blind, phase 1 trial of ascending doses of meningococcal group B rLP2086 vaccine. In: Program and abstracts of the 26th Annual Meeting of the European Society for Paediatric Infectious Diseases (Graz, Austria). **2008**.
15. von Gottberg A, du Plessis M, Cohen C, et al. Emergence of endemic serogroup W135 meningococcal disease associated with a high mortality rate in South Africa. Clin Infect Dis **2008**; 46:377–86.
16. Centers for Disease Control and Prevention. Active Bacterial Core Surveillance. Available at: http://www.cdc.gov/ncidod/dbmd/abcs/meth-surv-pop.htm. Accessed 8 June 2009.
17. Rouaud P, Perrocheau A, Taha MK, et al. Prolonged outbreak of B meningococcal disease in the Seine-Maritime department, France, January 2003 to June 2005. Euro Surveillance **2006**; 11:178–81.
18. Jolley KA, Chan MS, Maiden MC. mlstdbNet—distributed multi-locus sequence typing (MLST) databases. BMC Bioinformatics **2004**; 5:86.
19. Chenna R, Sugawara H, Koike T, et al. Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res **2003**; 31:3497–500.
20. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol **1987**; 4:406–25.
21. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol **1980**; 16:111–20.
22. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet **2000**; 16:276–7.
23. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol **2007**; 24:1596–9.
24. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol **2006**; 23:254–67.
25. Holmes EC, Urwin R and Maiden MC. The influenc of recombination on the population structure and evolution of the human pathogen *Neisseria meningitidis*. Mol Biol Evol **1999**; 16:741–9.
26. Mascioni A, Bentley BE, Camarda R, et al. Structural basis for the immunogenic properties of the meningococcal vaccine candidate LP2086. J Biol Chem **2009**; 284:8738–46.
27. Diermayer M, Hedberg K, Hoesly F, et al. Epidemic serogroup B meningococcal disease in Oregon: the evolving epidemiology of the ET-5 strain. JAMA **1999**; 281:1493–7.
28. Yazdankhah SP, Kriz P, Tzanakaki G, 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**; 42:5146–53.
29. Pillai S, Howell A, Alexander K, et al. Outer membrane protein (OMP) based vaccine for *Neisseria meningitidis* serogroup B. Vaccine **2005**; 23: 2206–9.
30. Jiang H-Q, Harris SL, Tan C, et al. Prediction of broad vaccine coverage for a bivalent rLP2086 based vaccine which elicits serum bactericidal activity agains a diverse collection of serogroup B meningococci. In: Program and abstract of the 16th International Pathogenic Neisseria Conference (Rotterdam, The Netherlands). **2008**.
31. Anderson AS, Xu R, Tan C, et al. Functional cross-reactive antibodies are elicited by a Group B *Neisseria meningitidis* bivalent recombinant lipidated LP2086 vaccine in Cynomolgus macaques [abstract P100]. In: Program and abstract of the 16th International Pathogenic Neisseria Conference (Rotterdam, The Netherlands). **2008**.
32. Beernink PT, Welsch JA, Harrison LH, Leipus A, Kaplan SL, Granoff DM. Prevalence of factor H–binding protein variants and NadA among meningococcal group B isolates from the United States: implications for the development of a multicomponent group B vaccine. J Infect Dis **2007**; 195:1472–9.
33. Jacobsson S, Thulin S, Molling P, et al. Sequence constancies and variations in genes encoding three new meningococcal vaccine candidate antigens. Vaccine **2006**; 24:2161–8.

34. Halperin SA, Langley JM, Smith B, et al. Phase 1 fi st-in-human studies of the reactogenicity and immunogenicity of a recombinant meningococcal NspA vaccine in healthy adults. Vaccine **2007**; 25:450–7.
35. Rokbi B, Mignon M, Caugant DA , Quentin-Millet MJ. Heterogeneity of *tbpB,* the transferrin-binding protein B gene, among serogroup B *Neisseria meningitidis* strains of the ET-5 complex. Clin Diagn Lab Immunol **1997**; 4:522–9.
36. Tondella ML, Popovic T, Rosenstein NE, et al. Distribution of *Neisseria meningitidis* serogroup B serosubtypes and serotypes circulating in the United States. J Clin Microbiol **2000**; 38:3323–8.
37. Sacchi CT, Whitney AM, Popovic T, et al. Diversity and prevalence of PorA types in *Neisseria meningitidis* serogroup B in the United States, 1992–1998. J Infect Dis **2000**; 182:1169–76.
38. Comanducci M, Bambini S, Brunelli B, et al. NadA, a novel vaccine candidate of*Neisseria meningitidis*. J Exp Med **2002**; 195:1445–54.
39. Koeberling O, Giuntini S, Seubert A, Granoff DM. Meningococcal outer membrane vesicle vaccines derived from mutant strains engineered to express Factor H binding proteins from antigenic variant groups 1 and 2. Clin Vaccine Immunol **2009**; 16:156–62.
40. McNeil LK, Murphy E, Zhao X-J, et al. Detection of LP2086 on the cell surface of *Neisseria meningitidis* and its accessibility in the presence of serogroup B capsular polysaccharide. Vaccine **2009**; 27:3417–21.