Variation in the *Neisseria meningitidis* FadL-like protein: an evolutionary model for a relatively low-abundance surface antigen

Daniel Yero,^{1†} Caroline Vipond,^{2†} Yanet Climent,¹ Gretel Sardiñas,³ Ian M. Feavers² and Rolando Pajón⁴

Correspondence

Caroline Vipond

Caroline.Vipond@nibsc.hpa.org.uk

¹Department of Molecular Biology, Division of Biotechnology, Finlay Institute, Havana, Cuba

²Division of Bacteriology, National Institute for Biological Standards and Control, Blanche Lane, South Mimms, Potters Bar, Hertfordshire, UK

³Division of Vaccines, Center for Genetic Engineering and Biotechnology, Havana, Cuba

⁴Center for Immunobiology and Vaccine Development, Children's Hospital Oakland Research Institute, Oakland, CA 94609, USA

The molecular diversity of a novel *Neisseria meningitidis* antigen, encoded by the ORF NMB0088 of MC58 (FadL-like protein), was assessed in a panel of 64 diverse meningococcal strains. The panel consisted of strains belonging to different serogroups, serotypes, serosubtypes and MLST sequence types, of different clinical sources, years and countries of isolation. Based on the sequence variability of the protein, the FadL-like protein has been divided into four variant groups in this species. Antigen variants were associated with specific serogroups and MLST clonal complexes. Maximum-likelihood analyses were used to determine the relationships among sequences and to compare the selection pressures acting on the encoded protein. Furthermore, a model of population genetics and molecular evolution was used to detect natural selection in DNA sequences using the non-synonymous : synonymous substitution ($d_N : d_S$) ratio. The meningococcal sequences were also compared with those of the related surface protein in non-pathogenic commensal *Neisseria* species to investigate potential horizontal gene transfer. The *N. meningitidis* *fadL* gene was subject to only weak positive selection pressure and was less diverse than meningococcal major outer-membrane proteins. The majority of the variability in *fadL* was due to recombination among existing alleles from the same or related species that resulted in a discrete mosaic structure in the meningococcal population. In general, the population structuring observed based on the FadL-like membrane protein indicates that it is under intermediate immune selection. However, the emergence of a new subvariant within the hyperinvasive lineages demonstrates the phenotypic adaptability of *N. meningitidis*, probably in response to selective pressure.

Received 1 July 2010

Revised 5 August 2010

Accepted 30 August 2010

INTRODUCTION

Neisseria meningitidis continues to be a major cause of meningitis and septicaemia worldwide, and meningococcal

disease is one of the most aggressive bacterial infectious diseases in humans. Of the 13 distinct *N. meningitidis* serogroups only five are associated with invasive disease; of these, serogroups A, B and C account for approximately 90 % of all cases (Rosenstein *et al.*, 2001). The disease continues to cause substantial morbidity and mortality in children worldwide despite the availability of effective antibiotics. Although early diagnosis and antibiotic treatment enhance survival, prevention through vaccination constitutes the most effective means to control meningococcal disease (Jódar *et al.*, 2002). Currently, there are successful glyconjugate vaccines against four (A, C, Y and W-135) of the five pathogenic serogroups (Mitka, 2005). However, prevention of serogroup B meningococcal disease represents a particularly difficult challenge in vaccine development. The use of capsular polysaccharide

†These authors contributed equally to this work.

‡Present address: Department of Biochemistry, Faculty of Biology, University of Havana. PC 10400, Cuba.

Abbreviations: CREE, Correia repeat-enclosed element; GARD, genetic algorithm for recombination detection; MLST, multilocus sequence type; VR, variable region.

The GenBank/EMBL/DDBJ accession numbers for the *fadL* sequences of the new strains of *N. meningitidis* examined are GU249513–GU249548.

A supplementary table, listing meningococcal strains used in this study, is available with the online version of this paper.

as the basis of a vaccine for prevention of meningococcal diseases has been problematic, since the serogroup B capsular polysaccharide is identical to a widely distributed human carbohydrate, making it poorly immunogenic in humans and raising the possibility that it might elicit an auto-antibody response (Finne *et al.*, 1987; Häyrynen *et al.*, 1995). One potential vaccine strategy in the fight against meningococcal disease involves the exploitation of surface proteins of *N. meningitidis* either contained in outer-membrane vesicles (Lewis *et al.*, 2009) or as recombinant proteins (Feavers & Pizza, 2009).

High levels of antigenic diversity among the immunogenic proteins of the meningococcus have hindered the development of a broadly cross-protective vaccine. However, epidemiological data have revealed that meningococcal populations are structured by the immunoselective pressure of the host response acting on variable cell-surface antigens (Gupta *et al.*, 1996; Gupta & Maiden, 2001). As a result hypervirulent strains have discrete antigenic phenotypes, which is consistent with the observed non-overlapping association of variants of major antigens such as PorA (Peeters *et al.*, 1999), PorB, FetA (Urwin *et al.*, 2004) and Opa (Callaghan *et al.*, 2008). This has implications for vaccine design: first, the sequence diversity of a meningococcal antigen provides evidence that it is indeed a target of the immune response and, second, it shows that a vaccine formulation containing as little as five protein variants has the potential to offer protection against the majority of hypervirulent lineages. An example of such a vaccine is the developmental NonaMen, which encompasses nine PorA protein types representing the major hypervirulent strains in circulation and should offer protection against the vast majority of invasive disease (van den Dobbelen *et al.*, 2007).

During the last decade, the availability of meningococcal genome sequence data (Tettelin *et al.*, 2000) has led to the identification of a number of less abundant (minor) cell-surface proteins with vaccine potential (Pajón *et al.*, 2009; Pizza *et al.*, 2000). The FadL-like protein encoded by the NMB0088 ORF in the MC58 genome is an example and has been identified in meningococcal outer-membrane vesicles prepared from different strains (Uli *et al.*, 2006; Vaughan *et al.*, 2006; Vipond *et al.*, 2006; Williams *et al.*, 2007). In experimental immunizations this protein induced the production of antibodies that were both bactericidal and protective in mice (Sardiñas *et al.*, 2009). Based on sequence similarity, it belongs to FadL family of proteins, which are responsible for the transport of hydrophobic molecules in some bacteria (van den Berg, 2005). A structural model of the protein, predicted using FadL from *Escherichia coli* as a template, has a 14-stranded β -barrel occluded by a central hatch domain (Sardiñas *et al.*, 2009). The protein has seven exposed loops, three of which show variability, the variable regions (VRs). Its function remains unknown.

In the present study, *fadL* and the amino acid sequence it encodes were examined for their diversity across an

expanded collection of isolates representing a substantial spectrum of different strains. The meningococcal sequences were also compared with the related surface protein of commensal *Neisseria* species, to investigate potential horizontal gene transfer. Additionally, phylogenetic analyses were conducted to identify the role of mutation and recombination in the evolutionary processes that diversified the meningococcal FadL-like membrane protein.

METHODS

Sequence data. A total of 64 *N. meningitidis* *fadL* gene sequences were examined: 26 from GenBank, including previously published sequences (Sardiñas *et al.*, 2009), and four from complete genome sequences (Bentley *et al.*, 2007; Parkhill *et al.*, 2000; Peng *et al.*, 2008; Tettelin *et al.*, 2000); a further 34 were determined *de novo*. For sequencing, bacterial DNA was amplified by PCR with *Taq* DNA polymerase (Applied Biosciences). Dye terminator sequencing using the BigDye terminator v3.1 cycling kit (Applied Biosystems) was used to sequence the PCR products generated from multiple primer combinations. Nine primers were used to generate PCR products and to sequence the gene and its upstream region. The sequences of the nine primers and their positions are given in Table 1. Previously unpublished sequences have been submitted to GenBank, with the accession numbers GU249513–GU249548 (excluding GU249534 and GU249536).

Of the 64 meningococcal source strains, 14 were serogroup A, 33 were serogroup B, eight were serogroup C, two were serogroup W-135, one was serogroup Y and one was serogroup Z; five strains were non-serogroupable. The panel includes 11 (17.2%) isolates from healthy carriers and 48 (75.0%) disease-causing meningococci (the source of five strains is unknown). The panel also includes strains isolated in different years originating from 19 countries. Multilocus sequence type (MLST) information was available for the whole set of strains (<http://pubmlst.org/neisseria/>), of which 12 were from the ST-32 clonal complex, 10 were from the ST-269 complex, nine were from the ST-41/44 complex, seven were from the ST-11 complex, six were from the ST-4 complex, four each were from the ST-1 and ST-5 complexes, three were from the ST-103 complex, and two each were from the ST-22 and ST-53 complexes. Five strains were from other clonal complexes and had sequence types ST-23, ST-344, ST-823, ST-3790 and ST-4821. The complete list of strains including their genotypic and phenotypic characteristics is given in Supplementary Table S1 available with the online version of this paper.

The *fadL* homologues from *Neisseria lactamica* strains 020-06 (ST-640) (<http://www.sanger.ac.uk/Projects/>) and Y92-1009 (Vaughan *et al.*, 2006) were used in the analysis. *Neisseria gonorrhoeae* strains FA1090 (GenBank accession no. NC_002946) and NCCP11945 (Chung *et al.*, 2008) were also explored for NMB0088 ORF homologues. *De novo* sequence data were obtained for the commensal bacteria *Neisseria sicca* (strain 02233) and *Neisseria polysaccharea* (NIBSC strain 2277) and included in the analysis.

Analysis of sequence data and statistics. To allow comparisons of the *de novo* *N. meningitidis* *fadL* sequences with the previously published meningococcal sequences (Sardiñas *et al.*, 2009), all were truncated to the same length to exclude the predicted signal peptide so that the amino acid sequences began at the 36th amino acid of the *N. meningitidis* MC58 sequence. Amino acid sequences were aligned using CLUSTALW (<http://www.ebi.ac.uk/clustalw/>) and corrected manually. The aligned amino acid sequences were used to generate nucleotide alignments in the MEGA program, version 3.1 (Kumar *et al.*, 2004). MEGA 3.1 was also used to calculate genetic distances between sequences and variants, and to generate approximate tree topologies

Table 1. Primers used to amplify *fadL* and flanking regions

Primer name	Sequence (5'–3')	Relative position (nt)*	Annealing temp. (°C)
CV0088F	GGCTACCACTTCGGCACAG	+76 to +96	50
CV0088F2	GCGTGTACGTCCCCTTCG	+377 to +394	50
CV0088F3	CACCACCATCACCCCAAC	+1038 to +1056	50
CV0088F4	GAACACAACATCCTGCATTCCG	–419 to –398	50
CV0088F6	CGGAAACCGAAATGACCCC	–11 to +8	48
CV0088R	CCTTTGCTGTCCACATCGTTG	+1331 to +1311	49
CV0088R2	CGCGTTGACCGACTGTGTG	+108 to +90	50
CV0088R3	GTTGACTTTGTATGCGCCG	+354 to +336	46
CV0088R4	CGTTTATTTGAATTTGTAGGTGTATTG	+1404 to +1378	47

*The position of the primer is relative to the reference sequence from MC58 genome (NP_273150). Position +1 defines the first base of the first codon; negative numbers are positions upstream of the *fadL* gene. Position of reverse primers is indicated from 3' to 5'.

using the neighbour-joining algorithm and the Kimura two-parameter model of nucleotide substitution. The percentage sequence identities within and between variable segment types of sequences were determined using CLUSTALW. The numbers of mutation sites were evaluated with DnaSP v5 (Librado & Rozas, 2009) and MEGA 3.1. Mutation rates were evaluated by measuring the nucleotide diversity per site (π) using DnaSP v5.

The program Modelgenerator (Keane *et al.*, 2006) was used to identify the optimal evolutionary model. Akaike information criteria (AIC) and hierarchical likelihood ratio test indicated that the general time-reversible (GTR) model best fit the sequence data. The phylogenetic trees were reconstructed using the maximum-likelihood method implemented in the PhyML program (v3.0 aLRT) (Guindon & Gascuel, 2003). The GTR model of nucleotide substitution (Lanave *et al.*, 1984; Yang, 1994) was used, with values for the nucleotide substitution matrix, the proportion of invariant sites, and the shape parameter (α) of a gamma distribution of rate variation among sites (with four categories) estimated during tree reconstruction (trees and parameter values are available on request). Statistical tests for branch support were performed using the bootstrapping method (1000 replicates). Phylograms were displayed and manually edited with MEGA 3.1.

Multiple $r \times c$ contingency tables of main *FadL* variants (r) versus clonal complexes, serogroups, source of isolation and PorA VR1 and VR2 sequence types (c) were analysed using the statistical software package SPSS, version 11.5.1 for the analysis of categorical data. Cramer's V coefficient was reported to test the null hypothesis of independence and to measure the strength of the observed correlations. Cramer's V values close to 1 indicate stronger association.

Detection of recombination. The data were also analysed using ClonalFrame version 1.1 to estimate recombination events (Didelot & Falush, 2007). Inference was performed in a Bayesian framework; the key assumption of this is that recombination events introduce a constant rate of substitutions to a contiguous region of sequence. In the present study, six independent runs, each with 50 000 burn-in iterations, were used. ClonalFrame computes a number of parameters, including the recombination rate (R) and the ratio of probabilities that a given site is altered through recombination and mutation (r/m). To screen multiple-sequence alignments rapidly for recombination we used the genetic algorithm for recombination detection (GARD) (Kosakovsky Pond *et al.*, 2006) implemented in the datamonkey web server (Kosakovsky Pond & Frost, 2005b). GARD identified break points and recombinant sequences whilst providing statistical support for the inferences.

Split decomposition analysis was used to assess the degree of tree-like structure present in the alleles found in the complete set of isolates. The sequence alignments were converted into NEXUS files using MEGA 3.1, and the split decomposition was performed with SplitsTree 4.10 (Huson & Bryant, 2006). The 'fit' parameter indicated how well the graph represented the distance matrix (uncorrected P distances) generated from the data. For graphical representation of the split networks, we excluded gap sites, constant sites and parsimony-uninformative sites.

Detection of selective pressure acting on the *fadL* gene. The distribution of nucleotide changes within the alleles was analysed with the Synonymous Non-synonymous Analysis Program (SNAP; available at <http://hiv.lanl.gov/>). This program calculated pairwise synonymous and non-synonymous distances according to the Nei–Gojobori method on the aligned meningococcal sequences (Nei & Gojobori, 1986).

The maximum-likelihood method of Nielsen, Yang and co-workers (Nielsen & Yang, 1998; Yang *et al.*, 2000) was used to test for positive selection and infer which amino acid sites were under positive selection. This was achieved using the CODEML program in PAML 4 (Yang, 2007). These likelihood models account for variable selective pressures among sites by assuming that there are different classes of sites in the reading frame with different ω [non-synonymous : synonymous substitution ($d_N : d_S$)] ratios. The analysis consisted of two major steps. The first step used the likelihood ratio test to test for positive selection, that is, for presence of sites with $\omega > 1$. The second major step of the analysis was to identify residues under positive selection when the likelihood ratio test suggests their presence. This was achieved by using the Bayes theorem to calculate the (posterior) probabilities that each site, given the data at that site, is from the different ω classes (Yang *et al.*, 2000, 2005). Sites with a high probability of coming from the class with $\omega > 1$ are likely to be under positive selection. The results obtained from the likelihood methods of Nielsen & Yang (1998) were confirmed using the fixed effects likelihood (FEL) analysis (Kosakovsky Pond & Frost, 2005a) in the datamonkey web server.

The OMEGAMAP program (Wilson & McVean, 2006) was also used to detect selection and recombination. Using a Bayesian method it estimates ω and the recombination rate ρ using a Monte Carlo Markov chain along the gene. Three independent OMEGAMAP runs with 100 000 iterations and a thinning interval of 100 were used to assess convergence. Five analyses were carried out to evaluate ω . The first included all 64 meningococcal sequences, and the subsequent analyses were carried out on the sequences from variants 1–4.

RESULTS

Diversity among neisserial NMB0088 ORF homologues

Nucleotide sequences of 64 meningococcal *fadL* genes were analysed in the present study. They had an overall nucleotide conservation of 96.7% [nucleotide diversity per site value (π)=0.033±0.0018] and 34 distinct alleles were detected. Nucleotide alleles had average *p*-distances varying from 0.001 to 0.064, with a total of 139 polymorphic nucleotide sites (Table 2). Many of the polymorphisms occurred within the three putative surface-exposed loops 3, 4 and 5. Phylogenetic analysis of the *fadL* genes clusters the isolates into four major groups (Fig. 1), which are largely congruent with the four protein variants described previously (Sardiñas *et al.*, 2009). The homologous sequence was also present in the four commensal isolates examined. However, in the gonococcus the homologous sequences contained a frame-shift mutation at nucleotide position 25. Fig. 1 also shows that the DNA sequences from three commensal strains clustered in separate branches with the exception of the *N. sicca* isolate, which clustered with three meningococcal strains. The alignment of *fadL* gene sequences revealed that nucleotide polymorphisms occurred in clusters that were common to more than one isolate, resulting in a mosaic structure with the reading frame composed of sequences found among various distantly related isolates (Fig. 2a).

The regions upstream of the new *fadL* genes described in the present work were also sequenced. In the commensal *N. polysaccharaea*, the gonococcal isolates and one of the meningococcal isolates (strain Z5035, ST-1) there is a deletion of approximately 100 nt approximately 30 bases upstream of the putative start codon. The deleted region contained the previously identified Correia repeat-enclosed element (CREE) (Liu *et al.*, 2002) flanking *fadL*. In *N.*

lactamica Y92-1009 this upstream repetitive element is also absent (Sardiñas *et al.*, 2009).

Diversity of the FadL-like protein

Analysis of the deduced FadL-like protein sequences using MEGA 3.1 divided the protein into four main variants. The subdivision of the variants was based on a cut-off in amino acid sequence similarity of ≤96% as previously defined by Sardiñas *et al.* (2009). The proteins in the strains analysed could be grouped into the same four variants: 35 (54.7%) were classified as variant 1, 7 (10.9%) as variant 2, 12 (18.7%) as variant 3 and 10 (15.7%) as variant 4. Although sequences from strains Z4673 (ST-41), Z4706 (ST-32) and Z5035 (ST-1) shared more than 96% amino acid sequence identity with the other variant 1 sequences, phylogenetic analysis based on the amino acid sequence clustered them into a separate group. Thus these three strains are designated here as subvariant 1a. The tree constructed using amino acid sequences (not shown) was similar in shape to that constructed using the nucleotide sequence. Therefore, the four variants and subvariant 1a were easily identified in the phylogenetic tree represented in Fig. 1.

Analysis of the 64 deduced FadL-like protein sequences indicated clusters of identity within the main variant branches. The greatest degree of sequence conservation was seen within variants 3 and 4, which shared at least 99.8% identity between amino acid sequences (Table 2). Variant 2 proteins also showed a high degree of conservation, with at least 98.1% identity between sequences. Amino acid sequences within variant 1 showed the highest degree of variability, with identities between amino acid sequences of at least 96.5%. When strains Z4673, Z4706 and Z5035 were removed from the analysis, the variant 1 sequences shared at least 98.6% identity. The amino acid sequence diversity was greatest between the four variants, with a minimum of

Table 2. Nucleotide and deduced amino acid sequence diversity among meningococcal *fadL* genes

Parameter	All	Variant 1*	Variant 2	Variant 3	Variant 4
<i>fadL</i> gene diversity					
No. of sequences	64	35	7	12	10
Length of sequence (bp)	1257–1269	1266–1269	1257–1260	1260	1260
No. of alleles	34	21	6	4	3
No. of variable sites	139	75	25	3	2
Mean <i>p</i> -distance	3.3 %	1.0 %	0.7 %	0.04 %	0.06 %
FadL protein diversity					
No. of amino acid sequences	29	17	6	4	2
No. of variable sites	54	31	10	3	1
Mean <i>p</i> -distance	3.7 %	0.9 %	0.9 %	0.1 %	0.05 %
VR1 (loop 3)	32.4 %	0.0 %	0.0 %	0.0 %	0.0 %
VR2 (loop 4)	25.8 %	8.1 %	1.4 %	0.0 %	0.0 %
VR3 (loop 5)	10.4 %	7.8 %	12.3 %	0.0 %	1.7 %

*Subvariant 1a is included within FadL sequence variant 1.

Fig. 1. Maximum-likelihood phylogenetic tree of the *fadL* gene in 64 meningococcal strains and four additional sequences from commensals. The serogroup and the MLST clonal complex, if available, are included in parentheses for each strain (NG, non-groupable). Strains isolated from healthy carriers are indicated with asterisks. The major groupings are indicated, as are branches with 100% bootstrap support.

Fig. 2. (a) Diagram illustrating the variability of VR1 and VR2 coding regions in *fadL* sequences from the four variants, subvariant 1a and the four commensals. Fragments with sequence identity higher than 85% are indicated with the same pattern. The potential breakpoint for recombination (position 765) is indicated by a vertical line. The positions of two sites with high evidence of positive selection are indicated by circles. (b, c) Neighbour-joining trees constructed from the aligned nucleotide sequences encoding loops 3 (VR1) (b) and 4 (VR2) (c) of the FadL protein.

93.7% identity between the amino acid sequences. Orthologous NMB0088 ORFs have been found in neisserial genomes from four commensal strains, with identities ranging from 87.95 to 97.59%.

Sequence comparison of the deduced amino acid sequences from the meningococcal and commensal strains revealed large areas of highly conserved sequence and several well-defined variable regions. After analysing the whole alignment, the variability was confined to three specific segments, previously designated VR1, VR2 and VR3 (Sardiñas *et al.*, 2009). The major differences between the protein variants are confined to these three variable regions (Table 2). Secondary-structure predictions, refined with the alignment analysis and homology modelling using previously crystallized proteins FadL from *Escherichia coli* (van den Berg *et al.*, 2004), TodX from *Pseudomonas putida* F1 and TbuX from *Ralstonia pickettii* PKO1 (Hearn *et al.*, 2008), revealed that all the variable regions correspond to predicted extracellular loop domains. The VR1 (loop 3), VR2 (loop 4) and VR3 (loop 5) are located in the surface-exposed loops 3, 4 and 5 respectively.

There is considerable variability in the VRs between the FadL-like protein variants; however, within a variant the VRs are relatively conserved. Principally VR1 is conserved within each of the four variants. VR2 is conserved within variants 3 and 4, and VR3 is conserved in variant 3 (Table 2). The major intra-variant variability was observed in variants 1 and 2 within VR2 and VR3. The diversity observed in VR1 within variant 1 is primarily due to the three strains that clustered into subvariant 1a (Fig. 1). Some of the variants shared the same VR sequences, e.g. VR1 between variants 2 and 3; VR2 between variants 1 and 2; and VR2 between variant 3 and subvariant 1a (Fig. 2a). These results

demonstrated that some variants have overlapping antigenic VR1/VR2 combinations of the FadL-like protein. Additionally, in some cases, the VR from the commensal strains showed sequence identity with various meningococcal FadL like protein variants (Fig. 2a). This suggests that recombination has occurred between the species.

Relationship between FadL variants, phenotypes and clonal complexes

Each of the four main branches of the dendrogram (Fig. 1) was assigned to a variant, which was tested for association with clonal complexes and serogroups. The association probability test Cramer's *V* index, which measures the strength of correlation, indicates moderate association between FadL variants and serogroups ($V=0.654$), and the strongest association was with the clonal complexes ($V=0.797$).

As shown in Fig. 1 and Supplementary Table S1, the majority of serogroup B strains and all the strains belonging to hypervirulent clonal complexes ST-32, ST-269 and ST-11 were FadL sequence variant 1. The variant 3 branch of the phylogenetic tree exclusively contained serogroup A strains, all with STs within the clonal complexes ST-1, ST-4 and ST-5. The meningococcal strains with the FadL sequence variant 4 belonged to ST-41/44 and ST-22 lineages. The variant 2 branch was the most diverse in terms of the clonal complexes and the strains' phenotypes; for the most part this variant is associated with carrier strains. Variant 1 also contained one isolate from each of the clonal complexes ST-103, ST-41/44 and ST-23 (serogroup Y) and two isolates of the serogroup A ST-1 lineage. Most of these atypical strains contained within the sequence variant 1 were clustered into subvariant 1a.

Fig. 3. Split decomposition network based on analyses of the *fadL* gene sequences representing the four main meningococcal variants and subvariant 1a. Gene sequences from four neisserial commensals were added to the analysis. The fit was 61.8 (with bootstrapping values of 100).

Recombinational basis of genetic diversity in *fadL*

Analysis of the recombination rate for meningococcal *fadL* gave a recombination parameter estimate of $R=3.24$ and the r/m value calculated by ClonalFrame was 1.81, indicating that this gene has been subject to an intermediate level of recombination events. The r/m value calculated for the 68 neisserial strains, including sequences from commensals, increased to 2.41 ($R=8.63$). GARD was applied to rapidly screen the multiple-sequence alignment for recombination. Further analysis using GARD identified one recombination break point located at nucleotide position 765 (MC58 sequence) with the change in the corrected AIC (ΔAIC_c) = 265.07 ($P < 0.001$).

The putative break point predicted by GARD is located between DNA sequences encoding loops 3 (VR1) and 4 (VR2) (Fig. 2a). Neighbour-joining trees were constructed from the nucleotide sequences encoding these loop regions from FadL like protein. For the two major variable loops containing VR1 and VR2, distinct clusters, grouping the different variants, were formed (Fig. 2b, c). The region encoding VR1 in the variant 1 sequence was identical to the DNA sequence from *N.*

sicca. VR1 regions in variants 2 and 3 also have identical sequences and clustered near the *N. lactamica* strain Y92-1009. In VR2 the FadL-like protein variants 1 and 2 formed a distinct group closely related to both *N. lactamica* strains, but the variant 3 and subvariant 1a were more closely related to *N. sicca* and *N. polysaccharae*. This demonstrates the mosaic structure of the gene and suggests that both inter- and intraspecies recombination has occurred over time. When the neighbour-joining analysis was done on loop 5 (VR3) sequences, the groups were indistinct and not well supported by bootstrap values (figure not included).

A subset of diverse FadL like protein sequences, containing an example of each of the four variants and the sequences from the commensal isolates, were analysed by the split decomposition method (Fig. 3). The resulting representation of FadL-like protein diversity had a net-like phylogeny providing further evidence that recombinational events played a role in its evolution (Fig. 3). Similar results were observed when the subset of sequences analysed by SplitsTree corresponded to the most variable strains within sequence variants 1 (Fig. 4a) and 2 (Fig. 4b) indicating that

Fig. 4. Split graphs based on analyses of the *fadL* gene sequences in the most diverse strains within variants 1 (a) and 2 (b).

recombination in the *fadL* gene is not limited to inter-species or inter-variant isolates.

Evidence of selection

Three methods were employed to evaluate whether the FadL like protein was under selective pressure. In the first method only the 34 distinct meningococcal sequences were used in the maximum-likelihood analysis with the CODEML program. The analysis provided evidence of positive selection. For this alignment ($n=34$), the best-supported M2a and M8 models ($l=-3015$) estimated ω ($d_N:d_S$) parameters >1 , indicative of positive selection. Estimation of parameters for model M2a showed that $\sim 6.2\%$ of sites fell into a relatively weakly positively selected class, where $\omega_2=4.39$ ($\sim 7\%$ with $\omega=4.11$ for model M8). The remaining $\sim 93\%$ of sites were highly conserved. Bayesian methods were used to identify the sites with the highest probability of falling into the positively selected class under the M2a and M8 models (Table 3).

In addition codon-by-codon analysis of selection on the gene was done using the Bayesian approach in the software package OMEGAMAP and the Nei–Gojobori method was used to estimate the ratio of synonymous and non-synonymous changes per site across the gene using the SNAP analysis tool. For these analyses all 64 meningococcal sequences were included; there was evidence of positive selection at 46 (10.8%) sites according to the Bayesian approach and at 38 sites (9%) using the Nei–Gojobori method. Positions predicted to be under positive selection for the FadL-like protein were mapped onto the topological structure previously predicted for this protein (Sardiñas *et al.*, 2009). All the residues identified as likely to be under positive selection are located in the exposed loops 3, 4 and 5 of the protein (Table 3). Using the Nei–Gojobori method to calculate the ω of each of the variable regions, the results were VR1=1.3, VR2=2.5 and VR3=14.4.

Finally, the question of whether recombination could have produced a false-positive signal for positive selection was addressed by repeating the analyses with sequences belonging to the same FadL variant. Using the maximum-likelihood model for variants 1, 3 and 4, comparison of model M1 with M2a suggests variations in the ω ratio between sites; however, no sign of positive selection was detected. Comparison of model M7 with M8 did not detect selection within these variants. Conversely, analysis of the variant 2 sequences produced significant evidence for positive selection ($P<0.05$ for M1 vs M2a and M7 vs M8), with the highest ω values at the selected sites. However, only one site (341K), also predicted in the original analysis (Table 3), fell into the positively selected class (full results are available from the authors on request). Using the Bayesian approach in OMEGAMAP to analyse variant 1 sequences, 24 amino acids (5.7%) had ω values greater than 1 (Table 3). This decrease from the value of 10.8% of amino acids under selection when analysing all 64 strains resulted from none of the amino acids in VR1, loop 3,

having ω values of >1 . When the analysis was carried out on the sequences from strains with the variant 2 protein the number of amino acids under selection was 7 (1.7%). In this analysis only amino acids in VR3, loop 5, had ω values >1 (Table 3). The analysis was not possible on the variant 3 or 4 proteins due to the high level of sequence conservation. In all the analyses undertaken only one site (341K) was classed as being under positive selection. Additionally, a recombination (GARD) screen run prior to performing selection analyses via the FEL program available through the Datamonkey server revealed only one positively selected site (276S).

A number of residues were identified which were conserved in the variable regions in all the sequences analysed. These are highlighted in Table 3.

DISCUSSION

The lack of an effective vaccine against serogroup B disease has fuelled the hunt for a broadly cross-protective meningococcal vaccine candidate for many years. To date only vaccines consisting of outer-membrane vesicles have been used to control outbreaks of serogroup B disease (Sadarangani & Pollard, 2010). A number of studies have examined the antigen content of the vesicles using proteomic-based technology, and the FadL-like protein encoded by ORF NMB0088 in MC58 was identified in four of these studies (Uli *et al.*, 2006; Vaughan *et al.*, 2006; Vipond *et al.*, 2006; Williams *et al.*, 2007). Here we compared the sequences of the gene encoding the protein in a panel of 64 different meningococcal isolates of clinical and epidemiological relevance.

The immunogenic outer-membrane epitopes of the meningococcus are subject to immunoselective pressure and are therefore variable. As the meningococcus only colonizes humans this means that the variability of these antigens is a reflection of potentially protective immunity in the human host. Antigenic variability, although a hindrance for the development of a cross-protective vaccine, can thus be considered the hallmark of a good vaccine candidate. As more evidence emerges on the antigenic variability of candidate protein antigens, it seems increasingly likely that a vaccine offering broad coverage will be multivalent, containing sufficient epitopes to offer protection against diverse meningococcal isolates (Feavers & Pizza, 2009). This study examined the evolutionary mechanisms acting on the FadL-like protein, the resultant levels of diversity and ultimately its potential as a vaccine candidate. Although the role of the FadL-like protein remains to be identified in the meningococcus, its homologue in the gonococcus is a pseudogene. That the protein is not essential for survival in the urinogenital tract provides a clue that may ultimately help elucidate its function. The predicted structure of the protein encoded by the meningococcal FadL-like protein was described by Sardiñas *et al.* (2009), using the FadL protein of *E. coli* as a

Table 3. Estimation of positive selection sites in variable regions of meningococcal FadL like protein

Protein region	Codon*	Amino acid	CODEML†		OMEGAMAP			SNAP		
			M2a	M8	Bayesian $d_N : d_S$			Nei-Gojobori $d_N : d_S > 1$		
			Bayesian method		All	V1	V2	All‡	V1‡	V2‡
			All‡	All‡						
Loop 3 VR1	193	K	+	+	2.29	0.06	0.09	+		
	194	S	+	+	2.28	0.06	0.09	+		
	195	K			2.22	0.06	0.09			
	196	A			2.20	0.06	0.09			
	197	E			2.20	0.06	0.09	+		
	198	I	+	+	2.19	0.06	0.08	+		
	199	L			2.13	0.06	0.08			
	200	T			2.12	0.06	0.08	+		
	201	A			2.06	0.06	0.08	+		
	202	K			1.91	0.06	0.08	+		
	203	P			1.77	0.06	0.08			
	204	P			1.74	0.06	0.08			
	205	K			1.69	0.06	0.08	+		
	206	P			1.64	0.06	0.08			
	207	N			1.61	0.06	0.08			
	208	G			1.60	0.06	0.08			
	209	V			1.59	0.06	0.08			
	210	A			1.60	0.06	0.08			
	211	E		+	1.61	0.06	0.08	+		
	212	A			1.06	0.06	0.08			
	213	A			0.98	0.06	0.08			
	214	K			0.95	0.06	0.08	+		
	215	I			0.77	0.06	0.08			
	216	Q			0.72	0.06	0.08	+		
Loop 4 VR2	268	G			1.08	0.17	0.17	+		+
	269	A			1.16	0.38	0.16			
	270	A	+	+	1.57	2.93	0.16	+	+	
	271	A			1.57	2.94	0.15			
	272	K			1.60	3.00	0.14			
	273	A	+	+	1.71	3.11	0.14	+	+	
	274	M	+	+	1.73	3.11	0.13	+	+	
	275	W			1.73	3.10	0.13			
	276	S	+	+	1.73	3.10	0.12	+	+	
	277	T	+	+	1.74	3.09	0.12	+		
	278	M			1.72	3.07	0.12	+		
	279	L			1.69	3.06	0.12			
	280	A	+	+	1.68	3.05	0.12			
	281	A			1.71	3.05	0.11	+	+	
	282	N			1.70	3.05	0.11	+	+	
	283	G			1.34	3.03	0.11			
	284	Y			1.19	2.77	0.11			
	285	T			1.13	2.70	0.10			
	286	A			1.11	2.68	0.10			
Loop 5 VR3	287	N			1.72	2.65	0.10			
	339	V			1.22	1.80	9.40	+	+	+
	340	V			1.28	1.79	9.98	+		+
	341	K	++	++	1.28	1.77	9.98	+	+	+
	342	G			1.16	1.34	6.53			
	343	K			1.13	1.17	5.91			
	344	S			1.13	1.11	5.67			
	345	D			0.93	0.47	5.23			

Table 3. cont.

*The codon position and matching amino acid correspond to NMB0088 ORF of MC58. Conserved residues in the variable regions in all the sequences analysed are shown in bold type.

†M2a and M8 were estimated using the CODEML program; estimates of parameters and the likelihood-ratio test are available from the authors upon request.

‡+ and ++ indicate an ω value of >1 or residue under positive selection when only variant 2 sequences were processed according to the M2a and M8 Bayesian method respectively. In some cases all sequences were evaluated or only variant 1 (V1) or variant 2 (V2) sequences were included in the analysis.

template. In accordance with other members of the FadL family, the meningococcal protein is predicted to be a transporter protein consisting of three domains: inner membrane, transmembrane and outer membrane (van den Berg, 2005).

The neisserial FadL-like proteins contained three variable regions; one mechanism known to generate this variability is diversifying selection. Diversifying selection pressure acting on highly immunogenic proteins in *Neisseria* spp. promotes allelic variation and has been described for a number of antigens, including PorA (Derrick *et al.*, 1999), PorB (Urwin *et al.*, 2002), FetA (Thompson *et al.*, 2003) and fHbp (Brehony *et al.*, 2009). Similarly the FadL homologue of *Haemophilus influenzae* has been shown to be under diversifying selective pressure (Mes & van Putten, 2007). A codon-by-codon analysis of the ratio of non-synonymous to synonymous base changes (ω parameter) in the gene encoding the FadL-like protein of the meningococcus identified that a number of residues located in cell-surface-exposed loops were under diversifying selection. Different models gave different predictions for the number of residues under positive selection, the highest predicting 48 and the lowest 10. Furthermore, the role of mutation in the diversification of the FadL-like protein differs between the VRs. The ω value calculated using the Nei–Gobojori method across the variable regions predicted VR1 and VR2 to be under weak diversifying selection whereas VR3 was under strong positive selection.

Recombination appears to be an important mechanism for generating the diversity within VR1 and VR2. The *fadL* gene consists of highly divergent regions flanked by highly conserved sequence, a structure that would facilitate homologous recombination between isolates and allow divergent sequences to be interchanged by horizontal genetic transfer. Furthermore, *fadL* is located in a chromosome region that may be genetically unstable, due to the presence of repeat sequences flanking the gene. Specifically *fadL* is flanked by CREE and neisserial DNA uptake sequences (DUS) (Sardiñas *et al.*, 2009). CREE sequences may serve to promote recombination with exogenously acquired DNA, increasing the rate of genetic exchange at the adjacent loci (Bentley *et al.*, 2007). Analysis of the distribution of base changes throughout *fadL* showed that nucleotide sequences were shared among several distantly related isolates. As a result the *fadL* gene has what is described as a mosaic structure (Spratt *et al.*,

1992), similar to genes encoding other meningococcal surface proteins such as *penA* (Spratt *et al.*, 1992), *porA* (Feavers *et al.*, 1992), *porB* (Dyet & Martin, 2005) and most recently *fHbp* (Beernink & Granoff, 2009). In the FadL-like protein, there is also evidence for interspecies recombination given the similarities between regions of amino acid sequence in *N. polysaccharaea*, *N. sicca* and *N. lactamica* with *N. meningitidis*. It is interesting to note that 71 % of the strains with the variant 2 protein are carriage isolates which live in the upper airways of healthy humans in association with other bacteria. The *N. lactamica* and variant 2 strains have homologous VR2 sequences, and it could be postulated that carriage of both species resulted in the acquisition by the meningococcus of DNA from *N. lactamica*. In common with other meningococcal antigens (Beernink & Granoff, 2009; Bennett *et al.*, 2009; Linz *et al.*, 2000), the interspecies horizontal genetic exchange has generated a family of antigenically diverse meningococcal FadL-like proteins.

Phylogenies based on the FadL-like membrane protein provided evidence for antigenic structuring, despite evidence of horizontal genetic exchange between isolates. Most isolates belonging to the same clonal complex tend to cluster together. However, the *fadL* alleles belonging to the ST-1, ST-41/44 and ST-103 complexes included two sequence variants each and exhibited more diversity. Associations of particular antigen types, namely PorA, PorB, FetA, Opa, fHbp, NadA, HmbR and NHBA, with clonal complexes have been identified in other studies (Bambini *et al.*, 2009; Brehony *et al.*, 2009; Callaghan *et al.*, 2006; Caugant *et al.*, 1987; Evans *et al.*, 2010; Suker *et al.*, 1994; Urwin *et al.*, 2004). This antigenic structuring offers the possibility of the development of a vaccine consisting of relatively few antigens to protect against hyperinvasive meningococci (Urwin *et al.*, 2004). Assuming that the FadL-like protein were to offer protection in humans and that there is no cross-protection between variants, the results of this study suggested that four variants would offer coverage against at least five major invasive meningococcal lineages.

As in previous studies of other meningococcal vaccine candidates (Brehony *et al.*, 2009; Evans *et al.*, 2010; Urwin *et al.*, 2004), we have adopted a bioinformatic approach to assess the vaccine potential of the meningococcal FadL-like protein. The limitations of using genomic data are that it does not provide definitive information on the functional

immune response to the protein or the phenotype of the strain in question. For example, expression levels or surface accessibility of the protein, both of which can affect protection conferred by an antigen, remain unknown (Welsch *et al.*, 2004; Martin *et al.*, 2006). Bactericidal killing in the presence of complement, as determined by the serum bactericidal assay (SBA), is the functional assay and accepted correlate of protection for meningococcal vaccines (Frasch *et al.*, 2009). Furthermore the demonstration of bactericidal killing by serum from immunized animals is a prerequisite for subsequent clinical studies. However, due to the differences in the immune systems of humans and animals, preclinical data based on animal models may not provide an accurate reflection of the potential of a vaccine candidate in humans. The FadL-like protein elicits a bactericidal antibody response in mice (Sardiñas *et al.*, 2009), but the comparative analyses of sequence data presented here suggest it is not a major target of protective immunity in humans. To further evaluate the potential of the FadL protein as a vaccine candidate, SBAs employing human serum should be undertaken across a panel of strains. Unfortunately standardization of SBAs is fraught with difficulties, and as such a comprehensive screen utilizing this method would be arduous.

This bioinformatic approach assumes that immunoselective pressure acting on a protein indicates a functional immune response. Availability of multiple genome sequences, a consequence of advances in high-throughput sequencing technology, will make it possible to undertake such an approach for any vaccine candidate before embarking on preclinical and clinical studies involving complex serological analyses.

ACKNOWLEDGEMENTS

The authors would like to thank Dr Daniel Wilson for his support and assistance with the use of the OMEGAMAP software package.

REFERENCES

- Bambini, S., Muzzi, A., Olcen, P., Rappuoli, R., Pizza, M. & Comanducci, M. (2009). Distribution and genetic variability of three vaccine components in a panel of strains representative of the diversity of serogroup B meningococcus. *Vaccine* **27**, 2794–2803.
- Beernink, P. T. & Granoff, D. M. (2009). The modular architecture of meningococcal factor H-binding protein. *Microbiology* **155**, 2873–2883.
- Bennett, J. S., Thompson, E. A., Kriz, P., Jolley, K. A. & Maiden, M. C. (2009). A common gene pool for the *Neisseria* FetA antigen. *Int J Med Microbiol* **299**, 133–139.
- Bentley, S. D., Vernikos, G. S., Snyder, L. A., Churcher, C., Arrowsmith, C., Chillingworth, T., Cronin, A., Davis, P. H., Holroyd, N. E. & other authors (2007). Meningococcal genetic variation mechanisms viewed through comparative analysis of serogroup C strain FAM18. *PLoS Genet* **3**, e23.
- Brehony, C., Wilson, D. J. & Maiden, M. C. (2009). Variation of the factor H-binding protein of *Neisseria meningitidis*. *Microbiology* **155**, 4155–4169.
- Callaghan, M. J., Jolley, K. A. & Maiden, M. C. (2006). Opacity-associated adhesin repertoire in hyperinvasive *Neisseria meningitidis*. *Infect Immun* **74**, 5085–5094.
- Callaghan, M. J., Buckee, C. O., Jolley, K. A., Kriz, P., Maiden, M. C. & Gupta, S. (2008). The effect of immune selection on the structure of the meningococcal opa protein repertoire. *PLoS Pathog* **4**, e1000020.
- Caugant, D. A., Mocca, L. F., Frasch, C. E., Froholm, L. O., Zollinger, W. D. & Selander, R. K. (1987). Genetic structure of *Neisseria meningitidis* populations in relation to serogroup, serotype, and outer membrane protein pattern. *J Bacteriol* **169**, 2781–2792.
- Chung, G. T., Yoo, J. S., Oh, H. B., Lee, Y. S., Cha, S. H., Kim, S. J. & Yoo, C. K. (2008). Complete genome sequence of *Neisseria gonorrhoeae* NCCP11945. *J Bacteriol* **190**, 6035–6036.
- Derrick, J. P., Urwin, R., Suker, J., Feavers, I. M. & Maiden, M. C. (1999). Structural and evolutionary inference from molecular variation in *Neisseria* porins. *Infect Immun* **67**, 2406–2413.
- Didelot, X. & Falush, D. (2007). Inference of bacterial microevolution using multilocus sequence data. *Genetics* **175**, 1251–1266.
- Dyet, K. H. & Martin, D. R. (2005). Sequence variation in the *porB* gene from B:P1.4 meningococci causing New Zealand's epidemic. *J Clin Microbiol* **43**, 838–842.
- Evans, N. J., Harrison, O. B., Clow, K., Derrick, J. P., Feavers, I. M. & Maiden, M. C. (2010). Variation and molecular evolution of HmbR, the *Neisseria meningitidis* haemoglobin receptor. *Microbiology* **156**, 1384–1393.
- Feavers, I. M. & Pizza, M. (2009). Meningococcal protein antigens and vaccines. *Vaccine* **27** ((Suppl. 2),), B42–B50.
- Feavers, I. M., Heath, A. B., Bygraves, J. A. & Maiden, M. C. (1992). Role of horizontal genetic exchange in the antigenic variation of the class 1 outer membrane protein of *Neisseria meningitidis*. *Mol Microbiol* **6**, 489–495.
- Finne, J., Bitter-Suermann, D., Goridis, C. & Finne, U. (1987). An IgG monoclonal antibody to group B meningococci cross-reacts with developmentally regulated polysialic acid units of glycoproteins in neural and extraneural tissues. *J Immunol* **138**, 4402–4407.
- Frasch, C. E., Borrow, R. & Donnelly, J. (2009). Bactericidal antibody is the immunologic surrogate of protection against meningococcal disease. *Vaccine* **27** ((Suppl. 2),), B112–B116.
- Guindon, S. & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol* **52**, 696–704.
- Gupta, S. & Maiden, M. C. (2001). Exploring the evolution of diversity in pathogen populations. *Trends Microbiol* **9**, 181–185.
- Gupta, S., Maiden, M. C., Feavers, I. M., Nee, S., May, R. M. & Anderson, R. M. (1996). The maintenance of strain structure in populations of recombining infectious agents. *Nat Med* **2**, 437–442.
- Häyrynen, J., Jennings, H., Raff, H. V., Rougon, G., Hanai, N., Gerardy-Schahn, R. & Finne, J. (1995). Antibodies to polysialic acid and its *N*-propyl derivative: binding properties and interaction with human embryonal brain glycopeptides. *J Infect Dis* **171**, 1481–1490.
- Hearn, E. M., Patel, D. R. & van den Berg, B. (2008). Outer-membrane transport of aromatic hydrocarbons as a first step in biodegradation. *Proc Natl Acad Sci U S A* **105**, 8601–8606.
- Huson, D. H. & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**, 254–267.
- Jódar, L., Feavers, I. M., Salisbury, D. & Granoff, D. M. (2002). Development of vaccines against meningococcal disease. *Lancet* **359**, 1499–1508.
- Keane, T. M., Creevey, C. J., Pentony, M. M., Naughton, T. J. & McInerney, J. O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc

- assumptions for choice of matrix are not justified. *BMC Evol Biol* 6, 29.
- Kosakovsky Pond, S. L. & Frost, S. D. (2005a). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol Biol Evol* 22, 1208–1222.
- Kosakovsky Pond, S. L. & Frost, S. D. (2005b). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21, 2531–2533.
- Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H. & Frost, S. D. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* 23, 1891–1901.
- Kumar, S., Tamura, K. & Nei, M. (2004). MEGA3: integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5, 150–163.
- Lanave, C., Preparata, G., Saccone, C. & Serio, G. (1984). A new method for calculating evolutionary substitution rates. *J Mol Evol* 20, 86–93.
- Lewis, S., Sadarangani, M., Hoe, J. C. & Pollard, A. J. (2009). Challenges and progress in the development of a serogroup B meningococcal vaccine. *Expert Rev Vaccines* 8, 729–745.
- Librado, P. & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.
- Linz, B., Schenker, M., Zhu, P. & Achtman, M. (2000). Frequent interspecific genetic exchange between commensal *neisseriae* and *Neisseria meningitidis*. *Mol Microbiol* 36, 1049–1058.
- Liu, S. V., Saunders, N. J., Jeffries, A. & Rest, R. F. (2002). Genome analysis and strain comparison of *Correia* repeats and *Correia* repeat-enclosed elements in pathogenic *Neisseria*. *J Bacteriol* 184, 6163–6173.
- Martin, D. R., Ruijter, N., McCallum, L., O'Hallahan, J. & Oster, P. (2006). The VR2 epitope on the PorA P1.7-2.4 protein is the major target for the immune response elicited by the strain-specific group B meningococcal vaccine MeNZB. *Clin Vaccine Immunol* 13, 486–491.
- Mes, T. H. & van Putten, J. P. (2007). Positively selected codons in immune-exposed loops of the vaccine candidate OMP-P1 of *Haemophilus influenzae*. *J Mol Evol* 64, 411–422.
- Mitka, M. (2005). New vaccine should ease meningitis fears. *JAMA* 293, 1433–1434.
- Nei, M. & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. *Mol Biol Evol* 3, 418–426.
- Nielsen, R. & Yang, Z. (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148, 929–936.
- Pajón, R., Yero, D., Niebla, O., Climent, Y., Sardiñas, G., García, D., Perera, Y., Llanes, A., Delgado, M. & other authors (2009). Identification of new meningococcal serogroup B surface antigens through a systematic analysis of *neisserial* genomes. *Vaccine* 28, 532–541.
- Parkhill, J., Achtman, M., James, K. D., Bentley, S. D., Churcher, C., Klee, S. R., Morelli, G., Basham, D., Brown, D. & other authors (2000). Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. *Nature* 404, 502–506.
- Peeters, C. C., Claassen, I. J., Schuller, M., Kersten, G. F., van der Voort, E. M. & Poolman, J. T. (1999). Immunogenicity of various presentation forms of PorA outer membrane protein of *Neisseria meningitidis* in mice. *Vaccine* 17, 2702–2712.
- Peng, J., Yang, L., Yang, F., Yang, J., Yan, Y., Nie, H., Zhang, X., Xiong, Z., Jiang, Y. & other authors (2008). Characterization of ST-4821 complex, a unique *Neisseria meningitidis* clone. *Genomics* 91, 78–87.
- Pizza, M., Scarlato, V., Masignani, V., Giuliani, M. M., Arico, B., Comanducci, M., Jennings, G. T., Baldi, L., Bartolini, E. & other authors (2000). Identification of vaccine candidates against serogroup B meningococcus by whole-genome sequencing. *Science* 287, 1816–1820.
- Rosenstein, N. E., Perkins, B. A., Stephens, D. S., Popovic, T. & Hughes, J. M. (2001). Meningococcal disease. *N Engl J Med* 344, 1378–1388.
- Sadarangani, M. & Pollard, A. J. (2010). Serogroup B meningococcal vaccines – an unfinished story. *Lancet Infect Dis* 10, 112–124.
- Sardiñas, G., Yero, D., Climent, Y., Caballero, E., Cobas, K. & Niebla, O. (2009). *Neisseria meningitidis* antigen NMB0088: sequence variability, protein topology and vaccine potential. *J Med Microbiol* 58, 196–208.
- Spratt, B. G., Bowler, L. D., Zhang, Q. Y., Zhou, J. & Smith, J. M. (1992). Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal *Neisseria* species. *J Mol Evol* 34, 115–125.
- Suker, J., Feavers, I. M., Achtman, M., Morelli, G., Wang, J. F. & Maiden, M. C. (1994). The *porA* gene in serogroup A meningococci: evolutionary stability and mechanism of genetic variation. *Mol Microbiol* 12, 253–265.
- Tettelin, H., Saunders, N. J., Heidelberg, J., Jeffries, A. C., Nelson, K. E., Eisen, J. A., Ketchum, K. A., Hood, D. W., Peden, J. F. & other authors (2000). Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. *Science* 287, 1809–1815.
- Thompson, E. A., Feavers, I. M. & Maiden, M. C. (2003). Antigenic diversity of meningococcal enterobactin receptor FetA, a vaccine component. *Microbiology* 149, 1849–1858.
- Uli, L., Castellanos-Serra, L., Betancourt, L., Dominguez, F., Barberá, R., Sotolongo, F., Guillén, G. & Pajón, F. R. (2006). Outer membrane vesicles of the VA-MENGOC-BC[®] vaccine against serogroup B of *Neisseria meningitidis*: analysis of protein components by two-dimensional gel electrophoresis and mass spectrometry. *Proteomics* 6, 3389–3399.
- Urwin, R., Holmes, E. C., Fox, A. J., Derrick, J. P. & Maiden, M. C. (2002). Phylogenetic evidence for frequent positive selection and recombination in the meningococcal surface antigen PorB. *Mol Biol Evol* 19, 1686–1694.
- Urwin, R., Russell, J. E., Thompson, E. A., Holmes, E. C., Feavers, I. M. & Maiden, M. C. (2004). Distribution of surface protein variants among hyperinvasive meningococci: implications for vaccine design. *Infect Immun* 72, 5955–5962.
- van den Berg, B. (2005). The FadL family: unusual transporters for unusual substrates. *Curr Opin Struct Biol* 15, 401–407.
- van den Berg, B., Black, P. N., Clemons, W. M., Jr & Rapoport, T. A. (2004). Crystal structure of the long-chain fatty acid transporter FadL. *Science* 304, 1506–1509.
- van den Dobbela, G. P. J. M., van Dijken, H. H., Pillai, S. & van Alphen, L. (2007). Immunogenicity of a combination vaccine containing pneumococcal conjugates and meningococcal PorA OMVs. *Vaccine* 25, 2491–2496.
- Vaughan, T. E., Skipp, P. J., O'Connor, C. D., Hudson, M. J., Vipond, R., Elmore, M. J. & Gorringe, A. R. (2006). Proteomic analysis of *Neisseria lactamica* and *Neisseria meningitidis* outer membrane vesicle vaccine antigens. *Vaccine* 24, 5277–5293.
- Vipond, C., Suker, J., Jones, C., Tang, C., Feavers, I. M. & Wheeler, J. X. (2006). Proteomic analysis of a meningococcal outer membrane vesicle vaccine prepared from the group B strain NZ98/254. *Proteomics* 6, 3400–3413.
- Welsch, J. A., Rossi, R., Comanducci, M. & Granoff, D. M. (2004). Protective activity of monoclonal antibodies to genome-derived *neisserial* antigen 1870, a *Neisseria meningitidis* candidate vaccine. *J Immunol* 172, 5606–5615.

Williams, J. N., Skipp, P. J., Humphries, H. E., Christodoulides, M., O'Connor, C. D. & Heckels, J. E. (2007). Proteomic analysis of outer membranes and vesicles from wild-type serogroup B *Neisseria meningitidis* and a lipopolysaccharide-deficient mutant. *Infect Immun* **75**, 1364–1372.

Wilson, D. J. & McVean, G. (2006). Estimating diversifying selection and functional constraint in the presence of recombination. *Genetics* **172**, 1411–1425.

Yang, Z. (1994). Estimating the pattern of nucleotide substitution. *J Mol Evol* **39**, 105–111.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586–1591.

Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. M. (2000). Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**, 431–449.

Yang, Z., Wong, W. S. & Nielsen, R. (2005). Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol* **22**, 1107–1118.

Edited by: P. van der Ley