Strong Positive Selection and Recombination Drive the Antigenic Variation of the PilE Protein of the Human Pathogen *Neisseria meningitidis*

T. Daniel Andrews*,†,[1] and Takashi Gojobori†

**The Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, United Kingdom and †Centre for Information Biology and DNA Databank of Japan, National Institute of Genetics, Mishima, Shizuoka-ken 411-8540, Japan*

Manuscript received June 8, 2003
Accepted for publication September 14, 2003

ABSTRACT

The PilE protein is the major component of the *Neisseria meningitidis* pilus, which is encoded by the *pil*E/*pil*S locus that includes an expressed gene and eight homologous silent fragments. The silent gene fragments have been shown to recombine through gene conversion with the expressed gene and thereby provide a means by which novel antigenic variants of the PilE protein can be generated. We have analyzed the evolutionary rate of the *pil*E gene using the nucleotide sequence of two complete *pil*E/*pil*S loci. The very high rate of evolution displayed by the PilE protein appears driven by both recombination and positive selection. Within the semivariable region of the *pil*E and *pil*S genes, recombination appears to occur within multiple small sequence blocks that lie between conserved sequence elements. Within the hypervariable region, positive selection was identified from comparison of the silent and expressed genes. The unusual gene conversion mechanism that operates at the *pil*E/*pil*S locus is a strategy employed by *N. meningitidis* to enhance mutation of certain regions of the PilE protein. The silent copies of the gene effectively allow "parallelized" evolution of *pil*E, thus enabling the encoded protein to rapidly explore a large area of sequence space in an effort to find novel antigenic variants.

THE *Neisseria meningitidis* bacterium is a human pathogen and a causative agent of meningitis and septicemia. *N. meningitidis* most commonly achieves asymptomatic infection of the nasopharynx, yet in a small but significant portion of these infections the bacteria gain entry to the bloodstream where they cause meningococcemia. A key component of the *N. meningitidis* infection machinery is the pilus, which aids binding of the bacterium to both epithelial and endothelial cells of the human host. The pilus is a filamentous structure that extends several micrometers from the bacterial cell surface and is composed primarily of a large number of identical subunits of the pilin protein encoded at the *pil*E/*pil*S locus (Heckels 1989; reviewed in Nassif 1999).

The crystal structure of the highly homologous PilE protein from *N. gonorrhoeae* shows that the protein forms an asymmetrical "ladle"-like structure. The handle of the ladle is formed by an unusually long α-helix and is attached to a globular α-β roll domain, which forms the ladle bowl (Parge *et al.* 1995). When polymerized, the PilE subunits are proposed to form their characteristic elongated fiber structure such that the long α-helical handles of each subunit pack to form a hydrophobic core, around which the globular domains wrap in a spiral. PilE itself appears not to be the primary mediator of host cell attachment. This role is performed by PilC, a protein that has been found to copurify with PilE and is thought to form the pilus tip (Rudel *et al.* 1995).

The exposed regions of PilE are subject to intense scrutiny by the host immune system and display high levels of antigenic variation (Diaz *et al.* 1984; Perry *et al.* 1988; Heckels 1989; see Hamrick *et al.* 2001 for an example from *N. gonorrhoeae*). The conserved structural elements of PilE, especially the N-terminal third of the sequence that forms the long α-helical handle, are buried by hypervariable residues that are exposed to the host immune system (Parge *et al.* 1995; Forest *et al.* 1996). Comparative sequence analysis shows that PilE from both *N. meningitidis* and *N. gonorrhoeae* contains three general regions. These are named according to their degree of conservation (highly conserved, semivariable, and hypervariable; Haas and Meyer 1986; Figure 2). The N-terminal third of the PilE sequence (residues 1–53) forms the long α-helical "handle" of the PilE protein and is highly conserved. The adjacent semivariable region (residues 54–114) displays more sequence diversity, but also contains five strongly conserved sequence elements (Sv1–5; Potts and Saunders 1988; Aho *et al.* 2000). The remainder of the sequence that extends to the C terminus is the hypervariable region. This region surrounds two highly conserved structural cysteine residues and their conserved flanking sequence (Potts and Saunders 1988). Within the hypervariable region, variation between different PilE sequences is extreme, often including short insertions and deletions. This region corresponds to the most exposed residues of the PilE protein subunit that form the surface of the

[1]*Corresponding author:* The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, United Kingdom.

pilus fiber and hence are most exposed to the host immune system (Parge *et al.* 1995).

The *pil*E gene of *N. meningitidis* lies in an unusual locus. Immediately upstream of the gene are eight truncated pseudogene-like homologs called the "silent" pilin genes (*pil*S; Perry *et al.* 1988; Parkhill *et al.* 2000). The *pil*S sequences are not expressed and lack varying amounts of the 5′ end of the homologous *pil*E gene. Hence the *pil*S genes are generally composed of a portion of the semivariable region plus all or almost all of the hypervariable region. The high level of antigenic variation observed between PilE proteins is generated through gene conversion between the *pil*S gene fragments and the corresponding portion of the *pil*E gene (Haas and Meyer 1986; Segal *et al.* 1986). The mechanism by which this occurs has not yet been fully elucidated for *N. meningitidis*, but is better understood for *N. gonorrhoeae*. Within *N. gonorrhoeae*, the gene conversion process is almost certainly mediated via the combined activity of conserved regions within the *pil*E gene and a series of repetitive elements that are positioned between the *pil*S genes. At least one possible recombinase binding site has been identified downstream of the *pil*E gene and this is also thought to contribute to the gene conversion mechanism (Haas and Meyer 1986; Haas *et al.* 1992; Seifert 1996; Howell-Adams and Seifert 2000). In *N. meningitidis*, gene conversion is probably achieved through a generally similar mechanism, due to the presence of conserved repetitive elements between the *pil*S fragments and a conserved Sma/Cla repeat downstream of the *pil*E gene (Parkhill *et al.* 2000).

The tandem array of *pil*S genes upstream of the *pil*E gene provides the *N. meningitidis* bacterium with a set of "contingency genes" (see Moxon *et al.* 1994) that often enable the bacterium to evade the host immune response. While the *pil*S genes are pseudogenes, in effect they are expressed through the proxy of gene conversion. Via this proxy these genes are subjected to evolutionary forces similar to normally expressed genes. Genes that are exposed to a foreign immune system often display elevated evolutionary rates and in many cases are subject to strong positive Darwinian selection (for example, Jiggins *et al.* 2002; Urwin *et al.* 2002). In this study, we have appraised this possibility through a comparative analysis of the evolutionary rates of the *pil*E and *pil*S genes extracted from two completely sequenced *N. meningitidis* genomes.

MATERIALS AND METHODS

Sequence data and alignment: The complete nucleotide sequence of the genomes of *N. meningitidis* strain Z2491 (serogroup A; Parkhill *et al.* 2000) and strain MC58 (serogroup B; Tettelin *et al.* 2000) were obtained from http://www.sanger.ac.uk/Projects/N_meningitidis/ and ftp://ftp.tigr.org/pub/data/Microbial_Genomes/ respectively. The sequences of the expressed and silent pilin genes were manually extracted using the genomic annotation coordinates as a guide. Where the end coordinate of a silent gene fragment was ambiguous, the fragment was arbitrarily truncated either where the sequence homology between the fragment and the expressed gene ended or where the end of the alignment with the expressed gene was reached. However, in three cases the sequence of the silent pilin gene fragments extended past the end of the expressed gene as they had an obvious stop codon within a short distance of the end of the expressed gene sequence. Terminal stop codons were removed from the expressed genes and from those gene fragments that possessed them. The extracted sequences were translated and aligned using CLUSTALW (Thompson *et al.* 1994; version 1.82) and the alignment was adjusted by hand. The amino acid alignments were transposed back to nucleotide sequences to gain a codon-based alignment (available from the authors on request).

Detection of recombination: Putative recombination breakpoints were detected using the method of Maynard-Smith (1992) as implemented by the Maximum Chi-Squared program (version 1.0; available from http://www.biols.susx.ac.uk/Biochem/Molbiol/index.html). Recombining blocks of sequence with discordant phylogenetic topologies were identified using the method of Grassly and Holmes (1997) as implemented by the PLATO software package (version 2.11).

Tree reconstruction: Maximum likelihood trees were estimated using fastDNAml (Felsenstein 1981; Olsen *et al.* 1994; version 1.2) with the F84 model and the t_s:t_v ratio set to 2. Kishino and Hasegawa (1989) tests were used to determine trees that were not significantly worse than the maximum likelihood tree. Majority-rule support for the branching pattern of the maximum likelihood tree was determined using the "consense" program from the Phylip package (version 3.6a).

Analysis of selection: The maximum likelihood method of Yang *et al.* (2000), as implemented in the PAML software package (Yang 2000; version 3.13), was used to test whether positive selection has taken place recently at sites within the *pil*E/*pil*S gene locus. Applying likelihood ratio tests (LRTs) that used different pairs of nested sequence evolution models tested the hypothesis that some codon sites have been positively selected. Following the example of Yang *et al.* (2000) and applying the same model-naming scheme, we conducted LRTs using the model pairs of M1 (neutral) *vs.* M2 (selection), M1 *vs.* M3 (discrete), and M7 (β) *vs.* M8 ($\beta + \omega$). Additionally, likelihood estimation was performed using model M0 (one ratio). In each of these tests, values of $\omega < 1$, $\omega = 1$, or $\omega > 1$ were interpreted as being indicative of purifying selection, neutral evolution, or positive selection, respectively. Model M0 assumes all codon sites have the same value of ω and as such allows a test of selection across all sites. Model M1 allows two codon site classes with values of ω fixed at 0 and 1. Model M1 is commonly compared to models M2 and M3, which have one more codon site class that allows $\omega > 1$. LRTs of M1 *vs.* M2 and M1 *vs.* M3 test whether a model that allows some sites to evolve under positive selection better describes the data. If the likelihood score of models M2 or M3 is significantly better than that of model M0 and if $\omega > 1$ for one or more of the three values as estimated by model M2 or M3, then this is evidence that positive selection may exist in the data. Models M2 and M3 differ in the number of site classes that allow $\omega > 1$. Model M2 allows just one site class where $\omega > 1$, whereas M3 has the freedom to set $\omega > 1$ for all of its three site classes. An LRT between models M2 and M3 tests whether a single class of positively selected sites describes the data better than multiple classes do. Model M7 allows 10 codon site classes (each with a restriction $\omega < 1$), whereas M8 has one extra site class that allows $\omega > 1$. Once again, good evidence for positive selection is found if the likelihood score gained from M8 is significantly better than that from M7 and $\omega > 1$. The

likelihood ratio test statistic used to determine the level of significance was calculated as twice the difference of the likelihood scores estimated by each model ($2\Delta l$). The null distribution of such test statistics can be approximated by a χ^2 distribution, with the number of degrees of freedom calculated as the difference in the number of estimated parameters between models. Hence, the degrees of freedom for the M1 *vs.* M2, M1 *vs.* M3, M2 *vs.* M3, and M7 *vs.* M8 tests were 2, 4, 3, and 2, respectively (Yang *et al.* 2000).

Should LRTs indicate the presence of positive selection in a data set, it is then possible to perform an empirical Bayesian analysis to calculate for each codon site the posterior probability that it belonged to a positively selected codon class (Nielsen and Yang 1998). Where the results of LRTs indicate that this analysis was appropriate, this calculation was conducted using "codeml" from the PAML software package.

Ad hoc pairwise comparison of synonymous and nonsynonymous substitutions between sequences was calculated using the method of Nei and Gojobori (1986) as implemented in the "codeml" program of the PAML package.

RESULTS

An elevated rate of protein evolution in the *pil*E/*pil*S gene locus was initially identified from exhaustive analysis of synonymous and nonsynonymous evolutionary rates of homologous genes in the two complete *N. meningitidis* genome sequences. Even at the crudest level, over the semi- and hypervariable regions of the sequence, identity between the two Neisseria serogroup PilE protein sequences is lower (86.5%) than that for the nucleotide sequences (89.8%), indicating that extreme and active diversification of the protein sequence has taken place.

Analysis of recombination: The process of frequent gene conversion of the *pil*E gene by *pil*S gene fragments appears almost entirely unidirectional (Seifert 1996). However, the propensity of Neisseria species for genome rearrangement, recombination, and horizontal transfer caused us to appraise the possibility that homologous recombination might have produced chimeric *pil*S gene fragments. The conservation pattern of the *pil*E/*pil*S genes is such that highly divergent regions are flanked by highly conserved sequence, which suggests a mechanism by which homologous recombination could easily pass novel divergent regions between sequences. Such chimeric gene fragments will perturb subsequent tests of positive selection that assume the absence of recombination (see Anisimova *et al.* 2003).

The method of Maynard-Smith (1992) was used to test for the existence of a recombination breakpoint within the data set. This method essentially looks for clustering of variant sites between a pair of sequences. Given that this is the pattern of variation displayed by the *pil*E/*pil*S sequences it seemed intuitively likely that a breakpoint would be found that separated the two main regions of variability (the hypervariable region being separated from variation in the semivariable region by a region of high conservation). As expected, the results indicated a highly significant breakpoint in the conserved region between the semivariable and the hypervariable regions.

A second method for detection of recombination was employed. The method of Grassly and Holmes (1997) analyzes a group of aligned sequences for blocks within the alignment that significantly deviate from an imposed phylogeny. We constructed a maximum likelihood tree using the sequence of just the hypervariable region, truncated to include the two conserved Cys1 and Cys2 regions and the variable region that lies between them (Figure 1). The full alignment was then analyzed for regions within it that significantly deviated from this phylogeny. The test detected five regions of the aligned *pil*E/*pil*S sequences that were deemed to significantly deviate from the imposed phylogeny. Four small blocks were identified in the semivariable region (196–210, 223–237, 256–261, and 292–300; using the coordinate scheme of Potts and Saunders 1988) and one larger block was identified in the hypervariable region (403–459). Importantly, these putative recombining blocks neatly account for almost all of the sequence variation within the *pil*E/*pil*S sequences. It is hard to determine whether this implies that all sequence variation between the sequences is the result of recombination or that the test is perturbed by the functionally restricted coordinate range within which substantial sequence variation is able to occur.

Given the possibility that the *pil*E/*pil*S sequences may contain a number of independently recombining sequence blocks, all subsequent analysis used the data set of truncated sequences described above and used to determine the tree in Figure 1. Although truncation of the full sequences to the hypervariable region reduced the length of the analyzed sequences to just 135 nucleotides, this data set still contained >60% (or 52 out of 83) of the informative sites of the full data set. Informative sites were counted as being those positions that were variant within at least two sequences. Further tests for breakpoints using the method of Maynard-Smith (1992) failed to find significant breakpoints within this truncated data set.

Tests of positive selection: *Ad hoc* pairwise comparison of the hypervariable region of the *pil*S sequences with *pil*E sequences from both *N. meningitidis* strains shows that nonsynonymous substitutions are more prevalent than synonymous substitutions in almost all comparisons (Table 1). A number of pairwise comparisons show a large excess of nonsynonymous substitutions compared to synonymous substitutions. Pairwise comparison of the two expressed *pil*E sequences shows a small excess of nonsynonymous substitutions compared to synonymous substitutions. Comparisons between pairs of *pil*S fragments showed results of similar character and magnitude as shown in Table 1, but for brevity the matrix of nonsynonymous and synonymous substitution rates is not presented.

This pairwise demonstration of rapid protein evolu-

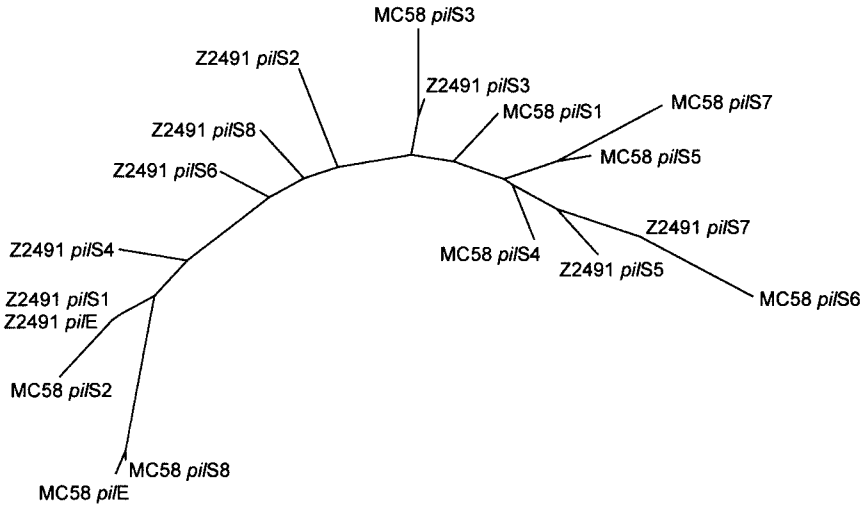

FIGURE 1.—Maximum likelihood (ML) tree reconstructed from the hypervariable region of the *pil*E/*pil*S genes from both *N. meningitidis* strains Z2491 and MC58. The length variable portion of the hypervariable region was excluded for the construction of this tree. Kishino-Hasegawa tests determined that 28 additional trees were not significantly less likely than this ML tree. Each node was supported by a consensus of 27/29 equally likely trees, except for the node grouping the sequences MC58 *pil*E and MC58 *pil*S8, which was maintained in each equally likely tree.

tion in the *pil*E/*pil*S locus motivated a further statistical analysis of whether positive selection could be an explanation. Figure 1 shows the reconstructed tree that was used for these tests. Table 2 shows the parameter estimates, likelihood scores, and the results of LRTs performed with the *pil*E and *pil*S genes. At the simplest level, the M0 model that allows just one class of codons shows that each site has an estimated value of $\omega = 1.51$. This alone indicates positive selection is acting in the hypervariable region of the *pil*E/*pil*S sequence. The results of the LRTs provide further support for this result. Each test (except that between models M2 and M3) strongly rejects the null hypothesis and indicates that positive selection may have taken place within this data set. For each of these tests, the test statistic was highly significant at $P < 0.0005$. The tests that employed models M1 *vs.* M2 and M1 *vs.* M3 showed clearly that a model that includes at least one site class that allows $\omega > 1$ (M2 and M3 models) describes the evolution of these sequences much better than a model that does not (model M1). Estimation of parameters for model M2 showed that the site class that allows $\omega > 1$ accounts for more than one-third of all sites and has a very high value of $\omega = 3.59$.

TABLE 1

Pairwise estimation of nonsynonymous and synonymous substitution rates between *pil*E sequences and between *pil*E and *pil*S sequences from *N. meningitidis* strains Z2491 and MC58

	$(d_N/d_S) = \omega$	
	*pil*E (Z2491 strain)	*pil*E (MC58 strain)
Z2491 strain		
*pil*E	—	(0.159/0.141) = 1.13
*pil*S1	(0.000/0.0340) = 0.00	(0.164/0.183) = 0.90
*pil*S2	(0.271/0.0606) = 4.47	(0.230/0.168) = 1.37
*pil*S3	(0.192/0.1776) = 1.08	(0.217/0.195) = 1.12
*pil*S4	(0.111/0.0379) = 2.94	(0.190/0.140) = 1.36
*pil*S5	(0.299/0.0504) = 5.94	(0.302/0.199) = 1.52
*pil*S6	(0.149/0.0114) = 13.1	(0.212/0.140) = 1.52
*pil*S7	(0.173/0.1237) = 1.40	(0.209/0.243) = 0.86
*pil*S8	(0.190/0.0568) = 3.34	(0.190/0.183) = 1.04
MC58 strain		
*pil*S1	(0.271/0.068) = 3.96	(0.289/0.221) = 1.31
*pil*S2	(0.069/0.035) = 1.97	(0.174/0.190) = 0.92
*pil*S3	(0.303/0.128) = 2.37	(0.306/0.251) = 1.22
*pil*S4	(0.221/0.169) = 1.31	(0.249/0.193) = 1.29
*pil*S5	(0.232/0.034) = 6.75	(0.256/0.166) = 1.54
*pil*S6	(0.296/0.097) = 3.05	(0.356/0.190) = 1.88
*pil*S7	(0.215/0.036) = 5.91	(0.193/0.177) = 1.09
*pil*S8	(0.146/0.142) = 1.02	(0.031/0.000)

The result of the likelihood estimation using model M3 shows extraordinarily that two of the three site classes have values $\omega > 1$. Between them, these two site classes account for almost two-thirds of all codons and have high estimated values of ω ($\omega_2 = 1.35$ and $\omega_3 = 3.76$). Furthermore, the result of the model M7 *vs.* M8 LRT also exhibits a similar pattern. The M8 model with its extra site class that allows for values of $\omega > 1$ describes the evolution of the pilin genes better than the M7 model does. More than one-third of all sites in the hypervariable region of the pilin gene are assigned to this positively selected site class, and the estimated value of ω for these sites is high at 3.21.

Table 2 also shows a listing of amino acid positions in the translated sequence that have strong support for belonging to a site class identified as being under possible positive selection (in the M2, M3, or M8 models). The concordance of the identity of the positively selected sites between each model is strong. With the M3 model, as two site classes are estimated to have values of $\omega > 1$, there are a greater number of sites with a high posterior probability of being positively selected. Between the M2 and M8 models, the positively selected sites predicted are in perfect agreement and are a subset of the sites predicted by model M3. The potentially

TABLE 2

Likelihood ratio tests of positive selection between *pil*E and *pil*S sequences

Model	Parameters in ω-distribution	l	$2\Delta l$	P	Positively selected codons[a]
M0 (one ratio)	ω1 = 1.51	−765.49			All
M1 (neutral)	ω1 = 0.000 $p1$ = 0.304 ω2 = 1.00 $p2$ = 0.696	−752.89			Not allowed
M2 (selection)	ω1 = 0.000 $p1$ = 0.269 ω2 = 1.00 $p2$ = 0.379 ω3 = 3.59 $p3$ = 0.352	−742.34	(M1 *vs.* M2) 21.09	<0.0005	125, 127, *131*, 137, *141*, *142*, *144*, 145
M3 (discrete)	ω1 = 0.100 $p1$ = 0.374	−741.75	(M1 *vs.* M3) 22.28	<0.0005	*125*, *127*, *128*, *129*, *130*, *131*, *132*, *133*, 134, *135*, *136*, *137*, *140*, *141*, *142*, *143*, *144*, *145*, *146*, *148*, 150
	ω2 = 1.35 $p2$ = 0.334 ω3 = 3.76 $p3$ = 0.293		(M2 *vs.* M3) 1.20	>0.5	
M7 (β)	p = 0.149 q = 0.099	−750.99			Not allowed
M8 (β + ω)	$p0$ = 0.631 p = 0.161 q = 1.67 $p1$ = 0.369 ω = 3.21	−742.09	(M7 *vs.* M8) 17.80	<0.0005	125, 127, *131*, 137, *141*, *142*, *144*, 145

Column headings l, $2\Delta l$, and P denote the likelihood score, test statistic, and level of significance of the test statistic, respectively.
[a] Codon positions in roman type indicate a posterior probability >0.95; codons in italic type indicate a posterior probability >0.99.

positively selected codons lie between the two conserved Cys regions and predominantly cluster closer to the Cys2 region (Figure 2). An alternative method of predicting the identity of positively selected sites (Suzuki and Gojobori 1999; Suzuki *et al.* 2001) was considered, but the relatively large number of sequences required for this method made it inappropriate for this analysis.

DISCUSSION

This analysis found that the *pil*E/*pil*S gene locus from *N. meningitidis* strains Z2491 and MC58 displays a strong case of positive selection. Given that the selection seems to work to produce novel antigens of the PilE protein, it may also be termed diversifying selection. Importantly, the selection among the *pil*E and *pil*S genes is detectable even though it is apparent that it must act through the proxy of gene conversion. This finding implies that novel amino acid changes in the *pil*E/*pil*S locus are highly important to *N. meningitidis* for evasion of the host immune response. We postulate that mutation and recombination within the silent pilin genes generates sequence diversity, which is then subject to strong selection should the silent fragment recombine with the expressed gene. Given that the mechanism for gene conversion does not alter the donor gene and that each of the *pil*S genes are preserved with valid reading frames, each silent fragment in the *pil*S locus possibly represents a "souvenir" of a gene conversion event that has led to a successful infection of a human host.

While gene conversion of the expressed gene by the silent copies is almost entirely a unidirectional process

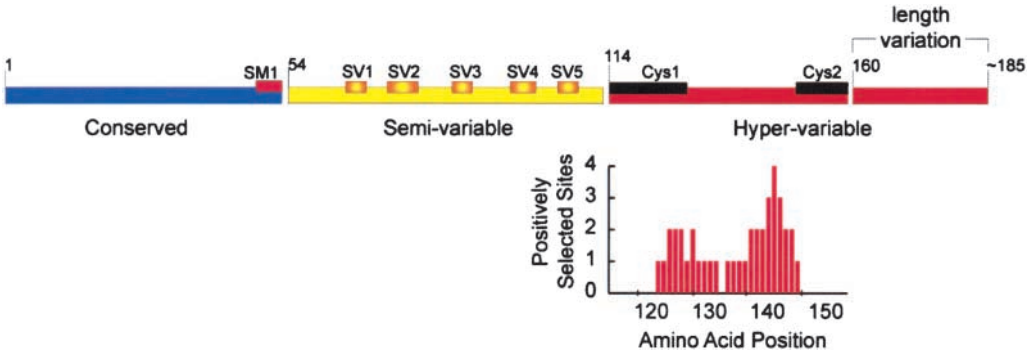

Figure 2.—Generalized primary structure and distribution of predicted positively selected sites of the PilE protein. Coordinates refer to those used by Potts and Saunders (1988). The number of positively selected sites was determined in a sliding window of five residues, using data presented in Table 2 for model M2. SM1 denotes the location of the SM1 epitope. SV1–5 denote the locations of the conserved motifs of the semivariable region. Cys1 and Cys2 denote the conserved residues that flank the conserved cysteine residues of the hypervariable region. The length variation at the carboxy terminus indicates the likely alteration in protein length that recombination of *pil*E with the extant *pil*S genes would produce.

(Seifert 1996), evidence derived from two different computational methods implies that homologous recombination between *pil*S fragments may occur. A low rate of reincorporation of *pil*E sequences into *pil*S sequences may also explain some of the observed recombination. The methods used for testing for the presence of positive selection assume an absence of recombination. When this assumption is violated, this greatly increases the occurrence of type I error (Anisimova *et al.* 2003). The data set used in this study was truncated to remove the effect of recombination or at least to isolate a single recombining region. Even so, the demonstration of positive selection in this data set must come with the warning that the full extent of recombination at this locus is not known. Anisimova *et al.* (2003) demonstrate that results of likelihood estimation under the M0 model are less sensitive to the effects of recombination. As shown in Table 2 the value of ω estimated using the M0 model was quite large (ω = 1.51). While the effect of recombination must be considered when the results of the LRTs are interpreted, overall it seems that positive selection is an important feature of the rapid protein evolution observed in the hypervariable region of the PilE protein.

Given the high evolutionary rate of the *pil*E genes, it is important to note that this rate may be enhanced by gene conversion with *pil*S genes from extrachromosomal DNA fragments. Neisseria species are well known for their autolytic behavior and their ability to selectively uptake other Neisserial DNA fragments (Sparling 1966; Sarubbi and Sparling 1974; Goodman and Scocca 1988). Furthermore, it has been shown *in vitro* that the presence of DNase greatly reduces Neisserial pilus variation (Seifert *et al.* 1988; Gibbs *et al.* 1989). Hence, much of the recombination that occurs with the *pil*E gene may be with extragenomic *pil*S sequences. This suggests that the value of a novel pilus antigenic form is so high that Neisseria species actively trawl their extracellular environment to find new *pil*E variants. It is interesting that the structure of the *N. gonorrhoeae* PilE protein indicates that the surface of the pilus nonspecifically binds DNA. The apparent lack of interstrain differentiation of genetic distances between *pil*E/*pil*S sequences analyzed in this study, given the speed with which these genes are evolving, could suggest that these genes have been passed between *N. meningitidis* strains in their recent past.

The Neisserial pilus consists mostly of PilE protein subunits, with the PilC protein located at the pilus tip to mediate host cell attachment (Rudel *et al.* 1995). Given that the PilE protein evolves quickly to evade the host immune system, it is probable that the PilC protein is subjected to similar levels of immune scrutiny. In *N. meningitidis*, the PilC protein is encoded by two genes, *pil*C1 and *pil*C2, although the latter appears to be absent from the predicted gene set of the Z2491 strain. We investigated the evolutionary rate of the PilC protein using two sequences from Rahman *et al.* (1997; accession nos. Y13020 and Y13021; annotated as *pil*C1 and *pil*C2, respectively) along with the *pil*C1 and *pil*C2 sequences from the MC58 strain and the sole *pil*C1 sequence from the Z2491 strain (data not shown). In pairwise analyses similar to those conducted with *pil*E, within the *pil*C2 sequences there was weak evidence for positive selection (ω = 1.13). Among the *pil*C1 sequences and between the *pil*C1 and *pil*C2 sequences, the synonymous substitution rate was marginally higher than the nonsynonymous rate (average ω = 0.843). While this information suggests that the *pil*C genes have a somewhat elevated rate of nonsynonymous evolution, it is generally less than that of the *pil*E/*pil*S genes. The apparent lack of silent *pil*C genes in the *N. meningitidis* genome implies that *pil*C1 and *pil*C2 do not evolve via the same gene conversion mechanism as the *pil*S/*pil*E locus and may evolve at a lower rate as a direct consequence of this. In addition, the structure of the PilC protein may not be as amenable to sustaining mutation, as compared to the hypervariable region of PilE protein, which seems to exist for the purpose of accommodating mutation.

If the gene conversion mechanism that operates within the *N. meningitidis pil*E/*pil*S locus is so effective at generating protein diversity for evading the host immune system, other organisms may also have employed this approach. Certainly, the *pil*E and *pil*S genes of the closely related pathogen *N. gonorrhoeae* are highly homologous to those found in *N. meningitidis*. Although the *pil*S fragments in *N. gonorrhoeae* are scattered throughout the genome, an almost identical mechanism of gene conversion between *pil*E and *pil*S has been shown to exist (see Seifert 1996). Using the same analytical method as applied here, the *pil*E and *pil*S genes of *N. gonorrhoeae* also display evidence of positive selection (our unpublished analysis).

Gene conversion has also been implicated as the mechanism for generating antigenic diversity among other proteins from other organisms. The outer membrane protein *msp*2 from *Anaplasma marginale* (Brayton *et al.* 2002), the hemagglutinin gene *vlhA* from *Mycoplasma synoviae* (Noormohammadi *et al.* 2000), and the surface-exposed lipoprotein *vls* from *Borrelia burgdorferi* (Zhang and Norris 1998) are a few good examples. In each case, these genes employ gene conversion to incorporate the genetic diversity of whole or fragmentary silent pseudogenes to generate antigenic variants. Further analysis is required to determine whether positive or diversifying selection has driven the evolution of these genes.

Compared to the gene loci of other organisms that employ gene conversion to generate antigenic variation, the *N. meningitidis pil*E/*pil*S locus is somewhat different in that the *pil*S genes are always found only as gene fragments. Genetic economy is possibly the main reason why the *pil*S genes are just fragments. However, if the *pil*S genes are always only fragmentary, this avoids the

potential problem where a silent gene may become "accidentally" expressed. If the silent genes were occasionally expressed, this would mean that not only would the *pil*E gene need to be antigenically novel, but also so would any expressed *pil*S genes. Due to the *pil*S genes being fragments and therefore never directly expressed, the effect of mutation and recombination at this locus is parallelized and concentrated. This parallelized evolution via the intermediacy of gene conversion may be an important factor that allows the *pil*E gene to evolve antigenic diversity at such a great rate.

The authors thank A. Wyndham for helpful discussions and two anonymous reviewers whose comments greatly improved this manuscript. This work was performed in part while T.D.A. was the recipient of a Japan Society for the Promotion of Science postdoctoral fellowship.

LITERATURE CITED

Aho, E. L., A. M. Keating and S. M. McGillivray, 2000 A comparative analysis of pilin genes from pathogenic and nonpathogenic Neisseria species. Microb. Pathog. **28:** 81–88.

Anisimova, M., R. Nielsen and Z. Yang, 2003 Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics **164:** 1229–1236.

Brayton, K. A., G. H. Palmer, A. Lundgren, J. Yi and A. F. Barbet, 2002 Antigenic variation of *Anaplasma marginale msp*2 occurs by combinatorial gene conversion. Mol. Microbiol. **43:** 1151–1159.

Diaz, J. L., M. Virji and J. E. Heckels, 1984 Structural and antigenic differences between two types of meningococcal pili. FEMS Microbiol. Lett. **21:** 181–184.

Felsenstein, J., 1981 Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17:** 368–376.

Forest, K. T., S. L. Bernstein, E. D. Getzoff, M. So, G. Tribbick *et al.*, 1996 Assembly and antigenicity of the *Neisseria gonorrhoeae* pilus mapped with antibodies. Infect. Immun. **64:** 644–652.

Gibbs, C. P., B. Y. Reimann, E. Schultz, A. Kaufmann, R. Haas *et al.*, 1989 Reassortment of pilin genes in *Neisseria gonorrhoeae* occurs by two distinct mechanisms. Nature **338:** 651–652.

Goodman, S. D., and J. J. Scocca, 1988 Identification and arrangement of the DNA sequence recognized in specific transformation of *Neisseria gonorrhoeae*. Proc. Natl. Acad. Sci. USA **85:** 6982–6986.

Grassly, N. C., and E. C. Holmes, 1997 A likelihood method for the detection of selection and recombination using nucleotide sequences. Mol. Biol. Evol. **14:** 239–247.

Haas, R., and T. F. Meyer, 1986 The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: evidence for gene conversion. Cell **44:** 107–115.

Haas, R., S. Veit and T. F. Meyer, 1992 Silent pilin genes of *Neisseria gonorrhoeae* MS11 and the occurrence of related hypervariant sequences among other gonococcal isolates. Mol. Microbiol. **6:** 197–208.

Hamrick, T. S., J. A. Dempsey, M. S. Cohen and J. G. Cannon, 2001 Antigenic variation of gonococcal pilin expression in vivo: analysis of the strain FA1090 pilin repertoire and identification of the *pil*S gene copies recombining with *pil*E during experimental human infection. Microbiology **147:** 839–849.

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

Howell-Adams, B., and H. S. Seifert, 2000 Molecular models accounting for the gene conversion reactions mediating gonococcal pilin antigenic variation. Mol. Microbiol. **37:** 1146–1158.

Jiggins, F. M., G. D. D. Hurst and Z. Yang, 2002 Host-symbiont conflicts: positive selection on an outer membrane protein of parasitic but not mutualistic Rickettsiaceae. Mol. Biol. Evol. **19:** 1341–1349.

Kishino, H., and M. Hasegawa, 1989 Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. Mol. Biol. Evol. **29:** 170–179.

Maynard-Smith, J., 1992 Analyzing the mosaic structure of genes. J. Mol. Evol. **34:** 126–129.

Moxon, E. R., P. B. Rainey, M. A. Nowak and R. E. Lenski, 1994 Adaptive evolution of highly mutable loci in pathogenic bacteria. Curr. Biol. **4:** 24–33.

Nassif, X., 1999 Interaction mechanisms of encapsulated meningococci with eucaryotic cells: What does this tell us about the crossing of the blood-brain barrier by *Neisseria meningitidis*? Curr. Opin. Microbiol. **2:** 71–77.

Nei, M., and T. Gojobori, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.

Nielsen, R., and Z. Yang, 1998 Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics **148:** 929–936.

Noormohammadi, A. H., P. F. Markham, A. Kanci, K. G. Whithear and G. F. Browning, 2000 A novel mechanism for control of antigenic variation in the haemagglutinin gene family of *Mycoplasma synoviae*. Mol. Microbiol. **35:** 911–923.

Olsen, G. J., H. Matsuda, R. Hagstrom and R. Overbeek, 1994 fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput. Appl. Biosci. **10:** 41–48.

Parge, H. E., K. T. Forest, M. J. Hickey, D. A. Christensen, E. D. Getzoff *et al.*, 1995 Structure of the fibre-forming protein pilin at 2.6 Å resolution. Nature **378:** 32–38.

Parkhill, J., M. Achtman, K. D. James, S. D. Bentley, C. Churcher *et al.*, 2000 Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491. Nature **404:** 502–506.

Perry, A. C., I. J. Nicolson and J. R. Saunders, 1988 *Neisseria meningitidis* C114 contains silent, truncated pilin genes that are homologous to *Neisseria gonorrhoeae pil* sequences. J. Bacteriol. **170:** 1691–1697.

Potts, W. J., and J. R. Saunders, 1988 Nucleotide sequence of the structural gene for class I pilin from *Neisseria meningitidis*: homologies with the *pil*E locus of *Neisseria gonorrhoeae*. Mol. Microbiol. **2:** 647–653.

Rahman, M., H. Kallstrom, S. Normark and A. B. Jonsson, 1997 *Pil*C of pathogenic Neisseria is associated with the bacterial cell surface. Mol. Microbiol. **25:** 11–25.

Rudel, T., I. Scheurerpflug and T. F. Meyer, 1995 Neisseria PilC protein identified as type-4 pilus tip-located adhesin. Nature **373:** 357–359.

Sarubbi, F. A., and P. F. Sparling, 1974 Transfer of antibiotic resistance in mixed cultures of *Neisseria gonorrhoeae*. J. Infect. Dis. **130:** 660–663.

Segal, E., P. Hagblom, H. S. Seifert and M. So, 1986 Antigenic variation of gonococcal pilus involves assembly of separated silent gene segments. Proc. Natl. Acad. Sci. USA **83:** 2177–2181.

Seifert, H. S., 1996 Questions about gonococcal pilus phase- and antigenic variation. Mol. Microbiol. **21:** 433–440.

Seifert, H. S., R. S. Ajioka, C. Marchal, P. F. Sparling and M. So, 1988 DNA transformation leads to pilin antigenic variation in *Neisseria gonorrhoeae*. Nature **336:** 392–395.

Sparling, P. F., 1966 Genetic transformation of *Neisseria gonorrhoeae* to streptomycin resistance. J. Bacteriol. **92:** 1364–1371.

Suzuki, Y., and T. Gojobori, 1999 A method for detecting positive selection at single amino acid sites. Mol. Biol. Evol. **16:** 1315–1328.

Suzuki, Y., T. Gojobori and M. Nei, 2001 ADAPTSITE: detecting positive selection at single amino acid sites. Bioinformatics **17:** 660–661.

Tettelin, H., N. J. Saunders, J. Heidelberg, A. C. Jeffries, K. E. Nelson *et al.*, 2000 Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58. Science **287:** 1809–1815.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Urwin, R., E. C. Holmes, A. J. Fox, J. P. Derrick and M. C. Maiden, 2002 Phylogenetic evidence for frequent positive selection and recombination in the meningococcal surface antigen PorB. Mol. Biol. Evol. **19:** 1686–1694.

Yang, Z., 2000 *Phylogenetic Analysis by Maximum Likelihood (PAML)*, Version 3.0. University College, London.

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

Zhang, J. R., and S. J. Norris, 1998 Genetic variation of the *Borrelia burgdorferi* gene *vls*E involves cassette-specific, segmental gene conversion. Infect. Immun. **66:** 3698–3704.

Communicating editor: Z. Yang