

# Population Genetics of the *porB* Gene of *Neisseria gonorrhoeae*: Different Dynamics in Different Homology Groups

David Posada,\* Keith A. Crandall,\* Man Nguyen,† James C. Demma,† and Raphael P. Viscidi†

\*Department of Zoology, Brigham Young University; and †Department of Pediatrics, The Johns Hopkins University School of Medicine

The *porB* locus codes for the major outer membrane protein of *Neisseria gonorrhoeae*. Alleles of this locus have been assigned to two homology groups based on close sequence and immunological relationships and are designated as either PIA or PIB. Several population parameters were estimated and compared among these two groups using a data set of 22 PIA sequences and 91 PIB sequences obtained from diverse geographic localities and from time periods spanning approximately 50 years. Recombination appears to be extensive in the *porB* gene. While the recombination rates are similar for the PIA and PIB sequences, the relative contribution of recombination to genetic diversity is higher for the PIA sequences. Alleles belonging to the PIB group show greater genetic diversity than do those in the PIA group. Although phylogenetic analysis did not reveal temporal or geographic clustering of sequences, estimates of gene flow and the fixation index suggested that PIB sequences exhibit population substructure based on geographic locality. Selection acts in these homology groups in a different way. While positive Darwinian selection is the dominant force driving the evolution of the PIA sequences, purifying selection operates also on the PIB sequences. These differences may be attributable to the greater propensity of PIA strains, as compared with PIB strains, to cause disseminated gonococcal infection, which would expose the former to intense selection pressure from the host immune system. The molecular evolution of *Neisseria gonorrhoeae* seems to be driven by the simultaneous action of selection and recombination, but under different rates and selection pressures for the PIA and PIB homology groups.

## Introduction

Gonorrhea is a common bacterial infection that is transmitted primarily by sexual contact or perinatally (for review, see Handsfield and Sparling 1995). The causative organism of gonorrhea, *Neisseria gonorrhoeae*, was described by Neisser in 1879 and first cultivated in 1882 by Leistikov and Loeffler. *Neisseria gonorrhoeae* is a fastidious Gram-negative diplococcus which closely resembles the related human pathogen *Neisseria meningitidis*, as well as several commensal species. All *Neisseria* species reside on the mucous membranes of mammals. Humans are the only host of the gonococcus, and spread of the organism occurs only through direct person-to-person contact.

Protein 1 (PI), encoded by the *porB* locus, is the major outer membrane protein of *Neisseria*. It functions as an anion-selective porin allowing the passage of small molecules through the outer membrane. The general structure of these porins consists of nine internal conserved regions separated by eight surface-exposed regions that are highly variable in both amino acid sequence and length (Carbonetti and Sparling 1987; Carbonetti et al. 1988; van der Ley et al. 1991; Feavers et al. 1992; Mee et al. 1993) (see fig 1). Protein 1 is constitutively expressed at high levels in all gonococci, is surface-exposed, and elicits a strong immune reaction during infection (Ison 1988). This indicates that PI may play a crucial role in gonococcal interaction with host cells and, in general, with the immune system (Butt,

Lambden, and Heckels 1990; Smith, Maynard Smith, and Spratt 1995), affecting the transmission probability and the length of the infectious state and consequently influencing the growth rate. It is not surprising, then, that PI is considered a potential vaccine target (Elkins et al. 1992; Heckels, Virji, and Tinsley 1990).

Alleles of this locus in *N. gonorrhoeae* have been assigned to two homology groups based on close sequence and immunological relationships and are designated as either PIA or PIB (Carbonetti et al. 1988). These two homology groups differ in molecular weight, susceptibility to proteolysis, and antigenic reactivity (Sandstrom, Chen, and Buchanan 1982). Alleles within each group are much more similar to each other than they are to members of the other group (i.e., PIA and PIB form distinct monophyletic groups which predate the splitting of species within the genus *Neisseria*), and individual *N. gonorrhoeae* strains express either a PIA or a PIB allele (Smith, Maynard Smith, and Spratt 1995). Although PIA/PIB recombinants can be constructed in the lab (Danielsson et al. 1986), such recombinants are virtually never found in natural isolates (Knapp et al. 1984). While it is likely that PI is subject to immunological pressure to undergo variation, the timescale of such variation is unknown (Cooke et al. 1997). The high nucleotide diversity of PIA alleles has been explained by Smith, Maynard Smith, and Spratt (1995) in terms of positive Darwinian selection acting primarily on the surface-exposed loop regions. On the other hand, it is not known which selective forces operate on the evolution of the PIB alleles. Furthermore, the relative frequencies of these two allelic classes differ greatly, with the PIB alleles showing a higher frequency for a given population. This indicates that different selection pressures may indeed be playing a role in the

Key words: *Neisseria gonorrhoeae*, *porB*, recombination, selection, genetic diversity, population genetics.

Address for correspondence and reprints: Keith A. Crandall, Department of Zoology, 574 Widtsoe Building, Brigham Young University, Provo, Utah 84602-5255. E-mail: keith\_crandall@byu.edu.

*Mol. Biol. Evol.* 17(3):423–436. 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

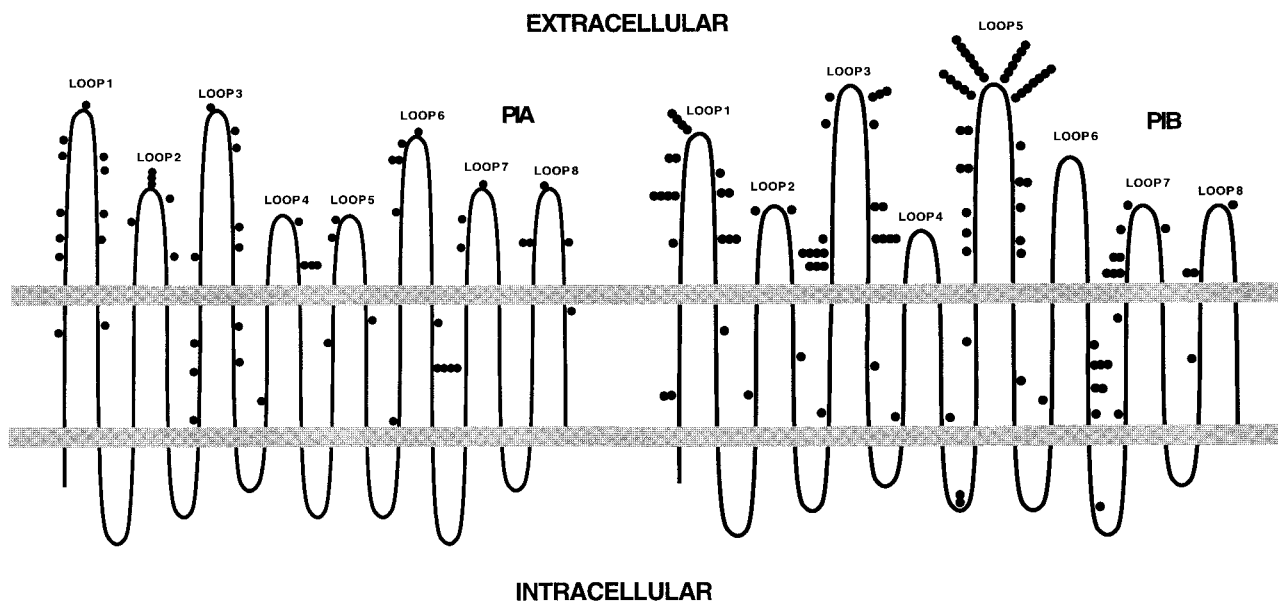


FIG. 1.—Amino acid replacement distribution on the *Neisseria gonorrhoeae* porin structure proposed by van der Ley et al. (1991). Each point represents one replacement.

maintenance of the divergence between PIA and PIB allelic classes.

The characterization of the molecular evolution of the PIA and PIB groups can offer great insights into the evolutionary processes acting in *N. gonorrhoeae* and into their interconnections with epidemiology. Knowledge of the population genetic structure is important for an understanding of the responses of pathogen populations to selective pressures imposed by host immunity and by antimicrobial drug therapy (Levin, Lipsitch, and Bonhoeffer 1999). Smith, Maynard Smith, and Spratt (1995) previously investigated the action of positive Darwinian selection on the evolution of the PIA and PIB alleles. They found, using only two sequences from PIA and two from PIB, that these homology groups showed different levels of selection, with the PIA sequences having an increased number of nonsynonymous substitutions compared with the PIB sequences. They also found these changes to be localized on the surface-exposed loops of the outer membrane of the protein. However, their study was based on a very limited number of sequences (four), and their estimate of nonsynonymous-to-synonymous substitution was based on a biased estimator (Crandall et al. 1999; Yang and Nielsen 2000). Finally, their study, due to the small sample sizes, did not include any estimates of important population genetic parameters such as nucleotide diversity and recombination rate, thus limiting their ability to explain the differences between the PIA and PIB homology groups. We are interested in simultaneously estimating several distinct population parameters, such as genetic diversity, gene flow, population structure, recombination rate, and growth rate, to describe the evolution of the *porB* gene from a multivariable perspective. Addressing the questions of whether there are differences in the molecular evolution of the PIA and PIB homology groups and whether epidemiological differences in PIA and PIB are

associated with changes in different population genetic parameters is the main goal of this work.

## Materials and Methods

### Source of *N. gonorrhoeae* Strains

Fifty-nine strains of *N. gonorrhoeae* were obtained from patients attending the Baltimore City Sexually Transmitted Diseases Clinics between 1995 and 1998. Twelve bacterial isolates, six from 1940–1941, two from 1971, and four from 1979, were obtained from the Bacterial Collection of the Walter Reed Army Institute (generously provided by Herman Schneider). All of the isolates came from patients living in the Washington, D.C., area. The *porB* sequences of these strains were compared with those of all *porB* sequences obtained from GenBank for which information on place and date of isolation was available. The PIA data set included 22 sequences, corresponding to four geographic localities (Baltimore, North Carolina, Washington, D.C., and the United Kingdom) and three time periods (the 1940s, the 1980s, and the 1990s) (see [http://bioag.byu.edu/zoology/crandall\\_lab/cranlabpcbs.htm](http://bioag.byu.edu/zoology/crandall_lab/cranlabpcbs.htm) for details on samples used in this study). The PIB data set included 91 sequences, corresponding to five geographic localities (Baltimore, North Carolina, Washington, D.C., the United Kingdom, and Singapore) and four time periods (the 1940s, the 1970s, the 1980s, and the 1990s).

### Sequencing of the *porB* Gene of *N. gonorrhoeae*

We established a polymerase chain reaction (PCR)-based method for sequencing a large fragment of the *porB* gene of *N. gonorrhoeae* starting with genomic DNA extracted from urine specimens or bacterial isolates.

For recovery of genomic DNA from urine, an aliquot of 1.8 ml of urine was centrifuged at  $10,000 \times g$

for 10 min in an Eppendorf microfuge, and the pellet was resuspended in 600  $\mu$ l of Tris-EDTA buffer (pH 7.4). For extraction of genomic DNA from bacterial isolates, gonococcal colonies were scraped off an agar plate of a primary culture and resuspended in 600  $\mu$ l of Tris-EDTA buffer (pH 7.4). After addition of sodium dodecyl sulfate (SDS) and proteinase K to final volumes of 0.5% and 100  $\mu$ g/ml, respectively, the suspension was mixed by inverting repeatedly and incubated for 1 h at 37°C. The lysate was adjusted to 0.7 M NaCl and 1% cetyltrimethylammonium bromide (CTAB), mixed thoroughly, and incubated for 10 min at 65°C in order to precipitate cell wall debris, denatured proteins, and polysaccharides. The sample was extracted once with chloroform/isoamyl alcohol (24:1), once with phenol/chloroform/isoamyl alcohol (25:24:1), and then a second time with chloroform/isoamyl alcohol. DNA was precipitated with isopropanol, and the pellet was washed once with 70% ethanol and resuspended in 15  $\mu$ l distilled water.

A seminested PCR was used to amplify two overlapping fragments of the *porB* gene. In the first round, a 2- $\mu$ l aliquot of the crude bacterial DNA was amplified in a 50- $\mu$ l reaction volume containing 200  $\mu$ M dNTPs, 0.5  $\mu$ M primers (POR-01, 5'-CTGACTTTG GCAGCCCTTCCTGTTG-3', nt 179–203 [MS11 strain, accession number M21289] and POR-14, 5'-CAGATAGAAATTTGTGGC GC-3', nt 1214–1195), 2.1 U Expand High Fidelity (Boehringer Mannheim, Indianapolis, Ind.), a mix of *Taq* and *pwo* DNA polymerases, and the manufacturer's recommended buffer with 1.5 mM MgCl<sub>2</sub>. A hot-start PCR protocol was performed using wax beads from PE Applied Biosystems, Inc. (Foster City, Calif.). Cycle conditions were 94°C for 2 min to melt the beads, then 30 cycles of 94°C for 40 s, 65°C for 40 s, and 72°C for 1 min, then a final extension reaction for 10 min at 72°C. First- and second-round reactions were performed in a 9700 thermal cycler (PE Applied Biosystems). The PCR products were diluted 1:100 in distilled water, and two nested reactions were performed using primers with 5' extensions encoding the M13 forward or M13 reverse sequencing primers. A 2- $\mu$ l aliquot was amplified in a 100- $\mu$ l reaction volume containing 200  $\mu$ M of dNTPs, 0.5  $\mu$ M primers (M13F-POR-01: 5'-GTCACGACGTTGTA<sup>CGA</sup>CGGCCAGTCTGACTTTGGCAGCCCTT-3' [M13F sequence is underlined] and M13R-POR-08: 5'-CACACAGGAAACAGCTATGACCGT ATTGTGCGAAGAAGC-3', nt 742–726 [M13R sequence is underlined]), or M13F-POR-11: 5'-GTCACGACGTTGTA<sup>CGA</sup>CGGCCAGTCTGACTTTGGCAGCCCTT AGTCTGTCCGTACGCTACG-3', nt 602–617, and M13R-POR14: 5'-CACACAGGAAACAGCTATGACCGA GATTA GAATTTGTGGC GC-3'), 1.4 U Expand High Fidelity enzyme mix, and the manufacturer's recommended buffer with 1.5 mM MgCl<sub>2</sub>. A hot-start PCR protocol was used with cycle conditions of 94°C for 2 min, followed by 35 cycles of 94°C for 40 s, 60°C for 20 s, and 68°C for 40 s, and a final extension reaction for 10 min at 72°C.

The PCR products were passed through a Gene-Clean spin column (Bio 101, Inc., Vista, Calif.) and eluted with 40  $\mu$ l dH<sub>2</sub>O, and the recovered DNA was measured by UV spectrometry. For dideoxy sequencing reactions, 60–80 ng of PCR product was added to a final reaction volume of 5  $\mu$ l containing 2  $\mu$ l of Big Dye Terminator RR mix (PE Applied Biosystems) and 1  $\mu$ M primer (M13 forward or M13 reverse sequencing primer). Cycle conditions were 95°C for 15 s, 50°C for 15 s, and 60°C for 4 min. After 25 cycles, reaction products were denatured by heating to 95°C for 30 s. The reaction volume was diluted with 15  $\mu$ l of distilled water and passed over a minicolumn (Spin-50, BioMax, Inc., Odenton, Mo.) equilibrated with distilled water. The labeled nucleic acids were dried in a Speed Vac concentrator (Savant, Farmingdale, N.Y.), resuspended in 7  $\mu$ l of loading buffer, and loaded into lanes of an ABI 377 automated DNA sequencer (Synthesis and Sequencing Facility, Department of Biological Chemistry, Johns Hopkins University School of Medicine). Trace data were edited and nucleotide sequences assembled with the SeqMan software program (DNASTAR, Inc., Madison, Wis.). The edited sequences typically spanned the region corresponding to amino acids 18–346 of the MS11 strain.

#### Analysis of *N. gonorrhoeae porB* Gene Sequences

Sequences were aligned using CLUSTAL X (Thompson et al. 1997). Alignments were then adjusted by eye as needed (the PIA sequences had no indels, and the PIB sequences had only a single region of 18 nt of questionable alignment that was removed from the analysis). The best-fit model of DNA substitution and the parameter estimates used for the tree reconstruction were chosen by performing hierarchical likelihood ratio tests (see Huelsenbeck and Crandall 1997) using PAUP\* beta 1 (Swofford 1998) and Modeltest 1.05 (Posada and Crandall 1998). A neighbor-joining tree (Saitou and Nei 1987) was estimated for each data set using PAUP\* beta 1 incorporating the best-fit maximum-likelihood model of evolution. Confidence in the tree relationships was assessed using 1,000 replicates of the bootstrap procedure (Felsenstein 1985).

Since recombination can affect the phylogenetic estimate of relationships among the *porB* sequences, we tested each data set for evidence of recombination using the likelihood approach of Grassly and Holmes (1997). Several genetic population parameters were also estimated. The recombination parameter  $C$  ( $=2N_{ei}c$ , where  $N_{ei}$  is the inbreeding effective population size and  $c$  is the recombination rate per site per generation) (Hey and Wakeley 1997; Hudson 1987) was estimated using a coalescent approach implemented in SITES (Hey and Wakeley 1997). Genetic diversity,  $\theta$  ( $=2N_{ei}\mu$ , where  $\mu$  is the mutation rate per site per generation), was estimated using the program FLUCTUATE (Kuhner, Yamato, and Felsenstein 1998). This coalescent-based method uses genealogical information and allows for variable population sizes when estimating genetic diversity (Kuhner, Yamato, and Felsenstein 1998).

We tested for genetic differentiation of populations using a categorical approach and a quantitative approach. The categorical analysis consisted of a  $\chi^2$  test of sequence absolute frequencies at each location (Hudson, Boos, and Kaplan 1992). Due to bias with low expected values, the  $P$  value was obtained by simulating the null distribution of no geographic subdivision (10,000 permutations) using the algorithm of Roff and Bentzen (1989) implemented in the program CHIPERM (D. Posada, available at <http://bioag.byu.edu/zoology/crandalllab/programs.htm>). The quantitative analysis consisted of a molecular analysis of variance (AMOVA) (Excoffier, Smouse, and Quattro 1992) performed using ARLEQUIN (Schneider et al. 1997). Gene flow can be easily estimated for recombining sequences by measuring  $F_{ST}$  (Hudson, Slatkin, and Maddison 1992) and using the standard relationship  $F_{ST} = 1/(1 + 2N_{ei}m)$  (Wright 1951) to obtain  $N_{ei}m$ , where  $m$  is the migration rate per generation.  $F$  statistics and  $N_{ei}m$  values were estimated using the program ARLEQUIN.

Finally, to infer the extent of selection in the PIA and PIB homology groups, we estimated the changes in nonsynonymous substitution rates (those resulting in an amino acid replacement) and synonymous substitution rates (those causing no change in the amino acid). Since the majority of nonsynonymous substitutions are eliminated by purifying selection, neutral evolution predicts a predominance of synonymous substitutions (Miyata and Yasunaga 1980). When positive Darwinian selection occurs, the nonsynonymous rate of substitution accelerates (Hughes and Nei 1988; Messier and Stewart 1997). Therefore, the relative rates of nonsynonymous to synonymous substitutions can be good indicators of the amount and types of selection affecting a gene (Sharp 1997). We estimated the rates of synonymous substitutions ( $K_s$ ) and nonsynonymous substitutions ( $K_a$ ) using the maximum-likelihood approach of Muse and Gaut (1994; Muse 1996). The maximum-likelihood estimates avoid many of the problems associated with estimates based on pairwise comparisons and allow the incorporation of more complex and realistic models of evolution (Nielsen and Yang 1998; Crandall et al. 1999).

## Results

DNA sequences collected by our group have been deposited in GenBank under accession numbers AF200745–AF200814. Combined with additional sequences from GenBank, our resulting data sets contained 22 sequences of 908 nt for the PIA homology group from four distinct geographic locations and 91 sequences of 956 nt for the PIB group representing five geographic locations. This alignment can be obtained in Nexus format at our website: <http://bioag.byu.edu/zoology/crandalllab/cranlabpubs.htm>.

The best-fit model of evolution for the PIA homology group was the Kimura two-parameter model (K80 or K2P) (Kimura 1980), with a transition/transversion (ti/tv) ratio of 2.1451, a significant proportion of invariable sites ( $I = 0.8061$ ), and rate heterogeneity among sites ( $G = 0.8619$ ). For the PIB group, the best-

fit model was HKY (Hasegawa, Kishino, and Yano 1985), with a ti/tv ratio of 2.2891, a significant proportion of invariable sites ( $I = 0.8313$ ), and rate heterogeneity among sites ( $G = 1.0973$ ). Thus, the optimization of a model of evolution for the two homology groups resulted in different models for each group. The major difference in models between the PIA and PIB groups was the incorporation of nucleotide frequency differences for the PIB sequences. We failed to reject the null hypothesis of equal base frequencies for the PIA sequences (table 1), whereas the PIB sequences have significantly different base frequencies ( $A = 0.27$ ,  $C = 0.28$ ,  $G = 0.24$ , and  $T = 0.21$ ). Also, the molecular-clock hypothesis was rejected for both data sets (table 1). The neighbor-joining trees estimated using these models are shown in figures 2 and 3. For both PIA (fig. 2) and PIB (fig. 3) phylogenies, the sequences did not cluster according to sampling time or locality. Bootstrap support was very low across the trees.

Using Grassly and Holmes' (1997) likelihood method for detecting recombination, we concluded that there were three noncontiguous significant recombinant fragments spanning 20 bp in the PIA sequences and six noncontiguous fragments spanning 112 bp in the PIB sequences. We cut these fragments out of the alignment and performed new phylogenetic analyses. The exclusion of these fragments changed the best-fit models for the data sets. For the PIA edited data set, the best-fit model was K80+G, while for the edited PIB data set, the best-fit model was the GTR+I+G (e.g., Rodríguez et al. 1990). The neighbor-joining trees (figs. 4 and 5) obtained after removing the potential recombinant fragments were significantly different from those obtained with the complete data (Kishino-Hasegawa test;  $P = 0.0094$  for PIA,  $P < 0.0001$  for PIB). Again in these new trees, we observed no geographical or temporal phylogenetic structure, and again bootstrap values were low. Excluding the regions associated with recombination leads to the discovery of greater differences in the evolutionary dynamics between these two homology groups. Now, they differ not only in nucleotide frequencies, but also in transition rates, in transversion rates, and in the proportions of invariable sites.

There are two standard approaches for measuring recombination rates. Both of these approaches attempt to estimate a population recombination parameter  $C$  ( $=2N_{ei}c$ ). The first method is based on the variance of the number of base pair differences between DNA sequences (Hudson 1987). This estimator has the disadvantage that large data sets are required to obtain accurate estimates. An alternative approach is based on coalescent theory (Hey and Wakeley 1997). In simulation studies, this last estimate showed low to moderate bias and gave reliable estimates of recombination rate, even for small data sets (Hey and Wakeley 1997). These approaches have not been widely used, especially in the context of infectious diseases. Since recombination rate plays a dominant role in the molecular evolution of genes undergoing moderate recombination, it becomes an essential parameter to estimate. This parameter affects, among other things, the distribution of linkage dis-

**Table 1**  
**Likelihood Ratio Tests of Models of Molecular Evolution for the PIA and PIB Sequences of *Neisseria gonorrhoeae***

Null Hypothesis	Models Compared <sup>a</sup>	-Ln Likelihoods	2(ln L <sub>1</sub> - ln L <sub>0</sub> )	df <sup>b</sup>	P <sup>c</sup>
<b>PIA</b>					
Equal base frequencies	H <sub>0</sub> : JC69	-ln L <sub>0</sub> : 1,914.1533	7.9472	3	0.047114
	H <sub>1</sub> : F81	-ln L <sub>1</sub> : 1,910.1797			
Equal ti/tv <sup>d</sup> rates	H <sub>0</sub> : JC69	-ln L <sub>0</sub> : 1,914.1533	38.9444	1	<0.000001
	H <sub>1</sub> : K80	-ln L <sub>1</sub> : 1,894.6811			
Equal ti and equal tv rates	H <sub>0</sub> : K80	-ln L <sub>0</sub> : 1,894.6811	10.5004	4	0.032790
	H <sub>1</sub> : SYM	-ln L <sub>1</sub> : 1,889.4309			
Equal rates among sites	H <sub>0</sub> : K80	-ln L <sub>0</sub> : 1,889.4309	54.7184	1	<0.000001
	H <sub>1</sub> : K80 + G	-ln L <sub>1</sub> : 1,862.0717			
Proportion of invariable sites	H <sub>0</sub> : K80 + G	-ln L <sub>0</sub> : 1,862.0717	19.5932	1	<0.000001
	H <sub>1</sub> : K80 + G + I	-ln L <sub>1</sub> : 1,852.2751			
Molecular clock	H <sub>0</sub> : K80 + G + Ic	-ln L <sub>0</sub> : 1,935.6774	166.8046	20	<0.000001
	H <sub>1</sub> : K80 + G + I	-ln L <sub>1</sub> : 1,852.2751			
<b>PIB</b>					
Equal base frequencies	H <sub>0</sub> : JC69	-ln L <sub>0</sub> : 4,335.5069	16.6358	3	0.000840
	H <sub>1</sub> : F81	-ln L <sub>1</sub> : 4,327.1890			
Equal ti/tv <sup>d</sup> rates	H <sub>0</sub> : F81	-ln L <sub>0</sub> : 4,327.1890	202.9406	1	<0.000001
	H <sub>1</sub> : HKY	-ln L <sub>1</sub> : 4,225.7187			
Equal ti and equal tv rates	H <sub>0</sub> : HKY	-ln L <sub>0</sub> : 4,225.7187	9.0770	4	0.059200
	H <sub>1</sub> : GTR	-ln L <sub>1</sub> : 4,221.1802			
Equal rates among sites	H <sub>0</sub> : HKY	-ln L <sub>0</sub> : 4,225.7187	954.1344	1	<0.000001
	H <sub>1</sub> : HKY + G	-ln L <sub>1</sub> : 3,748.6515			
Proportion of invariable sites	H <sub>0</sub> : HKY + G	-ln L <sub>0</sub> : 3,748.6515	136.3400	1	<0.000001
	H <sub>1</sub> : HKY + G + I	-ln L <sub>1</sub> : 3,680.4815			
Molecular clock	H <sub>0</sub> : HKY + G + Ic	-ln L <sub>0</sub> : 3,820.4817	280.0004	89	<0.000001
	H <sub>1</sub> : HKY + G + I	-ln L <sub>1</sub> : 3,680.4815			

<sup>a</sup> JC69 = Jukes-Cantor 69 (Jukes and Cantor 1969); F81 = Felsenstein 81 (Felsenstein 1981); K80 = Kimura 80 or Kimura two-parameter (Kimura 1980); SYM = symmetrical model (Zharkikh 1994); HKY = Hasegawa-Kishino-Yano 85 (Hasegawa, Kishino, and Yano 1985); GTR = general time reversible (Rodríguez et al. 1990); G = gamma; I = invariable sites.

<sup>b</sup> Degrees of freedom (the difference in the number of free parameters between the two models being tested). In the case of the molecular clock,  $df = n - 2$ , with  $n$  being the number of taxa.

<sup>c</sup> Due to the performance of multiple tests (six), the individual significance level of rejection of the null hypothesis should be adjusted via the Bonferroni correction to  $\alpha = 0.01$ .

<sup>d</sup> ti = transitions and tv = transversions.

equilibrium between sites and the variance of the number of segregating sites in samples (Hudson 1987). Using the approach of Hey and Wakeley (1997), we found similar overall recombination rates for the different homology groups and some differences among populations (table 2). Because the variances for these estimates are not defined, a statistical test to compare these values is not straightforward. Genetic diversity ( $\theta$ ) estimates for each population and for the total data set were calculated using a phylogenetic maximum-likelihood approach (table 2). The total variation in genetic diversity is higher for the PIB sequences than for the PIA sequences; however, it seems that the relative contribution of recombination to genetic diversity is greater for the PIA sequences. This greater relative contribution of recombination to genetic diversity may be due to a higher mutation rate in the PIB sequences. Recombination can contribute to genetic diversity by forming new combinations of existing sequence variants, thereby adding diversity to the population. The genetic diversity is clearly higher in the PIB sequences. Since the effective population size cancels out in the  $c/\mu$  statistic, the differences remaining seem to be driven by a faster mutation rate in the PIB (or a higher rate of recombination in PIA). Either way, this difference suggests that different population dynamics are occurring in these two homology

groups, suggesting the evolution of different pathogenic strategies.

Gene flow is another important parameter for understanding the evolution of *N. gonorrhoeae*. The estimated  $N_{ei}m$  values are shown in table 3. For the PIA sequences,  $N_{ei}m$  values ranged from 0.946 between North Carolina and Washington, D.C., to an incredibly high 14.693 between Washington, D.C., and the United Kingdom. This latter value may be an artifact of small sample sizes ( $n = 3$  for Washington, D.C., and  $n = 8$  for the United Kingdom). The PIB sequences showed low levels of gene flow across continents ( $N_{ei}m = 1.888$  between Washington, D.C., and the United Kingdom and 1.828 between the United Kingdom and Singapore). As expected, the PIB sequences showed extremely high levels of gene flow between Baltimore and Washington, D.C. ( $N_{ei}m = 31.820$ ), and moderately high levels between Baltimore and Washington, D.C., and North Carolina. The categorical analysis for detecting geographic subdivision was significant for the PIB alleles ( $P < 0.0001$ ), while we could not reject the hypothesis of overall geographic homogeneity for the PIA sequences ( $P = 0.5916$ ), presumably due to the small sample sizes for this homology group. Fixation indices partitioned this variation (table 4), indicating that for the PIB sequences, the distribution of haplotypes among popula-

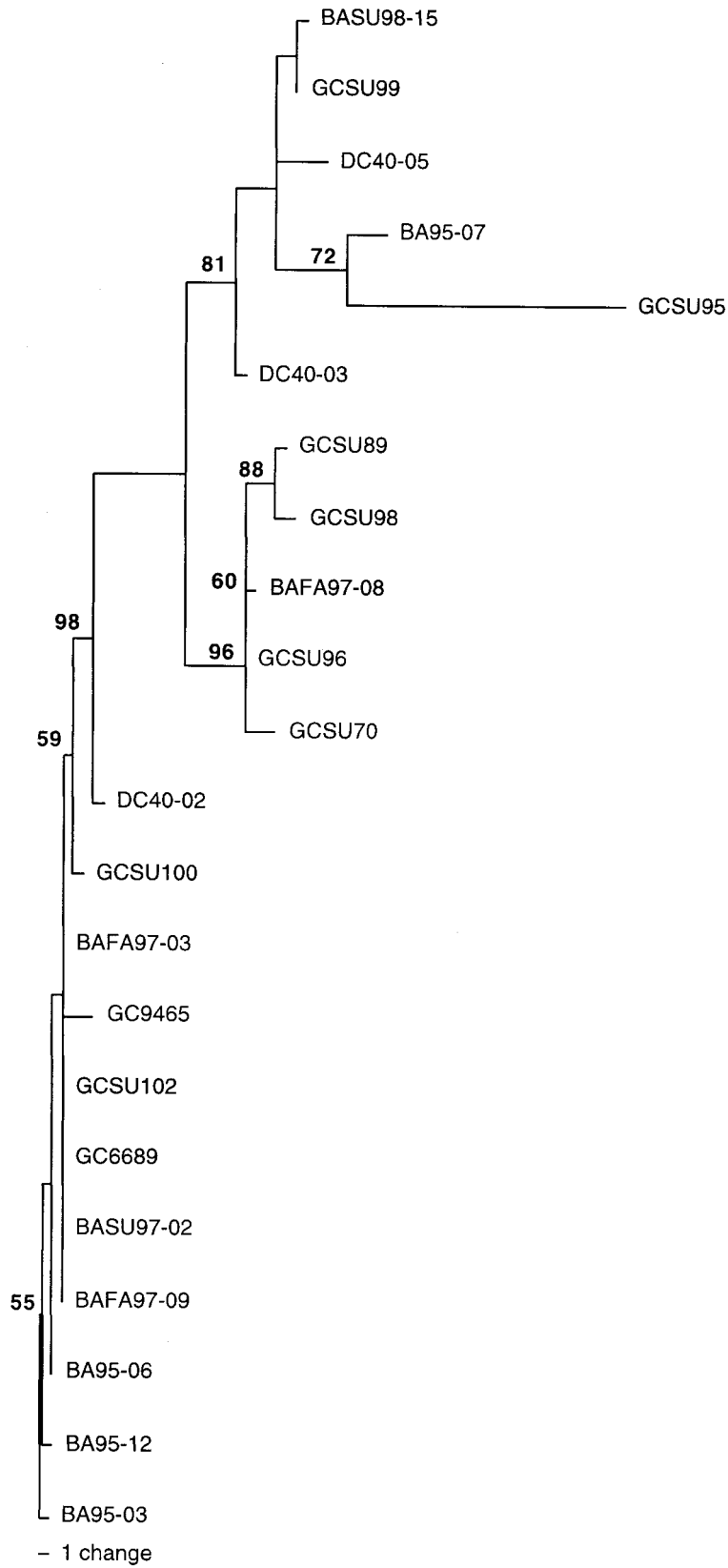


FIG. 2.—Neighbor-joining tree of the PIA sequences using a K2P model (Kimura 1980) (transition/transversion ratio = 2.145, I = 0.8061, and G = 0.8619). Bootstrap values are based on 1,000 replications of the bootstrap procedure. Only bootstrap values over 50% are shown. Branch lengths are shown proportional to the amount of evolutionary change.

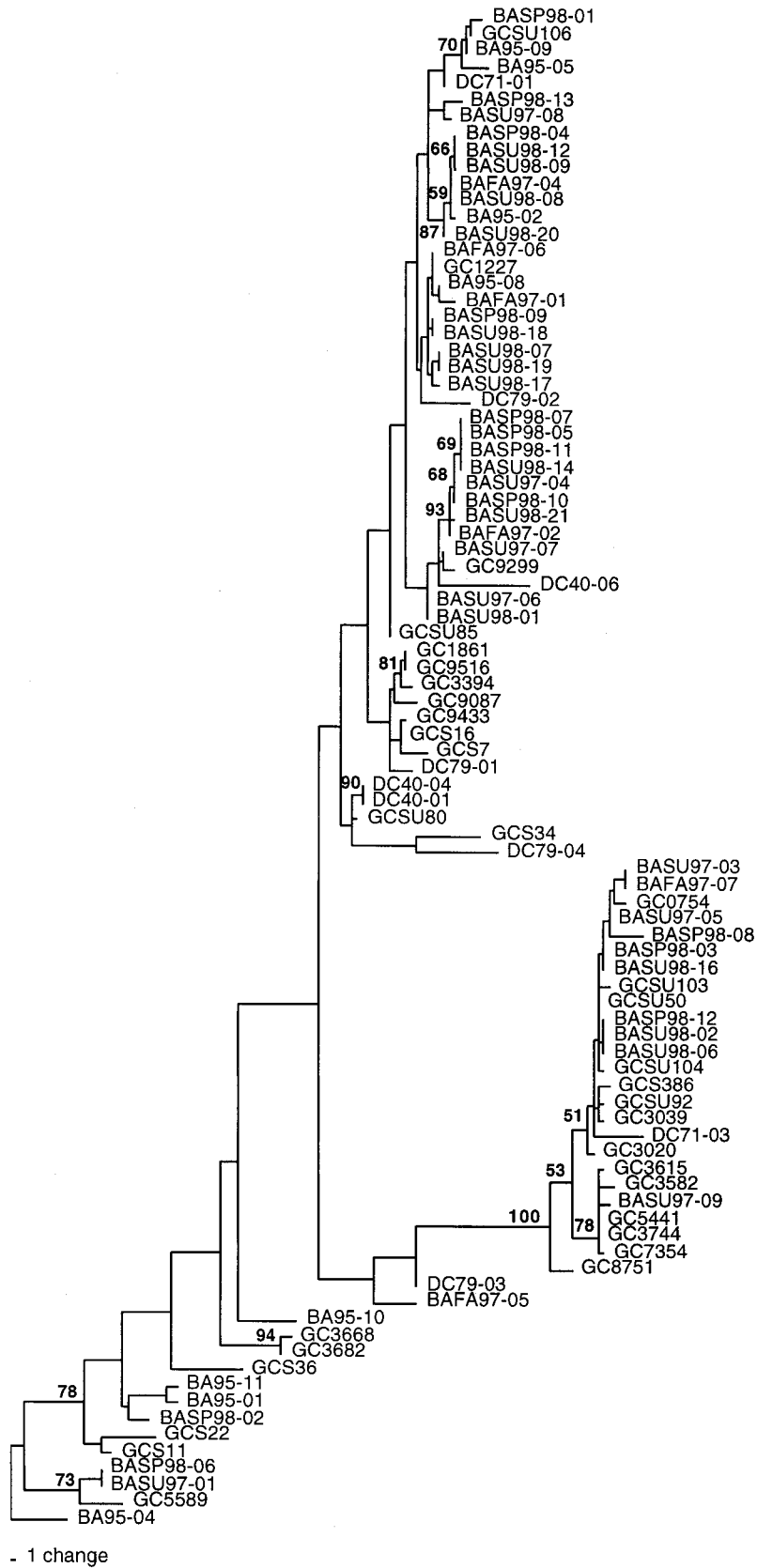


FIG. 3.—Neighbor-joining tree of the PIB sequences using an HKY model (Hasegawa, Kishino, and Yano 1985) (transition/transversion ratio = 2.2891, I = 0.8313, and G = 1.0973). Bootstrap values are based on 1,000 replications of the bootstrap procedure. Only bootstrap values over 50% are shown. Branch lengths are shown proportional to the amount of evolutionary change.

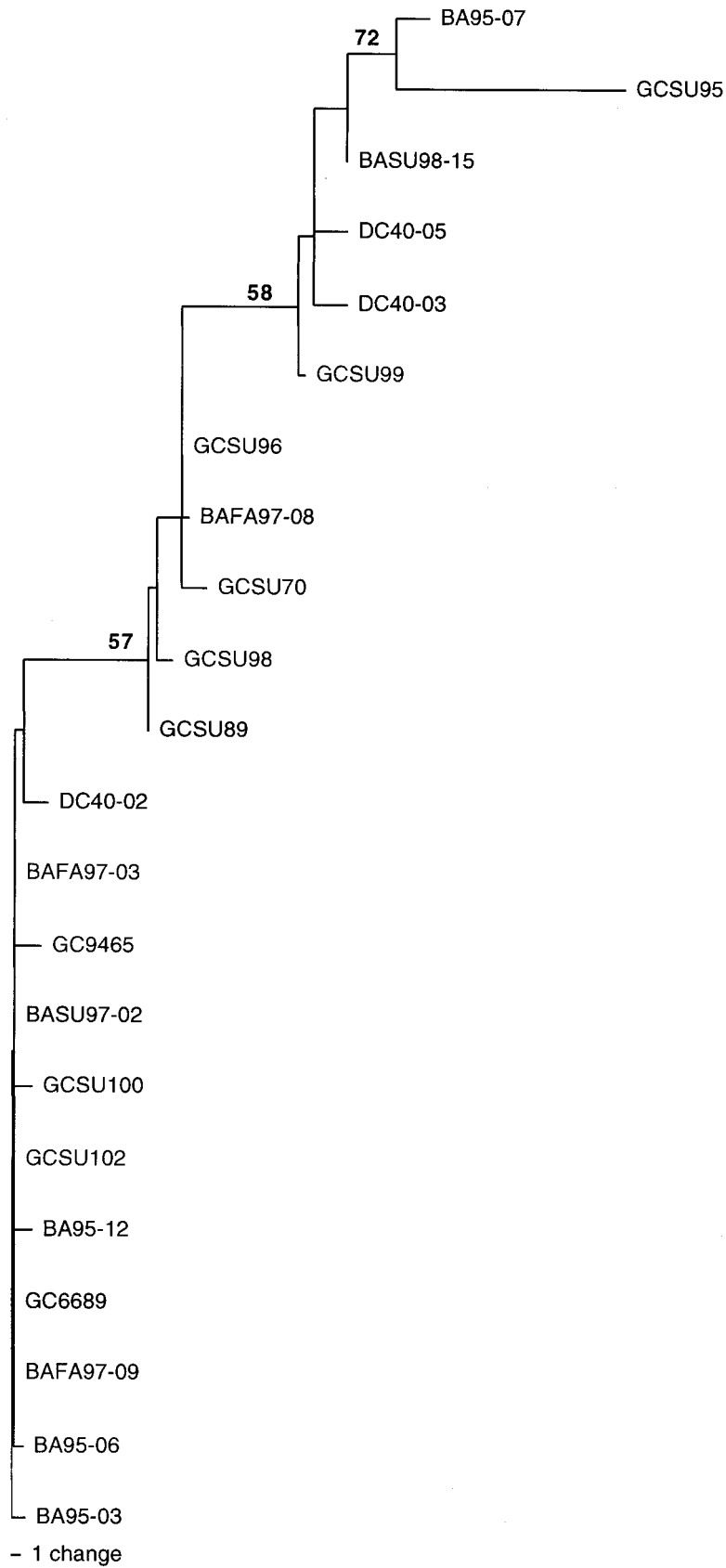


FIG. 4.—Neighbor-joining tree of the PIA sequences after removing three potential recombinant fragments expanding 20 bp. The model used was the K2P model (Kimura 1980) (transition/transversion ratio = 2.5403, G = 0.0067). Bootstrap values are based on 1,000 replications of the bootstrap procedure. Only bootstrap values over 50% are shown. Branch lengths are shown proportional to the amount of evolutionary change.

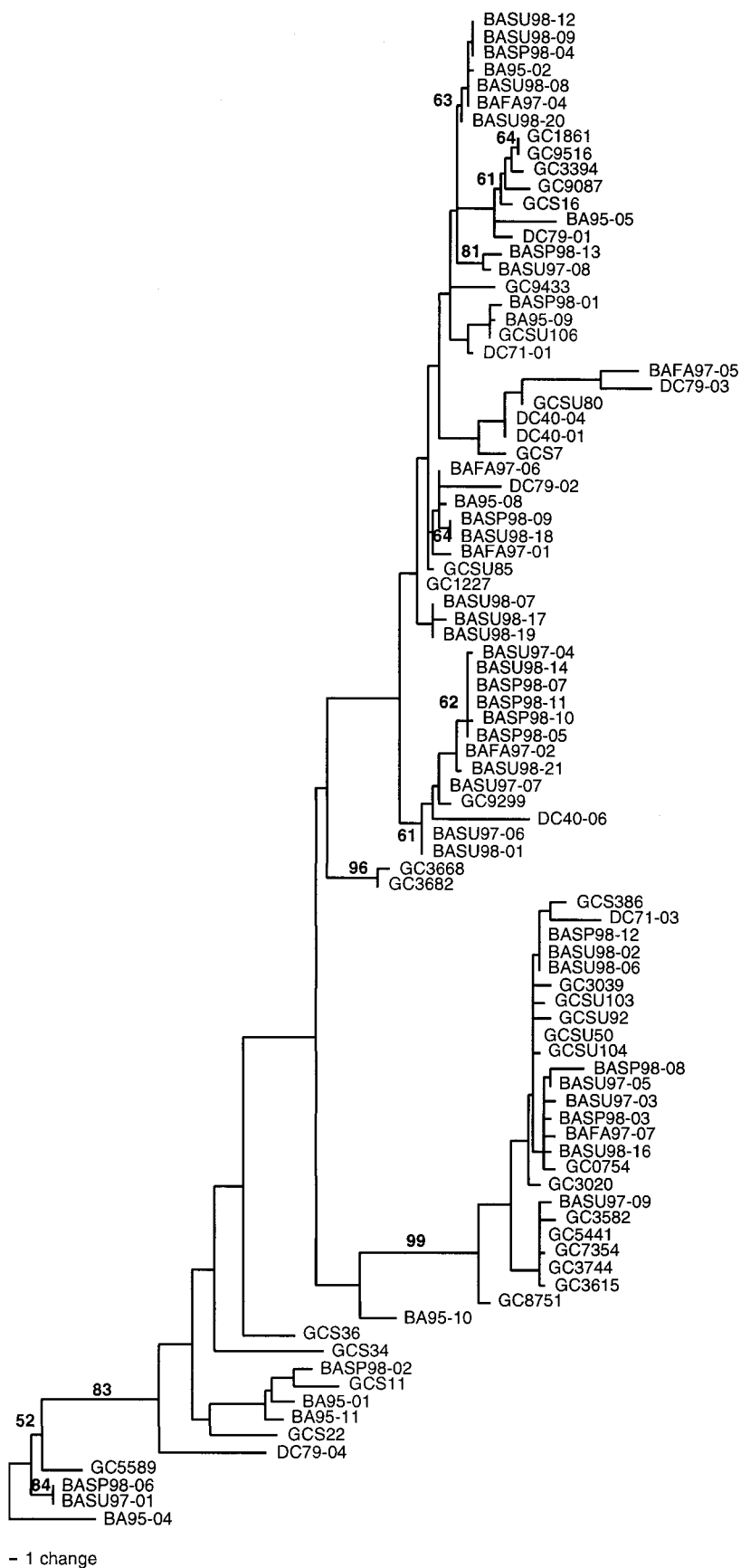


FIG. 5.—Neighbor-joining tree of the PIB sequences after removing six potential recombinant fragments expanding 112 bp. The model used was the GTR model (e.g., Rodriguez et al. 1990) ( $a = 0.6429$ ,  $b = 3.1401$ ,  $c = 0.3458$ ,  $d = 0.9258$ ,  $e = 7.0118$ ,  $f = 1.0000$ ;  $I = 0.7728$ , and  $G = 0.2933$ ). Bootstrap values are based on 1,000 replications of the bootstrap procedure. Only bootstrap values over 50% are shown. Branch lengths are shown proportional to the amount of evolutionary change.

**Table 2**  
**Estimates of Recombination, Genetic Diversity, and Nonsynonymous/Synonymous Substitution Rates for the PIA and PIB Sequences of *Neisseria gonorrhoeae***

Population	<i>N</i>	<i>C</i>	$\theta$	<i>c</i> / $\mu$	<i>K<sub>a</sub></i> / <i>K<sub>s</sub></i>
PIA	22	0.0327	0.0514	0.6362	2.3832
Baltimore	9	0.0224	0.0278	0.8057	4.1756
Washington, D.C.	3	*	*	*	*
North Carolina	2	*	*	*	*
United Kingdom	8	0.0614	0.0426	1.4413	2.2246
PIB	91	0.0583	0.4480	0.1301	0.6533
Baltimore	49	0.0462	0.0923	0.5005	0.5864
Washington, D.C.	9	0.0809	0.3353	0.2413	0.5415
North Carolina	13	0.0308	0.0550	0.5600	0.6655
United Kingdom	13	0.0314	0.0451	0.6962	0.7645
Singapore	7	0.0362	0.3164	0.1144	0.5973

NOTE.—*C* is the recombination parameter (Hey and Wakeley 1997).  $\theta$  is the estimate of genetic diversity (Kuhner, Yamato, and Felsenstein 1998). The ratio of the recombination rate to the mutation rate (*c*/ $\mu$ ) was obtained by dividing *C*/ $\theta$ . *K<sub>a</sub>*/*K<sub>s</sub>* is the nonsynonymous/synonymous substitution rate ratio (Muse 1996). Asterisks indicate small samples.

tions among continents ( $F_{ST}$ ) is not random, while for both groups there is no geographical association of the variation of haplotypes among populations within continents ( $F_{SC}$ ) or among continents ( $F_{CT}$ ). The analysis of molecular variance shows (table 5) that most of the variation for the PIA and PIB groups is located within populations (88% and 86% for PIA and PIB, respectively), while some of the variation corresponds among continents (8% and 11% for PIA and PIB, respectively). Of the total variation, only 4% and 3% corresponds to differences among populations within a continent for PIA and PIB, respectively.

The ratios of nonsynonymous sites to synonymous sites were 59/10 for the PIA sequences and 96/37 for the PIB sequences. For the PIA sequences, the average number of nonsynonymous sites was 704.39, while the average number of synonymous sites was 204.61. For the PIB sequences, the average number of nonsynonymous sites was 736.12, while the average number of synonymous sites was 219.56. The average *K<sub>a</sub>*/*K<sub>s</sub>* was 2.38 for the PIA and 0.59 for the PIB sequences (see table 2). Clearly, these two homology groups are under different selection pressures, presumably due to different selective constraints on them. We partitioned all the amino acid replacements on the different porin regions (external loops or internal regions) for better visualization of their distribution (fig. 1 and table 6). The distributions of amino acid replacements were not significantly different among the PIA and PIB sequences (two-way ANOVA using the length of the region as a covariate;  $P = 0.242$ ). The distributions of the replacements among internal or external regions were significantly different for both groups, accumulating more replacements in the exposed regions than in the internal regions (ANOVA;  $P$  values were 0.003 and 0.013 for the PIA and PIB sequences, respectively). Since the differences do not appear to be correlated with structural differences in the protein, we hypothesize that the selective differences between PIA and PIB have to do more with their epide-

**Table 3**  
**Gene Flow ( $N_e m$ ) Estimates for the PIA Sequences (above diagonal) and for the PIB Sequences (below diagonal)**

	Baltimore, Md.	Washing- ton, D.C.	North Carolina	United Kingdom
Baltimore, Md.	—	2.277	Infinite	5.058
Washington, D.C.	31.820	—	0.946	14.693
North Carolina	12.420	13.822	—	3.928
United Kingdom	2.867	1.888	7.603	—
Singapore	3.142	4.739	3.825	1.828

miology. We discuss this in more detail in the next section.

## Discussion

*Neisseria gonorrhoeae* has been proposed to have a nonclonal population structure, being effectively panmictic (Maynard Smith et al. 1993; O'Rourke and Stevens 1993). This is due mainly to recombination, presumably mediated by genetic transformation (O'Rourke et al. 1995), which has been shown to be extensive in *Neisseria* and in the *porB* gene (Feavers et al. 1992; Vázquez et al. 1993; Spratt et al. 1995; Hobbs et al. 1999) to the point of resulting in linkage equilibrium within populations (Vázquez et al. 1993). Recombination in bacterial sequences makes them particularly difficult to study from a phylogenetic perspective. Indeed, development of the first methods to detect recombination was stimulated by the analysis of bacterial gene sequences (e.g., Stephens 1985; Sawyer 1989). While recombination occurs in *N. gonorrhoeae* (Spratt et al. 1995), some phylogenetic signal still remains. Indeed, it is through the discordance of phylogenies from different gene regions that researchers have concluded recombination occurs (Spratt et al. 1995).

Researchers of infectious diseases typically use a Kimura two-parameter model (K80 or K2P) (Kimura 1980) to model molecular changes. The K80 model only accounts for differences between transitions and transversions. However, there are other parameters worth considering, e.g., nucleotide frequency differences, rate heterogeneity, etc. Therefore, an alternative model of evolution which takes into account more parameters may be more appropriate for a given data set. Indeed, the resulting topology and conclusions based on that to-

**Table 4**  
**Fixation Indices for the PIA and PIB Sequences**

Index	Value	<i>P</i>
PIA		
$F_{ST}$	0.11749	0.11535
$F_{SC}$	0.04362	0.10362
$F_{CT}$	0.07723	0.24340
PIB		
$F_{ST}$	0.14090	0.00391
$F_{SC}$	0.02991	0.20039
$F_{CT}$	0.11441	0.09384

NOTE.— $P$  values were calculated by permuting the data matrix 1,023 times.

**Table 5**  
**Analysis of Molecular Variance (AMOVA) for the PIA and PIB Sequences**

Source of Variation	df	Sum of Squares	Variance Components	Percentage of Variation
<b>PIA</b>				
Among continents	1	15.817	0.61943	7.72
Among populations within continents	2	16.508	0.32283	4.03
Within populations	18	127.403	7.07793	88.25
Total	21	159.727	8.02019	
<b>PIB</b>				
Among continents	2	112.867	2.08233	11.44
Among populations within continents	2	47.500	0.48211	2.65
Within populations	86	1,373.495	15.63590	85.91
Total	90	1,505.055	18.20034	

NOTE.—PIA—group 1: Baltimore ( $n = 9$ ), Washington, D.C. ( $n = 3$ ), and North Carolina ( $n = 2$ ); group 2: United Kingdom ( $n = 8$ ). PIB—group 1: Baltimore, Washington, D.C., and North Carolina; group 2: Singapore; Group 3: United Kingdom.

pology can be greatly influenced by the choice of model of evolution (Kelsey, Crandall, and Voevodin 1999). Thus, it is important to optimize models of evolution for particular data sets to infer phylogenies and then to use these phylogenies to test for recombination. As far as the detection of recombination depends on the estimated topologies, the use of a correct model of evolution will make its inference more reliable.

We have developed a hypothesis-testing framework to justify the choice of a model of evolution (Huelsenbeck and Crandall 1997). We have also developed software (Modeltest, freely distributed at our web site: [http://bioag.byu.edu/zoology/crandall\\_lab/programs.htm](http://bioag.byu.edu/zoology/crandall_lab/programs.htm)) that uses likelihood ratio tests to determine the model that best fits the data at hand (Posada and Crandall 1998). Once a model of evolution is chosen, phylogenetic relationships among sequences can be estimated using either the neighbor-joining algorithm (Saitou and Nei 1987) or the maximum-likelihood criterion (Felsenstein 1981). Our data sets were too large for maximum-likelihood analyses because of the computational expense of this method with large numbers of sequences. Therefore, we used the neighbor-joining method to estimate phylogenetic relationships. The PIA and PIB sequences show little phylogenetic structure due to geography or date of isolation, in concordance with previous results (Smith, Maynard Smith, and Spratt 1995). This lack of phylogenetic structure is likely to be the consequence of extensive recombination in the *porB* gene, since we have reduced the possibility of error due to inappropriate models of evolution. Even after removing recombinant fragments that were detected by the method of Grassly and Holmes (1997), we did not observe geographic or temporal phylogenetic structure. This is not surprising, because most of the recombination is probably still undetected by their method. Further progress in phylogenetic analysis of *N. gonorrhoeae* will require the development of improved methods to detect recombination or to deal with recombination in phylogenetic reconstructions. These are both areas of research that our group is actively pursuing.

The results of our population genetic analysis clearly reinforce the idea that recombination is extensive in *N. gonorrhoeae*. We also show that recombination appears to be similar in both the PIA and the PIB homol-

ogy groups. However, these groups differ greatly in their levels of genetic diversity. The diversity generated by point mutation is reflected by the estimates of the parameter  $\theta$  ( $2N_{ei}\mu$ ), while the estimates of the parameter  $C$  ( $2N_{ei}c$ ) indicate the diversity generated by recombination. Hence, if we divide one by the other, we will have an estimate of the ratio  $c/\mu$ , which can be interpreted as the relative chance of recombination per site versus the chance of point mutation per site. The high ratios of the recombination rate to the mutation rate ( $c/\mu$ ) for all of the populations (table 2) indicate that recombination is a major force generating diversity in *N. gonorrhoeae*. Again, there is a distinction between the PIA and the PIB groups in that the PIA group has a much higher  $c/\mu$  ratio than does the PIB group, suggesting differences in the roles of recombination and mutation in generating diversity in these two homology groups. This interaction among recombination and mutation and its contribution to the evolution of natural populations has been also described for plant viruses (Aranda et al. 1997).

The  $N_{ei}m$  estimates for the PIA sequences did not make geographic sense, in that those locations in closer geographic proximity showed lower levels of gene flow, most likely because of the small sample sizes of sequences from North Carolina (two) and Washington, D.C. (three). However, for the PIB sequences, closer locations showed higher levels of gene flow, suggesting that geographic distance may be a factor in the spread of genetic diversity among isolates of *N. gonorrhoeae*. This is contradictory to the interpretation by Smith, Maynard Smith, and Spratt (1995) that *N. gonorrhoeae* is a panmictic species. This is also reflected in the moderately high  $F_{ST}$  estimate and in the categorical analysis of geographic association, suggesting that there is geographic subdivision for the PIB sequences, although we probably do not have enough power (i.e., a large enough sample size) to reject geographic homogeneity for the PIA sequences. This subdivision is at the level of the population. As one of the main forces generating diversity is recombination, most of the variation is located within populations. As compared with the phylogenetic analysis, which failed to show structure due to geography, these population genetic analyses suggest that there is population substructure based on geographic locality.

**Table 6**  
**Distribution of Amino Acid Replacements Along the PIA and PIB Proteins**

REGION	PIA			PIB		
	AA Site	MNR	MNR/AA	AA Site	MNR	MNR/AA
Trans 1	1–11	1	0.091	1–11	2	0.182
Trans 2	35–62	1	0.036	35–62	2	0.071
Trans 3	76–99	3	0.125	76–99	2	0.083
Trans 4	124–145	3	0.136	128–149	2	0.091
Trans 5	156–178	1	0.043	159–181	4	0.174
Trans 6	188–210	2	0.154	207–229	2	0.087
Trans 7	228–251	5	0.208	243–266	10	0.417
Trans 8	265–286	0	0.000	280–301	1	0.045
Trans 9	299–302	1	0.250	314–318	0	0
Loop 1	12–34	10	0.434	12–34	17	0.739
Loop 2	63–75	6	0.461	63–75	2	0.154
Loop 3	100–123	6	0.250	100–127	20	0.714
Loop 4	146–155	4	0.400	150–158	0	0
Loop 5	179–187	2	0.222	182–206	39	1.560
Loop 6	211–227	5	0.294	230–242 <sup>a</sup>	0	0
Loop 7	252–264	3	0.230	267–279	8	0.615
Loop 8	287–298	4	0.333	302–313	3	0.250
Total Trans		17	0.094		25	0.137
Total Loop		40	0.330		89	0.654

NOTE.—AA = amino acid; MNR = minimum number of amino acid replacements; Trans = transmembrane region. See also figure 1.

<sup>a</sup> A region of 18 nt of questionable alignment was removed.

Directional positive selection appears to be operating to a significant extent in these sequences. This is not a surprise given the intensive selection pressures put on these bacteria by the immune system and antibiotic treatment (Smith, Maynard Smith, and Spratt 1995). The selective distribution of the replacements in the surface-exposed loops supports this idea. However, selection intensities as a whole appear to differ between the two homology groups. The PIA sequences are clearly evolving as a whole under positive selection, especially in the exposed regions. This supports the earlier conclusions of Smith, Maynard Smith, and Spratt (1995) based on a very limited data set. Given the reduced genetic variation associated with the PIA sequences, this suggests a role for selective sweeps in reducing the amount of genetic variation within a population. Although uncomplicated gonorrhoea is caused by isolates of either PIA or PIB homology groups, there are epidemiological differences among these groups. Blood isolates during disseminated gonococcal infection belong almost invariably to the PIA group (Sandstrom et al. 1984), whereas isolates from mucosal surfaces more often belong to the PIB group (Morse et al. 1982). These epidemiological correlates may provide an explanation for the apparent difference in selection pressures acting on the PIA and PIB sequences. Invasive disease would be expected to subject *N. gonorrhoeae* to more intense selection pressure from the host immune system than would occur during infections confined to mucosal surfaces. The PIB sequences, although they accumulate many replacements on the exposed regions, appear to be subject also to some extent of purifying selection. However, recombination is so high in these sequences that they maintain a great amount of diversity.

### Conclusions

We have shown that the evolution of *N. gonorrhoeae* is driven by the simultaneous action of selection,

mutation, and recombination. Recombination generates diversity, on which selection acts, favoring the differential spread of the fittest alleles, counterforcing the action of recombination. However, this relationship will change in different evolutionary scenarios. The PIA and PIB homology groups are evolving in different manners. The levels of genetic diversity are different between these two groups, with the PIB group showing higher diversity. The relative contribution of recombination to this diversity differs between these two groups, and the type and intensity of natural selection differs between them. Finally, the population structures also appear to differ between these two groups. Given these differences and the difference in the frequencies of the two homology groups in infected populations, there is a suggested difference in pathogenicity between the two groups driven by these differences in population genetic parameters. These differences also suggest an explanation for the differences in demography of the different homology groups. This discovery sets the stage for future work in two distinct areas: (1) in theoretical work modeling the population dynamics of the PIA and PIB groups with differences in these important parameters, and (2) in empirical studies combining analyses of population genetic parameters with more detailed epidemiological data. This approach will allow the investigation of correlations between population genetic parameters that differ between the homology groups and various epidemiological factors. For example, Hobbs et al. (1999) have suggested different transmission dynamics between male and female infected patients. Are these differences seen in both homology groups? How do the population dynamics of PIA and PIB differ by sex of host? Future work by our group will explore in more detail the epidemiological and clinical correlates with genetic diversity, recombination rates, and selection intensity to explain the distribution of variation in *N. gonorrhoeae*.

## Acknowledgments

This project was funded by NIH grant number RO1-HD33982 (R.P.V.) and the Alfred P. Sloan Foundation (K.A.C.).

## LITERATURE CITED

- ARANDA, M. A., A. FRAILE, J. DOPAZO, J. M. MALPICA, and F. GARCÍA-ARENAL. 1997. Contribution of mutation and RNA recombination to the evolution of a plant pathogenic RNA. *J. Mol. Evol.* **44**:81–88.
- BUTT, N. J., P. R. LAMBDEN, and J. E. HECKELS. 1990. The nucleotide sequence of the por gene from *Neisseria gonorrhoeae* strain P9 encoding outer membrane protein PIB. *Nucleic Acids Res.* **18**:4258.
- CARBONETTI, N. H., V. I. SIMNAD, H. S. SEIFERT, M. SO, and P. F. SPARLING. 1988. Genetics protein I of *Neisseria gonorrhoeae*: Construction of hybrid porins. *Genetics* **85**:6841–6845.
- CARBONETTI, N. H., and P. F. SPARLING. 1987. Molecular cloning and characterization of the structural gene for protein I, the major outer membrane protein of *Neisseria gonorrhoeae*. *Genetics* **84**:9084–9088.
- COOKE, S. J., H. DE LA PAZ, C. L. POH, C. A. ISON, and J. E. HECKELS. 1997. Variation within serovars of *Neisseria gonorrhoeae* detected by structural analysis of outer-membrane protein PIB and by pulsed-field gel electrophoresis. *Microbiology* **143**:1415–1422.
- CRANDALL, K. A., C. R. KELSEY, H. IMAMICHI, C. H. LANE, and N. P. SALZMAN. 1999. Parallel evolution of drug resistance in HIV: failure of nonsynonymous/synonymous substitution rate ratio to detect selection. *Mol. Biol. Evol.* **16**:372–382.
- DANIELSSON, D., H. FARUKI, D. DYER, and P. F. SPARLING. 1986. Recombination near the antibiotic resistance locus penB results in antigenic variation of gonococcal outer membrane protein I. *Infect. Immun.* **52**:529–533.
- ELKINS, C., N. H. CARBONETTI, V. A. VARELA, D. STIREWALT, D. G. KLAPPER, and P. F. SPARLING. 1992. Antibodies to N-terminal peptides of gonococcal porin are bactericidal when gonococcal lipopolysaccharide is not sialylated. *Mol. Microbiol.* **6**:2617–2628.
- EXCOFFIER, L., P. E. SMOUSE, and J. M. QUATTRO. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**:479–491.
- FEAVERS, I. M., A. B. HEATH, J. A. BYGRAVES, and M. C. J. MAIDEN. 1992. Role of horizontal genetic exchange in the antigenic variation of the class I outer membrane protein in *Neisseria meningitidis*. *Mol. Microbiol.* **6**:489–495.
- FELSENSTEIN, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**:368–376.
- . 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- 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.
- HANDSFIELD, H., and P. SPARLING. 1995. *Neisseria gonorrhoeae*. Pp. 1909–1926 in G. MANDELL, J. BENNETT, and R. DOLIN, eds. *Principles and practice of infectious diseases*. Churchill Livingstone, Philadelphia.
- HASEGAWA, M., K. KISHINO, and T. YANO. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**:160–174.
- HECKELS, J. E., M. VIRJI, and C. R. TINSLEY. 1990. Vaccination against gonorrhoeae: the potential protective effect of immunization with a synthetic peptide containing a conserved epitope of gonococcal outer membrane protein IB. *Vaccine* **8**:225–230.
- HEY, J., and J. WAKELEY. 1997. A coalescent estimator of the population recombination rate. *Genetics* **145**:833–846.
- HOBBS, M. M., T. M. ALCORN, R. H. DAVIS, W. FISCHER, J. C. THOMAS, I. MARTIN, C. ISON, P. F. SPARLING, and M. S. COHEN. 1999. Molecular typing of *Neisseria gonorrhoeae* causing repeated infections: evolution of porin during passage within a community. *J. Infect. Dis.* **179**:371–381.
- HUDSON, R. R. 1987. Estimating the recombination parameter of a finite population model without selection. *Genet. Res. Camb.* **50**:245–250.
- HUDSON, R. R., D. D. BOOS, and N. L. KAPLAN. 1992. A statistical test for detecting geographic subdivision. *Mol. Biol. Evol.* **9**:138–151.
- HUDSON, R. R., M. SLATKIN, and W. P. MADDISON. 1992. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**:583–589.
- HUELSENBECK, J. P., and K. A. CRANDALL. 1997. Phylogeny estimation and hypothesis testing using maximum likelihood. *Annu. Rev. Ecol. Syst.* **28**:437–466.
- HUGHES, A. L., and M. NEI. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**:167–170.
- ISON, C. A. 1988. Immunology of gonorrhoea. Pp. 95–116 in D. J. M. WRIGHT, ed. *Immunology of sexually transmitted diseases*. Kluwer Academic, London.
- JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 in H. M. MUNRO, ed. *Mammalian protein metabolism*. Academic Press, New York.
- KELSEY, C. R., K. A. CRANDALL, and A. F. VOEVODIN. 1999. Different models, different trees: the geographic origin of PTLV-I. *Mol. Phylogenet. Evol.* **13**:336–347.
- KIMURA, M. 1980. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**:111–120.
- KNAPP, J. S., M. R. RAM, R. C. NOWINSKI, K. K. HOLMES, and E. G. SANDSTROM. 1984. Serological classification of *Neisseria gonorrhoeae* with use of monoclonal antibodies to gonococcal outer membrane protein I. *J. Infect. Dis.* **150**:44–48.
- KUHNER, M. K., J. YAMATO, and J. FELSENSTEIN. 1998. Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics* **149**:429–434.
- LEVIN, B. R., M. LIPSITCH, and S. BONHOEFFER. 1999. Population biology, evolution and infectious disease: convergence and synthesis. *Science* **583**:806–809.
- MAYNARD SMITH, J., N. H. SMITH, M. O'ROURKE, and B. G. SPRATT. 1993. How clonal are bacteria? *Proc. Natl. Acad. Sci. USA* **90**:4384–4388.
- MEE, B. J., H. THOMAS, S. J. COOKE, P. R. LAMBDEN, and J. E. HECKELS. 1993. Structural comparison and epitope analysis of outer-membrane protein PIA from strains of *Neisseria gonorrhoeae* with differing serovar specificities. *J. Gen. Microbiol.* **139**:2613–2620.
- MESSIER, W., and C.-B. STEWART. 1997. Episodic adaptive evolution of primate lysozymes. *Nature* **385**:151–154.
- MIYATA, T., and T. YASUNAGA. 1980. Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitution from homologous nucleotide sequences and its application. *J. Mol. Evol.* **16**:23–36.
- MORSE, S. A., P. G. LYSKO, L. MCFARLAND, J. S. KNAPP, E. SANDSTROM, C. CRITCHLOW, and K. K. HOLMES. 1982.

- Gonococcal strains from homosexual men have outer membranes with reduced permeability to hydrophobic molecules. *Infect. Immun.* **37**:432–438.
- MUSE, S. V. 1996. Estimating synonymous and nonsynonymous substitution rates. *Mol. Biol. Evol.* **13**:105–114.
- MUSE, S. V., and B. S. GAUT. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol. Biol. Evol.* **11**:715–724.
- NIELSEN, R., and Z. YANG. 1998. Likelihood methods for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* **148**:929–936.
- O'ROURKE, M., C. A. ISON, A. M. RENTON, and B. G. SPRATT. 1995. Opa-typing: a high-resolution tool for studying the epidemiology of gonorrhoeae. *Mol. Microbiol.* **17**:865–875.
- O'ROURKE, M., and E. STEVENS. 1993. Genetic structure of *Neisseria gonorrhoeae* populations: a non-clonal pathogen. *J. Gen. Microbiol.* **139**:2603–2611.
- POSADA, D., and K. A. CRANDALL. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
- RODRÍGUEZ, F., J. F. OLIVER, A. MARÍN, and J. R. MEDINA. 1990. The general stochastic model of nucleotide substitutions. *J. Theor. Biol.* **142**:485–501.
- ROFF, D. A., and P. BENTZEN. 1989. The statistical analysis of mitochondrial DNA polymorphisms: chi-square and the problem of small samples. *Mol. Biol. Evol.* **6**:539–545.
- SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- SANDSTROM, E. G., K. C. S. CHEN, and T. M. BUCHANAN. 1982. Serology of *Neisseria gonorrhoeae*: co-agglutination serogroups WI and WII/III correspond to different outer membrane protein molecules. *Infect. Immun.* **38**:462–470.
- SANDSTROM, E. G., J. S. KNAPP, L. B. RELLER, S. E. THOMPSON, E. W. I. HOOK, and K. K. HOLMES. 1984. Serogrouping of *Neisseria gonorrhoeae*: correlation of serogroup with disseminated gonococcal infection. *Sex. Transm. Dis.* **11**:77–80.
- SAWYER, S. 1989. Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **6**:526–538.
- SCHNEIDER, S., J.-M. KUEFFER, D. ROESSLI, and L. EXCOFIER. 1997. Arlequin: a software for population genetic data analysis. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva.
- SHARP, P. M. 1997. In search of molecular Darwinism. *Nature* **385**:111–112.
- SMITH, N. H., J. M. MAYNARD SMITH, and B. G. SPRATT. 1995. Sequence evolution of the porB gene of *Neisseria gonorrhoeae* and *Neisseria meningitidis*: evidence of positive Darwinian selection. *Mol. Biol. Evol.* **12**:363–370.
- SPRATT, B. G., N. H. SMITH, J. ZHOU, M. O'ROURKE, and E. FEIL. 1995. The population genetics of the pathogenic *Neisseria*. Pp. 143–160 in S. BAUMBERG, J. P. W. YOUNG, E. M. H. WELLINGTON, and J. R. SAUNDERS, eds. *Population genetics of bacteria*. Press Syndicate of the University of Cambridge, Cambridge, England.
- STEPHENS, J. C. 1985. Statistical methods of DNA sequence analysis: detection of intragenic recombination or gene conversion. *Mol. Biol. Evol.* **2**:539–556.
- SWOFFORD, D. 1998. PAUP\* Phylogenetic analysis using parsimony and other methods. Sinauer, Sunderland, Mass.
- THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D. G. HIGGINS. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**:4876–4882.
- VAN DER LEY, P., J. E. HECKELS, M. VIRJI, P. HOOGERHOUT, and J. T. POOLMAN. 1991. Topology of outer membrane porins in pathogenic *Neisseria* spp. *Infect. Immun.* **59**:2963–2971.
- VÁZQUEZ, J. A., L. DE LA FUENTE, S. BERRON, M. O'ROURKE, N. H. SMITH, J. ZHOU, and B. G. SPRATT. 1993. Ecological separation and genetic isolation of *Neisseria gonorrhoeae* and *Neisseria meningitidis*. *Curr. Biol.* **3**:567–572.
- WRIGHT, S. 1951. The genetical structure of populations. *Ann. Eugen.* **15**:323–352.
- YANG, Z., and R. NIELSEN. 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* **17**:32–43.
- ZHARKIKH, A. 1994. Estimation of evolutionary distances between nucleotide sequences. *J. Mol. Evol.* **9**:315–329.

ANTONY DEAN, reviewing editor

Accepted November 18, 1999