JOURNAL OF CLINICAL MICROBIOLOGY, Feb. 1994, p. 464–468
0095-1137/94/$04.00+0

Vol. 32, No. 2

Nonrepresentative PCR Amplification of Variable Gene Sequences in Clinical Specimens Containing Dilute, Complex Mixtures of Microorganisms

CHRISTINA J. WRIGHT,[1] ANN E. JERSE,[2] MYRON S. COHEN,[3] JANNE G. CANNON,[2] AND H. STEVEN SEIFERT[1]*

Department of Microbiology-Immunology, Northwestern University Medical School, Chicago, Illinois 60611,[1] and Department of Microbiology and Immunology[2] and Department of Medicine,[3] University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27514

Received 2 September 1993/Returned for modification 28 October 1993/Accepted 16 November 1993

PCR amplification and DNA sequencing of the expression locus from *Neisseria gonorrhoeae* contained in urine sediments collected from experimentally infected human subjects produced two observations. First, different pilin sequences were obtained when separate aliquots of the same sample were amplified and sequenced. In contrast, the same pilin sequence was obtained when repeated amplifications were performed on individual colonies grown from the clinical samples. Second, mixed sequences (i.e., more than one nucleotide at variable positions in the pilin gene sequence) were observed in both the direct clinical isolates and individual cultures grown from the isolates. These results suggest that when clinical samples are directly examined by PCR amplification and sequencing, multiple amplifications may be required to detect sequence variants in the sample and minority variant sequences will not always be detected.

PCR with clinical specimens of tissue, hair, blood, and urine for templates has been used to generate many valuable diagnostic tests which can screen for the presence of infectious disease agents or the genetic predisposition for many congenital diseases. Because of the limiting amount of template in many clinical samples, it is often necessary to amplify target regions of DNA. Although DNA cloning procedures can provide a template for genetic studies, they are time-consuming, and PCR technology has improved the reliability and efficiency of template amplification for diagnostic applications. DNA sequencing has become an important tool in biomedical research, providing information about the molecular basis for many heritable and infectious human diseases, but a significant amount of purified template is required for DNA sequence analysis. Once sufficient template is isolated, a wide variety of methods that use either single- or double-stranded templates can be used to determine the DNA sequence of a gene (3, 6, 7).

Neisseria gonorrhoeae (gonococcus [GC]) is a gram-negative diplococcus and the causative agent of the sexually transmitted disease gonorrhea. Pili are filamentous-like structures emanating from the cell surface of the GC, which aid in attachment to host epithelial cells (8, 14, 15, 18, 21). Pili are also highly immunogenic and undergo antigenic variation to evade the human immune response and to change the functional properties of the pilus (4, 5, 11, 19, 22, 24). One or two expression loci (*pilE*) and multiple silent loci (*pilS*) exist within the GC chromosome (4, 17). *pilE* contains a complete pilin gene that is able to produce pilin monomers which can be assembled into a pilus, while the multiple, transcriptionally silent loci are missing the 5′ promoter sequences and conserved coding sequences (4, 16, 19). Through homologous recombination (10), variant pilin sequences from the silent loci are donated into the expression locus resulting in antigenic variation (4, 5, 19). We are studying pilin antigenic variation in *N. gonorrhoeae* through a human challenge study which uses strain FA1090 to produce clinical signs of disease (2, 20). Pilin variation has been examined in bacteria contained in urine sediments and from individual colonies grown in vitro from the clinical samples. PCR amplification of the expression locus from the GC chromosome allows us to study the changes that occur in pilin sequences during the colonization and infection of the male urethra (20). PCR amplification of target DNA and then cycle sequencing reactions with end-labeled primers have proven to be reliable and efficient means of performing this genetic analysis.

The presence of multiple antigenic variants within urine sediments and urethral swabs collected from the volunteers has created problems in analyzing the changes that occur during infection. We have found that mixed sequences can be observed when the minority sequence is as little as 33% of the sample. In addition, when low amounts of amplifiable template encoding multiple gene sequences are present, individual amplifications of the same sample can result in the determination of different gene sequences. Thus, when complex mixtures of variant gene sequences are present in a clinical sample, multiple amplifications may be required to fully characterize a mixed gene population.

MATERIALS AND METHODS

Origin and collection of clinical samples. Male volunteers between the ages of 18 and 35 years were inoculated with approximately 10^6 cells taken from 100 to 200 colonies of a piliated variant of strain FA1090 by using a pediatric catheter inserted 5 cm into the urethra (2). At 2 h postinoculation, urine samples were collected, and from then on, urine and genital swab samples were taken at three time points daily. In addition to urine sediments, single colony isolates were

* Corresponding author. Mailing address: Department of Microbiology-Immunology, Northwestern University Medical School, Mail drop W312, 303 East Chicago Avenue, Chicago, IL 60611. Phone: (312) 503-9788. Fax: (312) 503-1339. Electronic mail address: h-seifert@nwu.edu.

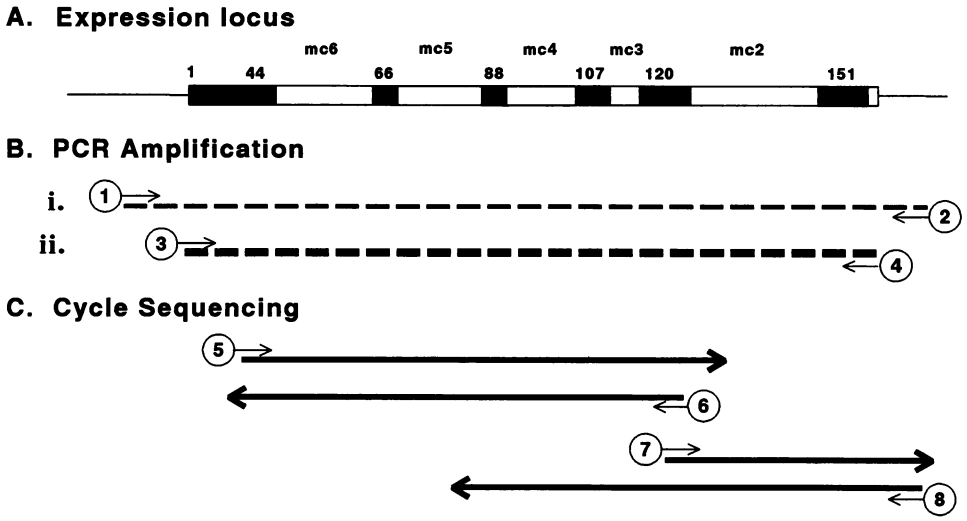

FIG. 1. PCR amplification and cycle sequencing of the pilin expression locus. (A) Schematic depiction of the pilin expression locus of *N. gonorrhoeae*. Listed above the drawing are the minicassettes (minicassettes [mc] 2 to 6; the 3′ minicassette 1 is not labeled) which are proposed regions of variability within the expression locus (4). The numbers above the drawing indicate the amino acid residue number. The black boxes are conserved regions, and the white boxes are variable regions of the expression locus. (B) Amplification of the pilin expression locus out of the GC chromosome. The dotted lines depict the product generated by PCR amplification. Primers are indicated by number, and the 5′ to 3′ orientation is indicated by a thin arrow. Primer 1 is PILSTART, and primer 2 is SP3A; these primers were used in the first PCRs of all samples (Table 1). Primer 3 is PILRBS, and primer 4 is SMACLA1; these primers were used for the nested amplification from the clinical samples (Table 1). (C) Sequencing strategy for the variable portions of the expression locus in GC. The numbers indicate the different primers used for the cycle sequencing reactions, and the thin arrows indicate their direction. The bold arrows show the approximate region of the expression locus which was sequenced by each primer. Primer 5 is CONSTF2, primer 6 is CYS1R, primer 7 is CYS1F, and primer 8 is PILEND (Table 1).

passaged twice in vitro at 37°C with 5% CO_2 on GCB medium (Difco, Detroit, Mich.) with Kellogg supplements (9) and VCN inhibitor (BBL, Cockeysville, Md.) before freezing. All volunteers were treated with ceftriaxone at the onset of clinical signs (purulent exudate) or at 5 days postinoculation. All samples were stored at −80°C in 3% Trypticase soy broth and 25% glycerol.

PCR amplification of DNA templates for sequencing. In order to prepare template for PCR amplification, 5 μl of thawed gonococcal samples were mixed with 5 μl of colony lysis buffer containing 1% Triton X-100, 2 mM EDTA, and 20 mM Tris (pH 8.5). Template-lysis mixtures were then vortexed for 1 min, heated at 94°C for 15 min, vortexed again for 1 min, and used as templates for the PCR.

Because of the lower amounts of bacteria present in the clinical samples, nested PCR was used to amplify *pilE* out of the bacterial chromosome (Fig. 1). When frozen stocks of isolated colonies were used as templates, only single PCR amplifications were necessary to produce an adequate amount of template for sequencing. However, nested PCR from these samples provided identical sequence information as single amplification. The primers used in PCR are listed in Table 1.

All amplifications were done in 100-μl volumes containing 50 pmol of each primer, 0.2 mM (each) deoxynucleoside triphosphates (Promega, Madison, Wis.), 50 mM KCl, 10 mM Tris-HCl, 0.1% Triton X-100, 3 mM $MgCl_2$, and 2 U of *Taq* DNA polymerase (Promega). Evaporation of the reaction mixture was prevented by using paraffin beads (25). The thermocycling profile used for the outer primer pair was 30 cycles of 94°C for 1 min, 60°C for 1 min, and 72°C for 2 min, producing a product of approximately 780 bp (PTC-100, MJ Research, Watertown, Mass.). The thermocycling profile used for the internal primer pair was 30 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 2 min, producing a product of approximately 630 bp.

DNA sequence analysis. Prior to sequencing, the entire PCR amplification mixture was passed through a Sepharose

TABLE 1. Oligonucleotides used for PCR and sequencing strategy

Primer	DNA sequence[a]	Description
PILSTART	GAGATAAACGCATAAAATTTCACC	Sequence in 5′ conserved promoter region of *pilE1* used with SP3A
SP3A	CCGGAACGGACGACCCCG	Sequence in 3′ conserved untranslated region of all *pil* loci
PILRBS	GGCTTTCCCCTTTCAATTAGGAG	Sequence at ribosome-binding site of *pilE* used with SMACLA1
SMACLA1	CAAACCCTTAAAAGACAAGC	Sequence in 3′ untranslated region of all *pil* loci
CONSTF2	TACCAAGACTACACCGCCCG	Sequence in 5′ conserved region of all *pil* loci used for sequencing
CYS1R	GTCCGCAGAACCATTTTACCG	Sequence in conserved *cys1* region of all *pil* loci used for sequencing
CYS1F	CGGTAAAATGGTTCTGCGGAC	Inverse complement of CYS1R used for sequencing
PILEND	CGCTTGATTTATTTAAAATTTAAGG	Sequence in 3′ untranslated region of some *pil* loci used for sequencing

[a] All DNA sequences are listed 5′ to 3′.

CL-6B (Sigma, St. Louis, Mo.) spin column containing 15 times the sample volume of a 60% slurry of resin in TE (10 mM Tris [pH 8.0], 1 mM EDTA). This removed both residual oligonucleotide primers and buffer components from the PCR amplification mixture. Approximately 10 to 50 ng of amplified template from the spin-dialyzed PCR mixture was then used for sequencing by using [γ-^{32}P]ATP-end-labeled primers and the *fmole* DNA Sequencing System (Promega). The end-labeled primers corresponded to conserved regions within the *pilE* and generated complete double-stranded sequence data for the variable portions of each pilin gene (Table 1 and Fig. 1). The thermocycling profile used with these primers was initial denaturation at 95°C for 2 min and then 30 cycles of 95°C for 30 s, 60°C for 30 s, and 70°C for 1 min. The samples were heated at 75°C for 2 min and run on a 5% Long Ranger denaturing gel (J. T. Baker, Phillipsburg, N.J.) at 60 W, dried at 80°C for 40 min, and exposed to X-OMAT-AR film (Kodak, Rochester, N.Y.) overnight.

Sequencing of mixed templates. Purified PCR products of two different pilin variants were diluted, and their concentrations were equalized by matching band intensities on ethidium bromide-stained agarose gels. These two templates were then mixed at the following ratios: 1:1, 2:1, 4:1, and 8:1. One template was kept at 10 ng, and the concentration of the other template was varied by diluting at twofold increments, from 10 to 1.25 ng. Once the templates were combined, the mixture was used as the starting template for sequencing reactions following the protocol described above. To analyze mixed sequences after PCR amplification, serial dilutions of the two starting templates were analyzed by PCR with the nested primers until the minimal concentration that produced a visible product on an ethidium bromide-stained agarose gel was found. Approximately 1.6 pg of starting template was consistently able to produce a visible PCR product. One template was used at 1.6 pg, and the other template was diluted at twofold increments from 1.6 to 0.2 pg. Templates were then mixed at ratios of 1:1, 1:2, 1:4, and 1:8 and amplified by PCR; this was followed by DNA sequencing.

RESULTS AND DISCUSSION

During our studies of gonococcal pilin variation, we collected urine sediments containing various numbers of viable bacteria (Table 2). In addition, bacteria were isolated from genital swab specimens from the time that clinical signs appeared. The *pilE* locus of the bacteria was amplified by PCR by using either nested or single rounds of PCR. The resultant PCR products were then sequenced by using the *fmole* DNA Sequencing kit (Promega) and four oligonucleotides that produced the DNA sequence of both strands of the variable gene sequences.

Prior to these studies, the number of different variant gene sequences that could be expressed at any one time during infection was unknown. We assumed that for any one sample one or two predominant sequences would be expressed in the population. Our initial observations showed that when amplification was done from the urine sediments, one or two pilin variant sequences were observed. However, when we returned to an identical urine sample and reamplified it to create a new sequencing template, gene sequences different from those found originally were observed (Table 2). Therefore, it appears that the urine sediments contain mixed populations of bacteria expressing different pilin genes. The initial uniform gene sequences gained from the urine samples presumably represent one or two variant genes in the population that were amplified early in the nested PCRs. The different sequences found were not a result of errors produced by the *Taq* polymerase since the changes occurred only in variable positions of the pilin gene (20). Additionally, the number of differences was much greater than that which could be produced by polymerase misincorporation (20). Thus, each time the sample was used to produce a new template, a new member of the population was amplified early and produced a different sequence. Multiple amplifications and sequencing reactions may be required to obtain a better representation of the population present within any mixed clinical sample.

TABLE 2. Number of sequences generated from repeated amplifications of direct clinical versus colony isolate samples

Isolate	No. of:		
	Viable bacteria	Amplifications	Different sequences
Colony isolates			
Inoculum[a]	$>10^6$	3	1
1-81-S2[b]	$>10^6$	2	1
Clinical samples			
1-2-BU	1.9×10^3/ml	3	3
1-57-BU	1.6×10^2/ml	2	2
2-2-BU	4.7×10^2/ml	2	2

[a] The FA1090 inoculum was the sample used to initiate infection in the volunteers (20).

[b] The numbering system for the rest of the samples may be read as follows: the first number indicates the volunteer from whom the sample was collected, the second number indicates the time postinoculation (in hours) that the sample was collected, B indicates a direct clinical specimen, U indicates a urine sample, and S indicates a sample from a genital swab.

In addition to the problem of clinical samples yielding different sequence information upon reamplification, sequences were also mixed because of the presence of two or more variant templates that contributed to the sequence information. To determine the percentage of the total template at which a particular pilin variant must be present to provide mixed sequences, we performed template mixing experiments. Low levels of two different templates were equalized on an ethidium bromide-stained agarose gel and were then mixed at ratios of 1:1, 2:1, 4:1, and 8:1. The mixed templates were then used for cycle sequencing reactions; the minority sequence became visible at template ratios of 4:1 (Fig. 2B). In other words, the template producing the secondary sequence must be present at levels of 20% of the total template or greater in order to produce a visible band on the autoradiogram. However, it was not until the 2:1 ratio (33% of total template) that the secondary sequence began to rise to levels above what we would consider normal background levels (Fig. 2B). Template mixing prior to PCR amplification was done in two ways, keeping one template constant and diluting the other template. Both mixing experiments produced similar results. When the different templates were diluted to the minimal level that produced a detectable amplification product and were mixed prior to both PCR amplification and DNA sequencing, the secondary sequence was again observable at the 1:4 ratio of template mixing and above background levels at the 1:2 ratio (Fig. 2A). Secondary sequences resulting from template mixing prior to PCR amplification were less intense than those resulting from template mixing and then direct sequence analysis. Similar results from template mixing experiments were seen by

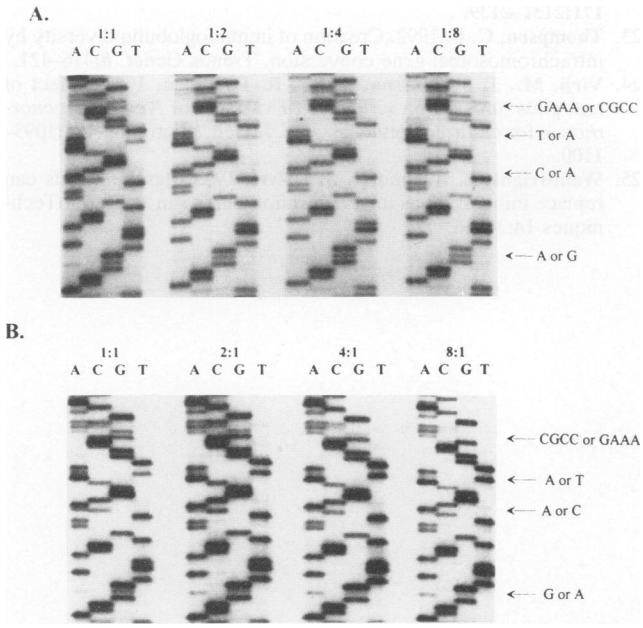

FIG. 2. Template mixing and its effect on PCR amplification and DNA sequencing. (A) Templates were mixed at the ratios listed above the panel, with template B at 1.6 pg and template A at twofold dilutions from 1.6 to 0.2 pg. Points of sequence variation between the two genes are indicated on the right. Mixed templates were used for PCR amplification and then DNA sequencing, generating the data shown. (B) By using the same stock of template as was used in panel A, template A remained at 10 ng and template B was diluted at twofold dilutions from 10 to 1.25 ng. Mixed templates were sequenced directly, and points of sequence variation are indicated on the right.

Leitner et al. (12), who used solid-phase purification of single-stranded template and fluorescence-based automated sequencing. Detection of secondary sequences in this system was more sensitive, with 10 to 25% of total template being detectable (12).

When PCR amplification is used to produce sequencing template from a uniform source, a pure culture, or a cell line, the possibility of mixed sequences being present is remote. Alternatively, when clinical or environmental samples are used to generate sequencing templates, it is often possible that more than one gene sequence is present in the sample. The gonococcal pilin gene is unusual in that it is highly variable, with multiple sequence variants possible in a single strain. The observations reported here would also relate to sequence analysis of other variable genes such as *Borrelia* variable major proteins (1), *Trypanosoma* variable surface glycoproteins (13), or rabbit or chicken immunoglobulin genes (23). In addition, mixed populations of different strains of a microorganism would create a similar situation when more than one version of a gene sequence was present. These observations may also be pertinent when a gene family is studied in a clonal cell population. Our data show that if single amplifications are performed, only one of the gene sequences might be recorded and others might be missed. Therefore, in order to obtain an accurate representation of variant gene sequences expressed in populations, multiple PCR and sequence analyses may be required. While this type of selective amplification is an important factor when samples are sequenced directly, it could also occur when a variable gene sequence is cloned from a PCR-amplified sample.

ACKNOWLEDGMENTS

The research described here was supported by the following grants: R29AI27195, MO1R0046, UO1AI31494, and UO1AI31496. H.S.S. is a recipient of a Junior Faculty award from the American Cancer Society.

REFERENCES

1. **Barbour, A. G.** 1990. Antigenic variation of a relapsing fever *Borrelia* species. Annu. Rev. Microbiol. **44**:155–171.
2. **Cohen, M. S., J. G. Cannon, A. E. Jerse, L. Charniga, S. Isbey, and L. Whicker.** J. Infect. Dis., in press.
3. **Gyllensten, U. B., H. A. Erlich, H. H. Li, X. F. Cui, R. K. Saiki, and N. Arnheim.** 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus: amplification and analysis of DNA sequences in single human sperm and diploid cells. Proc. Natl. Acad. Sci. USA **335**:414–417.
4. **Haas, R., and T. F. Meyer.** 1986. The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: evidence for gene conversion. Cell **44**:107–115.
5. **Hagblom, P., E. Segal, E. Billyard, and M. So.** 1985. Intragenic recombination leads to pilus antigenic variation in *Neisseria gonorrhoeae*. Nature (London) **315**:156–158.
6. **Higuchi, R. G., and H. Ochman.** 1989. Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. **17**:5865.
7. **Hultman, T., S. Stahl, E. Hornes, and M. Uhlen.** 1989. Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as a solid support. Nucleic Acids Res. **17**:4937–4946.
8. **Johnson, A. P., D. Taylor-Robinson, and Z. A. McGee.** 1977. Species specificity of attachment and damage to oviduct mucosa by *Neisseria gonorrhoeae*. Infect. Immun. **18**:833–839.
9. **Kellogg, D. S., Jr., W. L. Peacock, W. E. Deacon, L. Brown, and C. I. Pirkle.** 1963. *Neisseria gonorrhoeae*. I. Virulence genetically linked to colonial variation. J. Bacteriol. **85**:1274–1279.
10. **Koomey, M., E. C. Gotschlich, K. Robbins, S. Bergström, and J. Swanson.** 1987. Effects of *recA* mutations on pilus antigenic variation and phase transitions in *Neisseria gonorrhoeae*. Genetics **117**:391–398.
11. **Lambden, P. R., J. E. Heckels, and P. J. Watt.** 1982. Effect of anti-pilus antibodies on survival of gonococci within guinea pig subcutaneous chambers. Infect. Immun. **38**:27–30.
12. **Leitner, T., E. Halapi, G. Scarlatti, P. Rossi, J. Albert, E.-M. Fenyo, and M. Uhlen.** 1993. Analysis of heterogeneous viral populations by direct DNA sequencing. BioTechniques **15**:120–127.
13. **Lex, H. T., V. der Ploeg, K. Gottesdiener, and M. G. S. Lee.** 1992. Antigenic variation in African trypanosomes. Trends Genet. **8**:452–456.
14. **McGee, Z. A., A. P. Johnson, and D. Taylor-Robinson.** 1981. Pathogenic mechanisms of *Neisseria gonorrhoeae*: observations on damage to human fallopian tubes in organ culture by gonococci of colony type 1 or type 4. J. Infect. Dis. **143**:413–422.
15. **Melly, M. A., C. R. Gregg, and Z. A. McGee.** 1981. Studies of toxicity of *Neisseria gonorrhoeae* for human fallopian tube mucosa. J. Infect. Dis. **143**:423–431.
16. **Meyer, T. F., E. Billyard, R. Haas, S. Storzbach, and M. So.** 1984. Pilus genes of *Neisseria gonorrhoeae*: chromosomal organization and DNA sequence. Proc. Natl. Acad. Sci. USA **81**:6110–6114.
17. **Meyer, T. F., N. Mlawer, and M. So.** 1982. Pilus expression in *Neisseria gonorrhoeae* involves chromosomal rearrangement. Cell **30**:45–52.
18. **Rudel, T., J. P. M. van Putten, C. P. Gibbs, R. Haas, and T. F. Meyer.** 1992. Interaction of two variable proteins (PilE and PilC) required for pilus-mediated adherence of *Neisseria gonorrhoeae* to human epithelial cells. Mol. Microbiol. **6**:3439–3450.
19. **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.
20. **Seifert, H. S., C. Wright, A. Jerse, M. Cohen, and J. G. Cannon.** Submitted for publication.
21. **Stephens, D. S., J. W. Krebs, and Z. A. McGee.** 1984. Loss of pili and decreased attachment to human cells by *Neisseria meningitidis* and *Neisseria gonorrhoeae* exposed to subinhibitory concentrations of antibiotics. Infect. Immun. **46:**507–513.
22. **Swanson, J., S. Morrison, O. Barrera, and S. Hill.** 1990. Piliation changes in transformation-defective gonococci. J. Exp. Med. **171:**2131–2139.
23. **Thompson, C. B.** 1992. Creation of immunoglobulin diversity by intrachromosomal gene conversion. Trends Genet. **8:**416–421.
24. **Virji, M., J. S. Everson, and P. R. Lambden.** 1982. Effect of anti-pilus antisera on virulence of variants of *Neisseria gonorrhoeae* for cultured epithelial cells. J. Gen. Microbiol. **128:**1095–1100.
25. **Wainwright, L. A., and H. S. Seifert.** 1993. Paraffin beads can replace mineral oil as an evaporation barrier in PCR. BioTechniques **14:**34–36.