APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Jan. 2001, p. 469–472 Vol. 67, No. 1
0099-2240/01/$04.00+0 DOI: 10.1128/AEM.67.1.469–472.2001

Microvariation Artifacts Introduced by PCR and Cloning of Closely Related 16S rRNA Gene Sequences†

ARJEN G. C. L. SPEKSNIJDER,[1,2]* GEORGE A. KOWALCHUK,[3] SANDER DE JONG,[2] ELIZABETH KLINE,[4] JOHN R. STEPHEN,[4]‡ AND HENDRIKUS J. LAANBROEK[2]

Department of Zoology, Natural History Museum, South Kensington, London SW7 5BD, United Kingdom[1]; Department of Plant-Microorganism Interactions, Centre for Terrestrial Ecology, Netherlands Institute of Ecology, 6666 ZG Heteren,[3] and Department of Microbial Ecology, Centre for Limnology, Netherlands Institute of Ecology, 3136 AC Nieuwersluis,[2] The Netherlands; and Center for Environmental Biotechnology, University of Tennessee, Knoxville, Tennessee 37932-2575[4]

Received 3 August 2000/Accepted 6 October 2000

A defined template mixture of seven closely related 16S-rDNA clones was used in a PCR-cloning experiment to assess and track sources of artifactual sequence variation in 16S rDNA clone libraries. At least 14% of the recovered clones contained aberrations. Artifact sources were polymerase errors, a mutational hot spot, and cloning of heteroduplexes and chimeras. These data may partially explain the high degree of microheterogeneity typical of sequence clusters detected in environmental clone libraries.

The use of PCR targeting the 16S rRNA gene has provided a powerful, culture-independent means of characterizing microbial communities, which has advanced our understanding of microbial diversity and evolution (16). In addition to revealing novel lineages, the analysis of environmental 16S rDNA clone libraries often produces unresolved "bushes" of closely related clones (7, 22). It is not known to what extent these bushes relate to true microheterogeneity of ribotypes versus noise introduced by PCR and cloning procedures. We therefore sought to determine the potential for multiple competitive PCR and cloning of the products thereof to induce an elevated occurrence of sequence errors under conditions as close to optimal as possible. This study for the first time uses a controlled experiment to track potential origins of microvariation, including intermolecular interactions, within 16S rDNA clone libraries.

The experiment utilized a multiple competitive PCR amplification, designed to simulate 16S rDNA retrieval of closely related sequences from environmental samples. Seven closely related environmental clones, Gm3 (accession number AJ003751), Vm6 (accession no. AJ003758), Gm9 (accession no. AJ003752), Vm10 (accession no. AJ003760), Vm11 (accession no. AJ003761), Vm12 (accession no. AJ003762), and Ws23 (accession no. AJ003775), containing 16S rDNA sequences related to ammonia-oxidizing bacteria of the β-subgroup *Proteobacteria* (20), were mixed in equal quantities. A mixed-plasmid template was chosen to maximize control of DNA quality and quantity while avoiding potential variables of chromosomal DNA templates, such as chromosome size, *rrn* copy number, sequences flanking the priming sites, and intrastrain rRNA heterogeneity (5). PCR was performed using 0.5 ng of the mixed-plasmid template, the Expand High Fidelity (Expand H-F) DNA polymerase system (Boehringer, Mannheim, Germany), primers matching the terminal ends of the 16S rDNA insert (13), and 25 thermocycles, as described previously (21). Expand H-F polymerase has a reported error rate of 8.5×10^{-6} mismatches bp^{-1} and consists of a mixture of *Taq* and *Pwo* polymerases, the latter containing 3′ exonuclease activity. The 1,180-bp PCR fragment was recovered following standard agarose electrophoresis using QIAquick columns (Qiagen, Hilden, Germany) and ligated into a pGEM-T vector (Promega, Madison, Wisc.) for transformation of Epicurian Coli XL1-Blue MRF′ cells (Stratagene, La Jolla, Calif.). Seventy white colonies were chosen randomly from Luria broth–5-bromo-4-chloro-3-indolyl β-D-galactopyranoside–isopropyl-β-D-thiogalactopyranoside 1.5% agar plates for colony PCR using the 357f-GC/518r primers, a touch-down thermocycling program (14), and *Taq* polymerase (2 U) (Gibco Laboratories, Detroit, Mich.), and 66 produced a 180-bp product. Denaturing gradient gel electrophoresis (DGGE) screening (27) revealed that 14% (9 clones) exhibited novel DGGE migration (designated Nclone). All original clone mobilities were observed in the clone library, with each original mobility being found in 6 to 18% of the clones examined (Fig. 1).

All clones with novel DGGE mobility plus two additional clones (Pclone4 and Pclone29) were chosen for full-length sequencing of both DNA strands (21) by two or more researchers to eliminate the possibility of misreading faithful sequence data, and a compressed alignment and distance matrix are shown in Fig. 2. The most closely related original clone sequence was deemed the putative parent sequence. Novel clone sequences (Nclone) showed a divergence from their putative parent sequences ranging between 0.2 and 1.2%. Pclone4 and Pclone29 revealed deviations of 0.5 and 0.4% from their putative original clone sequences, with all sequence differences falling outside the region examined by DGGE.

Random base changes were defined as unique base-pair differences that were not represented in any of the original

* Corresponding author. Present address: Dept. of Molecular and Cell Biology, IMS Building, Foresterhill, University of Aberdeen, Aberdeen AB25 2ZD, United Kingdom. Phone: 44 1224 273149. Fax: 44 1224 273144. E-mail: a.speksnijder@abdn.ac.uk.

† Publication 2679 of the NIOO Centre for Limnology, Nieuwersluis, The Netherlands.

‡ Present address: Crop and Weed Science, Horticulture Research International, Wellesbourne, UK CV35 9EF.

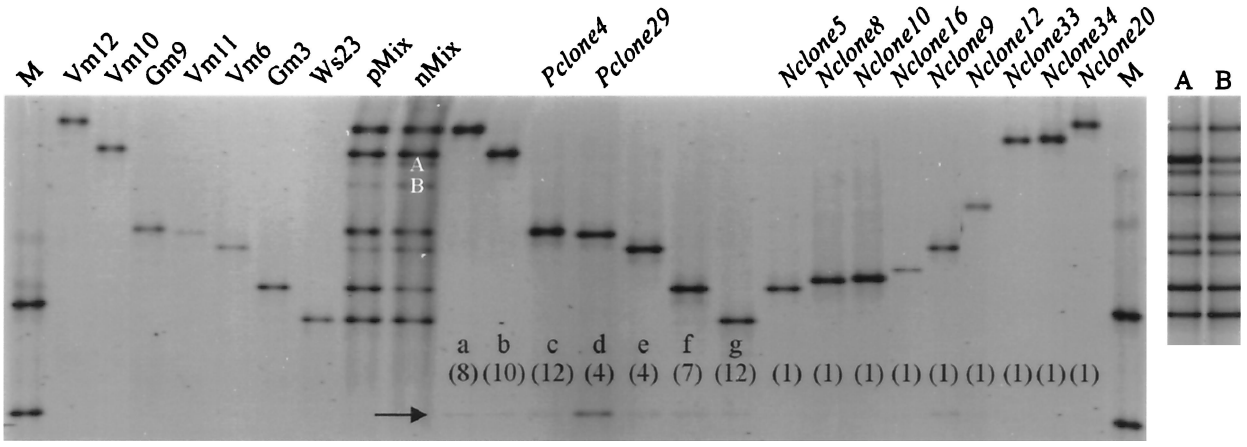

FIG. 1. DGGE analysis of individual clones and mixed PCR products. Vm, Gm, and Ws designations indicate PCR products derived from the original seven clones used to create the template mix (21). Lanes labeled a to g give examples of recovered clones whose DGGE mobilities match those of one of the original clones. The numbers between parentheses indicate the number of recovered clones showing the given DGGE mobility. pMix indicates that the original plasmid mix was used as a template in an amplification reaction using the 357f-GC and 518 primers, and the nMix sample used products of the βAMOf/βAMOr PCR as a template. The arrow indicates *E. coli* contamination. The poor amplification of clone Vm11 is due to a single mismatch with the 518r primer. The marker lanes (M) contained PCR-amplified 16S rDNA fragments of *Lactococcus lactis* and *E. coli* according to the method of Zwart et al. (27). The bands labeled A and B in the nMix sample were excised for DNA elution and reamplification using the 357f-GC and 518 primers. The DGGE analysis of these products is shown in the right panel, with lanes labeled according to the original excised band.

clone sequences. Nclone10 had a change at position 449 (G), Nclone12 at positions 433 (A) and 798 (C); Nclone16 at 474 (G) and 716 (G); Pclone29 at 1062 (C); Nclone20 at 435 (C); Nclone34 at 443 (C), 816 (C), and 986 (C); Nclone33 at 996 (T); and Nclone8 at 300 (G) and 511 (T). Nclone16 also had a 3-bp stretch of foreign sequence from position 581 to 583 (GCA→CAG). Pclone4 contained two one-base deletions and a single one-base insertion at positions 797, 811, and 823, respectively. The introduction of random base anomalies depends on a number of factors, including the DNA polymerase and the number of PCR cycles used (9). However, even when using a proofreading polymerase system, as described here, this source of sequence error study may be high enough to affect genetic diversity estimates.

The majority of the aberrant clones contained an insertion of an A, a T, or an AT between *Escherichia coli* positions 389 and 390 (Fig. 2). No evidence for such a mutational hot spot was observed upon inspection of environmental clones from the database, and the origin of these sequence aberrations is not known.

In addition to errors introduced by polymerase and sequencing mistakes (9, 26), the formation of chimeric DNA molecules during the PCR has been recognized as a source of sequence infidelity (12, 24, 25). The Pclone4 sequence could be the result of a chimera between clones Gm9 and Vm11, which occurred between positions 1021 and 1036. DGGE analysis did not span this region, which would explain why this putative chimera was not detected by the screening procedure.

The frequency of heteroduplex formation (4, 8, 19) would be expected to increase in the later cycles of mixed PCR amplifications, when the amplified DNA species reach concentrations high enough to compete with primers for binding sites (23). Cloned heteroduplex molecules may be subjected to *E. coli* DNA repair mechanisms (2, 3, 11, 17, 18), resulting in hybrid plasmid inserts. Nclone8 contains marker nucleotides from clones Ws23 and Vm10. Vm10-like sequence spans positions 332 to 359, 629 to 645, and 932 to 1053, suggesting the possible role of heteroduplex formation followed by mismatch resolution in *E. coli*. In addition, Nclone8 appeared to contain two random base anomalies. Nclone34 contained five positions where heteroduplex-induced changes could have generated sequence differing from that of the putative parental clone, Vm12. Four of these positions, 203, 389, 629, and 862, are present in the original clone, Vm10, with two of these being identical to the consensus sequence. In addition, several clones differ from their putative parent sequence at positions where several clones might be implicated in sequence donation by intermolecular interactions. In such cases, it is not possible to determine which original clone was most likely to have been the source of the aberrant base pair.

An identical experiment using only a single environmental clone as a template (Vm6) resulted in only a 2% (1 of 50) recovery rate of clones showing novel DGGE mobility. This is in good agreement with the expected frequency of the error-containing fragments (1 to 2%) based upon the length of DNA fragment analyzed by DGGE, the number of cycles performed in the cloning experiment (25 cycles), and the enzyme system used (Expand H-F). Thus, the majority of aberrant clones seem to have been due to the interaction of the different template molecules during the PCR-cloning procedure.

To check heteroduplex formation, PCR-DGGE profiles were generated using the original plasmid mix and its PCR product as a DNA template (Fig. 1). Two novel DGGE bands were observed for both of these mixed products. Excision, re-amplification, and DGGE analysis of these two bands produced all the bands of the original mixed products, with differences in relative band intensities (Fig. 1, right panel), suggesting that these new bands contained more than a single homoduplex molecule (6, 15).

The fact that several of the clones, including Pclone4 and

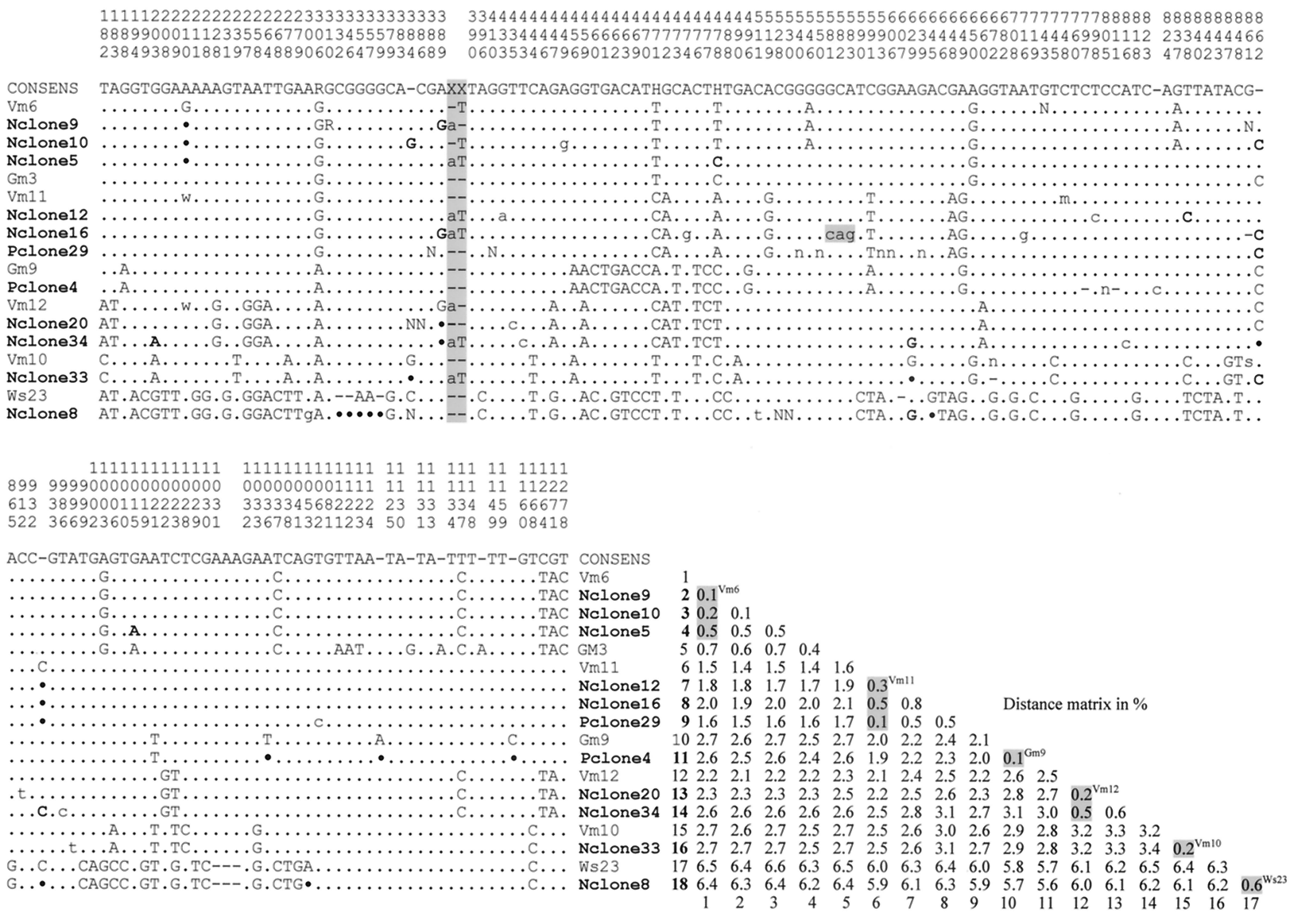

FIG. 2. Compressed alignment and distance matrix of original and novel clones. The names of clones derived in this study are shown in boldface type. The alignment was compressed by deleting constant characters (except positions adjacent to gaps). Dots denote no difference from the consensus sequence. Base numbering is according to *E. coli* convention (1) and is given vertically above the consensus sequence. Bases given in lowercase indicate characters that are not present in any of the original clones used in the template mixture. Boldface type, including heavy dots, designates positions for a potential chimera or heteroduplex correction. The grey shading represents the region of the mutational hot spot and the introduction of a foreign 3-bp sequence (CAG). Clones were ordered according to the relationship between novel clones and the original clones to which they are most closely related, and these levels of sequence divergence are shaded in grey and labeled with the appropriate original clone designation in the distance matrix. Both the alignment and the distance matrix were constructed with the SeqPup program (D. B. Gilbert, SeqPup, a biosequence editor and analysis application [ftp://iubio.bio.indiana.edu/molbio/seqpup/], 1996).

Pclone29, contain introduced errors outside the region used in the DGGE screening argues that the observed aberrant clone frequency of 14% is most certainly an underestimation. Despite a theoretical detection of more than 97% of all sequence variants for the size of DNA fragment screened (20), DGGE may fail to detect all sequence variants, thus leading to a further underestimation of unique clones. The overall phylogenetic placement of the recovered sequences was not affected by the introduced sequence anomalies (not shown). However, all the clones used in this study were affiliated with the same narrow *Nitrosomonas*-like sequence cluster (21). The exchange of sequence regions between more phylogenetically diverse sequences could impair phylogenetic placement. The frequency of occurrence of minor sequence artifacts found here supports the common practice of grouping clusters of sequences with less than 3% sequence variation when interpreting 16S rDNA clone data, and further relaxation in assignment of operational taxonomic units at the 95% similarity level may be warranted. Although different ecotypes may exhibit nearly identical 16S rRNA sequences (6), the identification of ecologically distinct microbial populations clearly demands additional evidence, beyond the recovery of closely related sequences from clone libraries. Detection of sequence hybrids between closely related 16S rDNA molecules is problematic (11; N. Larsen, Check__Chimera program of the Ribosomal Database Project, 1999), making it difficult to estimate the amount of real sequence microvariation in extant databases. The identified sources of aberrant sequence information merit consideration during both primer and probe design, as do the ecological and evolutionary interpretation of microheterogeneity within such environmental data, since a significant proportion of such variation may be artifactual.

John R. Stephen and Elizabeth Kline were funded by the National Science Foundation (grant number DEB 9814813) and Department of Energy, Office of Energy Research (grant number DE-FC02-96ER62278White).

REFERENCES

1. **Brosius, J., T. L. Dull, D. D. Sleeter, and H. F. Noller.** 1981. Gene organisation and primary structure of a ribosomal RNA operon from Escherichia coli. J. Mol. Biol. **148:**107–127.
2. **Caraway, M., and M. G. Marinus.** 1993. Repair of heteroduplex DNA molecules with multibase loops in *Escherichia coli*. J. Bacteriol. **175:**3972–3980.
3. **Cariello, N. F., W. G. Thilly, J. A. Swenberg, and T. R. Skopek.** 1991. Deletion mutagenesis during polymerase chain reaction: dependence on DNA polymerase. Gene **99:**105–108.
4. **Espejo, R. T., C. G. Feijo, J. Romero, and M. Vasquez.** 1998. PAGE analysis of the heteroduplexes formed between PCR-amplified 16S rRNA genes: estimation of sequence similarity and rDNA complexity. Microbiology (United Kingdom) **144:**1611–1617.
5. **Farrelly, V., F. A. Rainey, and E. Stackebrandt.** 1995. Effect of genome size and rrn gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. Appl. Environ. Microbiol. **61:**2798–2801.
6. **Ferris, M. J., G. Muyzer, and D. M. Ward.** 1996. Denaturing gradient gel electrophoresis profiles of 16S rRNA-defined populations inhabiting a hot spring microbial mat community. Appl. Environ. Microbiol. **62:**340–346.
7. **Fuhrman, J. A., and L. Campbell.** 1998. Microbial microdiversity. Nature **393:**410–411.
8. **Jensen, M. A., and N. Straus.** 1993. Effect of PCR conditions on the formation of heteroduplex and single-stranded DNA products in the amplification of bacterial ribosomal DNA spacer regions. PCR Methods Appl. **3:**186–194.
9. **Keohavong, P., and W. G. Thilly.** 1989. Fidelity of DNA polymerases in DNA amplification. Proc. Natl. Acad. Sci. USA **86:**9253–9257.
10. **Komatsoulis, G. A., and M. S. Waterman.** 1997. A new computational method for detection of chimeric 16S rRNA artifacts generated by PCR amplification from mixed bacterial populations. Appl. Environ. Microbiol. **63:**2338–2346.
11. **Learn, B. A., and R. H. Grafstrom.** 1989. Methyl-directed repair of frameshift heteroduplex in cell extracts from *Escherichia coli*. J. Bacteriol. **173:**6473–6481.
12. **Liesack, W., H. Weyland, and E. Stakebrandt.** 1991. Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed culture of strict barophilic bacteria. Microb. Ecol. **21:**192–198.
13. **McCaig, A. E., T. M. Embley, and J. I. Prosser.** 1994. Molecular analysis of enrichment cultures of marine ammonium oxidisers. FEMS Microbiol. Lett. **120:**363–368.
14. **Muyzer, G., E. C. de Waal, and A. G. Uitterlinden.** 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. **59:**695–700.
15. **Nagamine, C. M., K. Chan, and Y. C. Lau.** 1989. A PCR artifact: generation of heteroduplexes. Am. J. Hum. Genet. **45:**337–339.
16. **Pace, N. R.** 1997. A molecular view of microbial diversity and the biosphere. Science **276:**734–740.
17. **Parker, B. O., and M. G. Marinus.** 1992. Repair of DNA heteroduplexes containing small heterologous sequences in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **89:**1730–1734.
18. **Sancar, A., and G. B. Sancar.** 1988. DNA repair enzymes. Annu. Rev. Biochem. **57:**29–68.
19. **Separovic, E., and S. A. Nadindavis.** 1995. Secondary structure within PCR target sequences may facilitate heteroduplex production. PCR Methods Appl. **3:**248–251.
20. **Sheffield, V. C., D. R. Cox, L. S. Lerman, and R. M. Myers.** 1989. Attachment of a 40-base-pair G + C-rich sequence (GC-clamp) to genomic DNA fragments by the polymerase chain reaction results in improved detection of single-base changes. Proc. Natl. Acad. Sci. USA **86:**232–236.
21. **Speksnijder, A. G. C. L., G. A. Kowalchuk, K. Roest, and H. J. Laanbroek.** 1998. Recovery of a *Nitrosomonas*-like 16S rDNA sequence group from freshwater habitats. Syst. Appl. Microbiol. **21:**321–330.
22. **Stephen, J. R., A. E. McCaig, Z. Smith, J. I. Prosser, and T. M. Embley.** 1996. Molecular diversity of soil and marine 16S rRNA gene sequences related to β-subgroup ammonia-oxidizing bacteria. Appl. Environ. Microbiol. **62:**4147–4154.
23. **Suzuki, M. T., and S. J. Giovannoni.** 1996. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. **62:**625–630.
24. **Wang, G. C.-Y., and Y. Wang.** 1996. The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. Microbiology (United Kingdom) **142:**1107–1114.
25. **Wang, G. C.-Y., and Y. Wang.** 1997. The frequency of formation of chimeric molecules as a consequence of PCR coampification of 16S rRNA genes from mixed bacterial genomes. Appl. Environ. Microbiol. **63:**4645–4650.
26. **Weisburg, W. G., S. M. Barns, D. A. Pelletier, and D. J. Lane.** 1991. 16S ribosomal DNA amplification for phylogenetic study. J. Bacteriol. **173:**697–703.
27. **Zwart, G., R. Huismans, M. P. Van Agterveld, Y. Van de Peer, P. De Rijk, H. Eenhoorn, G. Muyzer, E. J. Van Hannen, H. J. Gons, and H. J. Laanbroek.** 1998. Divergent members of the bacterial division Verrucomicrobiales in a temperate freshwater lake. FEMS Microbiol. Ecol. **25:**159–169.