Applied and Environmental Microbiology, Dec. 2005, p. 8966–8969 Vol. 71, No. 12
0099-2240/05/$08.00+0 doi:10.1128/AEM.71.12.8966–8969.2005
Copyright © 2005, American Society for Microbiology. All Rights Reserved.

PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample†

Silvia G. Acinas, Ramahi Sarma-Rupavtarm, Vanja Klepac-Ceraj, and Martin F. Polz*

Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

Received 31 March 2005/Accepted 26 August 2005

The contribution of PCR artifacts to 16S rRNA gene sequence diversity from a complex bacterioplankton sample was estimated. *Taq* DNA polymerase errors were found to be the dominant sequence artifact but could be constrained by clustering the sequences into 99% sequence similarity groups. Other artifacts (chimeras and heteroduplex molecules) were significantly reduced by employing modified amplification protocols. Surprisingly, no skew in sequence types was detected in the two libraries constructed from PCR products amplified for different numbers of cycles. Recommendations for modification of amplification protocols and for reporting diversity estimates at 99% sequence similarity as a standard are given.

Estimation of the extent of PCR-induced artifacts in microbial diversity studies remains an important task in the search for patterns and extent of microbial diversity. The basic types of PCR artifacts have been shown in controlled laboratory studies and can be divided into two categories: those resulting in sequence artifacts (PCR errors), and those skewing the distribution of PCR products due to unequal amplification (PCR bias) or cloning efficiency. Sequence artifacts may arise due to (i) the formation of chimerical molecules (3, 10, 14, 15, 25, 26, 37, 38), (ii) the formation of heteroduplex molecules (25, 27, 29, 32), and (iii) *Taq* DNA polymerase error (4, 25). PCR bias is thought to be due to intrinsic differences in the amplification efficiency of templates (23) or to the inhibition of amplification by the self-annealing of the most abundant templates in the late stages of amplification (31). However, it remains difficult to translate these results to environmental samples in which target genes are orders of magnitude more highly concentrated than in the simple mixtures of templates generally used in controlled laboratory studies.

Here, we address the following questions. (i) To what extent do different PCR errors contribute to overestimation of microbial diversity? (ii) Do these PCR errors suggest differences in community structure? (iii) To what extent does PCR bias result in different template distributions after various cycle numbers? Finally, we derive and reiterate recommendations to minimize PCR artifacts.

We have recently generated two large 16S rRNA gene libraries (~1,000 sequences each) from a single bacterioplankton sample (1), providing an opportunity to evaluate PCR artifacts in a realistic setting. The first (standard) library was constructed using 35-cycle amplification to mimic commonly used protocols. The second (modified) library was based on the following amplification protocol to reduce the accumulation of PCR artifacts: limitation to 15 cycles of amplification to decrease PCR bias (23) and accumulation of *Taq* DNA polymerase errors and chimerical sequence formation (25), followed by 3 additional cycles in a fresh reaction mixture (reconditioning PCR step) to minimize the formation of heteroduplex and *Taq* DNA polymerase errors (32). In addition, we identified *Taq* DNA polymerase errors in sequences from the modified library by manual reconstruction of 16S rRNA secondary structures (1). This accounted for the occurrence of *Taq* DNA polymerase errors to a high degree, with an error rate of 3.3×10^{-5} per nucleotide per duplication, which approximated the theoretically expected error rate of 2×10^{-5} per nucleotide per duplication for the *Taq* DNA polymerase used (1). Finally, a combination of three bioinformatics tools was used to identify putative chimeras in both libraries (1).

Table 1 demonstrates the strong effect of a decrease in cycle numbers from 35 to 15 + 3 on the accumulation of PCR artifacts (1). This is evident from several results, such as (i) a decrease in unique 16S rRNA sequences (ribotypes) from 76% to 48% in the standard and modified libraries after correction for chimeras and *Taq* DNA polymerase errors (1); (ii) a greater-than-twofold decrease in sequence diversity, from 3,881 to 1,633 sequences, among libraries based on the Chao-1 nonparametric richness estimator (1); and (iii) an increase in coverage index from 24% in the standard library to 64% in the modified library (Table 1), suggesting that much fewer clones would have to be sequenced to obtain a representative sample of the modified library.

An important question is to what extent chimeras and heteroduplex molecules contribute to diversity estimates in clone libraries, as these represent amalgams of existing sequences that would be scored as novel lineages. While our experimental design does not allow differentiation between chimeras and heteroduplex molecules, we have shown previously that the inclusion of a "reconditioning" step reduces the occurrence of heteroduplex amplicons to a negligible level (undetectable at a detection level of 1%) (32). Moreover, the combination of three bioinformatics tools suggests that the incidence of chimeras dropped from 13% in the standard library to only 3% in the modified library (Table 1), again suggesting a strong effect of modified amplification protocols (32).

The strong effect of *Taq* DNA polymerase errors can be seen from a comparison of the frequencies of shared and unique sequence types among the two-16S rRNA gene clone libraries

* Corresponding author. Mailing address: Massachusetts Institute of Technology, 48-421, 77 Massachusetts Avenue, Cambridge, MA 02139. Phone: (617) 253-7128. Fax: (617) 258-8850. E-mail: mpolz@mit.edu.

† Supplemental material for this article may be found at http://aem.asm.org/.

TABLE 1. Compositions of the standard and modified 16S rRNA gene clone libraries

Clone library	No. of PCR cycles	No. of sequences analyzed	% Chimeric sequences	No. of ribotypes (100%; 99%[a])	Estimated total no. of sequences[b] (100%; 99%[a])	% Unique 16S rRNA sequences (100%)	Library coverage[c] (100%; 99%[a])
Standard	35	909	13	692; 302	3,881; 962	76	24; 66.5
Modified (+)[d]	18	1,131	ND[g]	686; 326	3,087; 699	61	39.4; 71.2
Modified (−)[e]	18	1,033[f]	3	516; 240	1,633; 520	48	64; 89.3

[a] Values calculated for 100 and 99% sequence similarity cutoffs, respectively.
[b] The total number of sequences was estimated using the Chao-1 nonparametric richness estimator (11).
[c] Library coverage was calculated by the method of Good (9).
[d] Modified library (+) data set analyzed before removal of chimeras and *Taq* error correction.
[e] Modified library (−) data set analyzed after removal of chimeras and *Taq* error correction.
[f] Number of sequences after removal of chimeric sequences and sequences shorter than 600 bp used for rarefaction and phylogenetic analyses. Thus, the number differs slightly from those in Table 2, which includes shorter sequences.
[g] ND, not determined.

(Table 2). For example, there was an almost-twofold-higher incidence of singletons (unique sequences occurring only once) in the standard library (61.5%) compared to that in the modified library (36%) (Table 2). Indeed, if sequences are clustered into 99% consensus groups, ~80% of lineages are shared among the libraries (Table 2).

Given that the two libraries showed considerable differences in sequence diversity, we explored whether a statistical comparison would identify both libraries as being drawn from the same sample. Coverage curves generated with the LIBSHUFF program (Fig. 1) indicated that the modified library would be judged significantly different from the standard library ($P = 0.001$) and that therefore both libraries would be interpreted as samples from different communities ($P < 0.05$). However, the analysis demonstrates that the differences are due to genetic distances of less than 0.01 and that both libraries share most phylogenetic lineages at evolutionary distances (D) greater than 0.02. This result supports the likelihood that *Taq* DNA polymerase errors are the major contributor to artificial sequence divergence. Similar results arose from lineage-versus-time plots (data not shown). We conclude that the incidence of *Taq* DNA polymerase errors is sufficiently accounted for by the clustering of sequences into 99% identity groups, and we recommend that sequence diversity be reported at this similarity cutoff.

The construction of two well-sampled libraries from the same sample gave us the unique opportunity to examine the effect of PCR cycle numbers on the relative distribution of sequence types (i.e., PCR bias). We examined the relative abundances of distinct phylogenetic groups in both libraries by (i) grouping of sequences using neighbor-joining and parsimony methods implemented in the ARB sequence analysis package (17) and (ii) statistical comparison of differences in taxonomic composition of the two libraries using the Ribosomal Database Project Classifier (5). Surprisingly, neither approach showed a significant difference in distribution of major phylogenetic lineages among the libraries (Fig. 2; see also S3 in the supplemental material). The three most abundant groups (*Bacteroidetes*, α-*Proteobacteria*, and γ-*Proteobacteria*) were all similarly represented (32.8% versus 32.5%, 29.3% versus 28.7%, and 29.3% versus 28.7%, respectively) (Fig. 2). These three groups comprise more than 80% of the total clones in each library. However, even the less-well-represented *Actinobacteria* showed only very minor differences in abundance, constituting 8 and 5.5% of the total clones (Fig. 2). Well-represented clades such as SAR11 or *Roseobacter*, which allow comparison with finer taxonomic resolution, also display similar distributions among the libraries (see S3 in the supplemental material).

In order to explore potential reasons for the unexpected similarity of the relative abundances of the ribotypes in the two libraries, we investigated whether the amplification process essentially stopped soon after the 18th cycle due to reagent limitation. Using quantitative PCR (qPCR), we determined the kinetics of product accumulation mimicking PCR conditions utilized for clone library construction (100 nM primer concentration and 5 to 10 ng DNA) (see S4 in the supplemental material). The results show that the reaction was saturated around the 30th cycle, indicating that the modified library underwent ~12 additional cycles of amplification (see S4 in the supplemental material). Because the slopes of the product accumulation curves indicate that the amplification efficiency was between 83 and 88%, and assuming an initial target concentration of $\sim 10^7$ rRNA gene copies ($\sim 4.9 \times 10^6$ genomic templates × ~2.5 rRNA operons per genome [2]), these 12

TABLE 2. Occurrences of shared and unique sequence types within the two 16S rRNA gene clone libraries

Clone library	No. of sequences analyzed	% (no.) of repeated sequences shared within libraries at a sequence similarity cutoff of:		% (no.) of singleton sequences at a sequence similarity cutoff of[a]:		% (no.) of unique clusters at 100% similarity cutoff[b]
		100%	99%	100%	99%	
Standard	902	32.5 (293)	80 (722)	61.5 (555)	13.6 (122)	6 (54)
Modified	1,082	48 (520)	81 (876)	36 (389)	10.7 (116)	16 (173)

[a] Singleton sequence, single sequence unique to one of the libraries.
[b] Unique cluster, cluster containing more than one sequence unique to either one of the libraries.

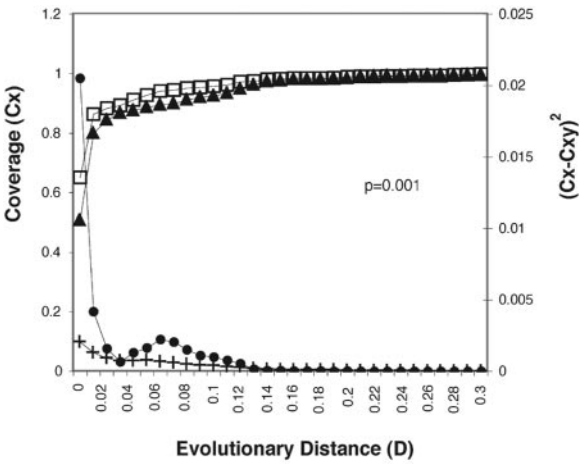

FIG. 1. LIBSHUFF coverage curve comparison for the modified and standard libraries. Shown are the homologous coverage curve (Cx) of the modified library (squares), the heterologous coverage curve (Cxy) of the modified library versus the standard library (triangles), and the value of $(Cx\text{-}Cxy)^2$ (crosses) for the samples at each *D* value. *D* is equal to the Jukes-Cantor evolutionary distance determined using PAUP software. Filled circles indicate the *P* value of 0.05 for $(Cx\text{-}Cxy)^2$ for the randomized samples.

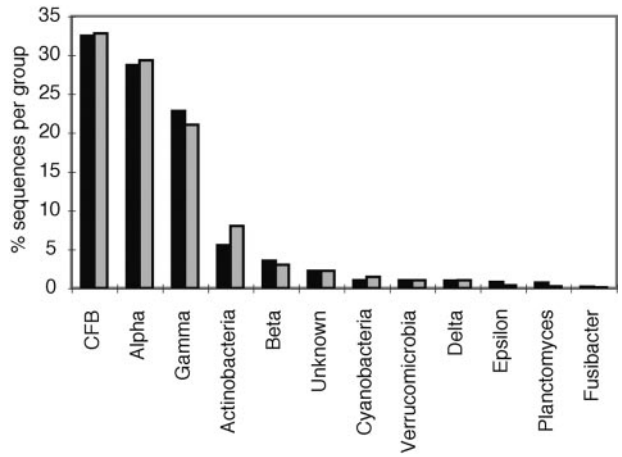

FIG. 2. Relative frequency distribution of major phylogenetic groups detected among the environmental 16S rRNA sequences from the standard (gray bars) and modified (black bars) libraries, respectively.

extra cycles led to an approximately 1,410- to 1,949-fold-higher product concentration. Thus, even a minor difference in amplification efficiency among different templates would have resulted in a noticeable difference in ribotype abundance in the two libraries.

PCR bias was previously attributed to intrinsic differences in amplification efficiencies as a consequence of differences in primer binding energy (12, 23, 35, 36) and to inhibition of amplification due to the reannealing of templates that occurs once they reach saturation concentration (30, 31). Indeed, a primer binding efficiency different from that of genomic templates may still play a significant role in biased amplification but may be lessened during later cycles when amplification proceeds primarily from PCR amplicons, which display a perfect match to the primers. Further, Suzuki et al. already discussed the possibility that PCR bias due to reannealing should be small in environmental DNA samples composed of highly diverse templates because no template may reach saturation concentration (30, 31). Thus, it may not be surprising that several other studies comparing amplicon distributions in community fingerprints (terminal restriction fragment length polymorphism and automated ribosomal intergenic spacer analyses) generated at different cycles similarly suggested an absence of bias (7, 18, 22).

These considerations thus focus attention on the first few cycles of the PCR as a major source of bias to libraries generated from complex environmental samples.

Although we have no direct way of estimating bias during the first few cycles, the distribution of different taxa observed in our clone libraries is strikingly similar to previous quantitative estimates of different bacterioplankton groups. For example, representatives of the *Cytophaga-Flavobacterium-Bacteroides* group (*Bacteroidetes*) accounted for 32% of the sequences in our library, while they were previously detected at ~30% by fluorescent in situ hybridization (FISH) in other coastal samples (6, 13). Similarly, the SAR11 and *Roseobacter* clades comprised 12% and 8% of the total clones in the libraries, respectively. These values are within the ranges observed for 16S rRNA genes based on genomic shotgun sequencing of Sargasso Sea bacterioplankton, where the SAR11 group comprised 4.7 to 37.3% (20, 34), and in FISH enumerations (1 to 32%) (21). The *Roseobacter* clade accounted for 10 to 40% of the total bacterioplankton cells determined by FISH analysis (8) and between 7 to 20% of the total bacterial counts determined using a dilution culture approach in combination with other molecular methods (28). For less-abundant taxa, we confirmed the abundance of *Vibrio*-related sequences in our library by qPCR estimation in the same environment. These made up a total of 1.8% of the clones, which corresponds well to their abundance within total bacterioplankton (33). Nonetheless, the current data stem from a single experimental system with a single primer set, and there are cases where results from qPCR or probing did not agree with the distribution of clones in libraries in similarly complex systems (19, 24).

We summarize and reiterate the following suggestions for minimizing PCR artifacts in environmental clone library construction. (i) To minimize PCR drift, several replicate PCR amplifications should be combined (23, 36). (ii) To minimize chimeras and *Taq* DNA polymerase errors, the smallest possible number of PCR amplification cycles should be carried out (e.g., until a band is barely visible on agarose gels) (23, 30, 31). (iii) To reduce PCR bias, high ramp rates between the denaturation and annealing steps and low annealing temperatures should be used, while long extension times (>180 s) should be avoided (12, 16). (iv) To minimize the presence of heteroduplex molecules, a reconditioning PCR step (i.e., three to five additional PCR cycles using fresh reagent mixture) should be included (32). Finally, our analysis suggests that *Taq* DNA polymerase errors are sufficiently constrained when sequences are clustered into 99% similarity groups so that these groups should always be reported in estimates of sequence diversity, particularly since higher similarity cutoffs (e.g., the commonly used 97%) may mask microdiverse clusters, which we have recently suggested as representing important units of differentiation among marine bacterioplankton (1).

This work was supported by research grants from NSF-OCE and DOE Genomes-to-Life to M.F.P. and by a partial postdoctoral fellowship from the Spanish Ministry of Education (Ministerio de Educacion, Cultura y Deporte [MECD]) to S.G.A.

We are indebted to David Singleton for his help with and suggestions for running the LIBSHUFF program and to Rima Ann Upchurch for writing PRELIBSHUFF, which allows LIBSHUFF to operate with PAUP outputs on Macintosh OSX. We also thank Thomas Huber for help with running Bellerophon to detect chimeric sequences, Ivan Ceraj for writing the chimera (ChimeraBuster) and clustering programs, and Daniel Distel for his input and discussion on lineage-per-time plots. Finally, we acknowledge Jim Cole and Qiong Wang, who ran the Ribosomal Database Project Classifier program with our data set.

REFERENCES

1. **Acinas, S. G., V. Klepac-Ceraj, D. E. Hunt, C. Pharino, I. Ceraj, D. L. Distel, and M. F. Polz.** 2004. Fine-scale phylogenetic architecture of a complex bacterial community. Nature **430:**551–554.
2. **Acinas, S. G., L. A. Marcelino, V. Klepac-Ceraj, and M. F. Polz.** 2004. Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. J. Bacteriol. **186:**2629–2635.
3. **Brakenhoff, R. H., J. G. Schoenmakers, and N. H. Lubsen.** 1991. Chimeric cDNA clones: a novel PCR artifact. Nucleic Acids Res. **19:**1949.
4. **Cline, J., J. C. Braman, and H. H. Hogrefe.** 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res. **24:**3546–3551.
5. **Cole, J. R., B. Chai, R. J. Farris, Q. Wang, S. A. Kulam, D. M. McGarrell, G. M. Garrity, and J. M. Tiedje.** 2005. The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res. **33:**D294–D296.
6. **Cottrell, M. T., and D. L. Kirchman.** 2000. Community composition of marine bacterioplankton determined by 16S rRNA gene clone libraries and fluorescence in situ hybridization. Appl. Environ. Microbiol. **66:**5116–5122.
7. **Fisher, M. M., and E. W. Triplett.** 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. Appl. Environ. Microbiol. **65:**4630–4636.
8. **González, J. M., R. Simó, R. Massana, J. S. Covert, E. O. Casamayor, C. Pedrós-Alió, and M. A. Moran.** 2000. Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. Appl. Environ. Microbiol. **66:**4237–4246.
9. **Good, I. J.** 1953. The population frequencies of species and the estimation of population parametres. Biometrika **40:**237–264.
10. **Hugenholtz, P., and T. Huber.** 2003. Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. Int. J. Syst. Evol. Microbiol. **53:**289–293.
11. **Hughes, J. B., J. J. Hellmann, T. H. Ricketts, and B. J. M. Bohannan.** 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. Appl. Environ. Microbiol. **67:**4399–4406.
12. **Ishii, K., and M. Fukui.** 2001. Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. Appl. Environ. Microbiol. **67:**3753–3755.
13. **Kirchman, D. L., L. Yu, and M. T. Cottrell.** 2003. Diversity and abundance of uncultured *Cytophaga*-like bacteria in the Delaware estuary. Appl. Environ. Microbiol. **69:**6587–6596.
14. **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.
15. **Kopczynski, E. D., M. M. Bateson, and D. M. Ward.** 1994. Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. Appl. Environ. Microbiol. **60:**746–748.
16. **Kurata, S., T. Kanagawa, Y. Magariyama, K. Takatsu, K. Yamada, T. Yokomaku, and Y. Kamagata.** 2004. Reevaluation and reduction of a PCR bias caused by reannealing of templates. Appl. Environ. Microbiol. **70:**7545–7549.
17. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K.-H. Schleifer.** 2004. ARB: a software environment for sequence data. Nucleic Acids Res. **32:**1363–1371.
18. **Lueders, T., and M. W. Friedrich.** 2003. Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and *mcrA* genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. Appl. Environ Microbiol. **69:**320–326.
19. **Luyten, Y. A., J. R. Thompson, Wendy Morrill, M. F. Polz, and D. L. Distel.** Extensive variation in intracellular symbiont community composition among members of a single population of the wood-boring bivalve *Lyrodus pedicellatus* (Bivalvia: Teredinidae). Appl. Environ. Microbiol., in press.
20. **Moran, M. A., A. Buchan, J. M. Gonzalez, J. F. Heidelberg, W. B. Whitman, R. P. Kiene, J. R. Henriksen, G. M. King, R. Belas, C. Fuqua, L. Brinkac, M. Lewis, S. Johri, B. Weaver, G. Pai, J. A. Eisen, E. Rahe, W. M. Sheldon, W. Ye, T. R. Miller, J. Carlton, D. A. Rasko, I. T. Paulsen, Q. Ren, S. C. Daugherty, R. T. Deboy, R. J. Dodson, A. S. Durkin, R. Madupu, W. C. Nelson, S. A. Sullivan, M. J. Rosovitz, D. H. Haft, J. Selengut, and N. Ward.** 2004. Genome sequence of *Silicibacter pomeroyi* reveals adaptations to the marine environment. Nature **432:**910–913.
21. **Morris, R. M., M. S. Rappe, S. A. Connon, K. L. Vergin, W. A. Siebold, C. A. Carlson, and S. J. Giovannoni.** 2002. SAR11 clade dominates ocean surface bacterioplankton communities. Nature **420:**806–810.
22. **Osborn, A. M., E. R. Moore, and K. N. Timmis.** 2000. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ. Microbiol. **2:**39–50.
23. **Polz, M. F., and C. M. Cavanaugh.** 1998. Bias in template-to-product ratios in multitemplate PCR. Appl. Environ. Microbiol. **64:**3724–3730.
24. **Polz, M. F., C. Harbison, and C. M. Cavanaugh.** 1999. Diversity and heterogeneity of epibiotic bacterial communities on the marine nematode *Eubostrichus dianae*. Appl. Environ. Microbiol. **65:**4271–4275.
25. **Qiu, X., L. Wu, H. Huang, P. E. McDonel, A. V. Palumbo, J. M. Tiedje, and J. Zhou.** 2001. Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Appl. Environ. Microbiol. **67:**880–887.
26. **Robison-Cox, J. F., M. M. Bateson, and D. M. Ward.** 1995. Evaluation of nearest-neighbor methods for detection of chimeric small-subunit rRNA sequences. Appl. Environ. Microbiol. **61:**1240–1245.
27. **Ruano, G., and K. K. Kidd.** 1992. Modeling of heteroduplex formation during PCR from mixtures of DNA templates. PCR Methods Appl. **2:**112–116.
28. **Selje, N., M. Simon, and T. Brinkhoff.** 2004. A newly discovered Roseobacter cluster in temperate and polar oceans. Nature **427:**445–448.
29. **Speksnijder, A. G. C. L., G. A. Kowalchuk, S. De Jong, E. Kline, J. R. Stephen, and H. J. Laanbroek.** 2001. Microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences. Appl. Environ. Microbiol. **67:**469–472.
30. **Suzuki, M., M. S. Rappé, and S. J. Giovannoni.** 1998. Kinetic bias in estimates of coastal picoplankton community structure obtained by measurements of small-subunit rRNA gene PCR amplicon length heterogeneity. Appl. Environ. Microbiol. **64:**4522–4529.
31. **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.
32. **Thompson, J. R., L. A. Marcelino, and M. F. Polz.** 2002. Heteroduplexes in mixed-template amplifications: formation, consequence and elimination by 'reconditioning PCR'. Nucleic Acids Res. **30:**2083–2088.
33. **Thompson, J. R., S. Pacocha, C. Pharino, V. Klepac-Ceraj, D. E. Hunt, J. Benoit, R. Sarma-Rupavtarm, D. L. Distel, and M. F. Polz.** 2005. Genotypic diversity within a natural coastal bacterioplankton population. Science **307:**1311–1313.
34. **Venter, C. J., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. H. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, P. C., Y.-H. Rogers, and H. O. Smith.** 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304:**66–74.
35. **Von Wintzingerode, F., U. B. Gobel, and E. Stackebrandt.** 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. **21:**213–229.
36. **Wagner, A., N. Blackstone, P. Cartwright, M. Dick, B. Misof, P. Snow, G. P. Wagner, J. Bartels, M. Murtha, and J. Pendleton.** 1994. Surveys of gene families using polymerase chain reaction: PCR selection and PCR drift. Syst. Biol. **43:**250–261.
37. **Wang, G. C., and Y. Wang.** 1996. The frequency of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from different bacterial species. Microbiology **142:**1107–1114.
38. **Wang, G. C., and Y. Wang.** 1997. Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. Appl. Environ. Microbiol. **63:**4645–4650.