APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Feb. 1994, p. 746–748
0099-2240/94/$04.00+0

Vol. 60, No. 2

Recognition of Chimeric Small-Subunit Ribosomal DNAs Composed of Genes from Uncultivated Microorganisms

ERIC D. KOPCZYNSKI, MARY M. BATESON, AND DAVID M. WARD*

Department of Microbiology, Montana State University, Bozeman, Montana 59717

Received 23 July 1993/Accepted 27 November 1993

When PCR was used to recover small-subunit (SSU) rRNA genes from a hot spring cyanobacterial mat community, chimeric SSU rRNA sequences which exhibited little or no secondary structural abnormality were recovered. They were revealed as chimeras of SSU rRNA genes of uncultivated species through separate phylogenetic analysis of short sequence domains.

PCR has become a popular tool for retrieval from natural environments of small-subunit (SSU) rRNA genes that represent native microbial species (18). However, a potential hazard associated with PCR use is the creation of recombinant ("shuffle-gene") or chimeric products (4, 10, 11, 13, 17). Detection of chimeric SSU rRNA genes is of the utmost importance since they could otherwise lead to erroneous reporting of novel lines of evolutionary descent, give a false impression of biodiversity within microbial communities, and reduce the quality of centralized sequence data bases (10). One method of recognizing chimeras is through abnormal base pairing in helices formed from the two different genes (5, 7, 10). We have found that analyzing the secondary structure of SSU ribosomal DNA alone does not necessarily detect chimeric sequences, since conservation in helices between mixed populations of organisms can be high enough to produce normal SSU rRNA secondary structure. We report two chimeric SSU ribosomal DNA sequences, one of which was not detectable by analysis of secondary structure, from a library of PCR-amplified SSU ribosomal DNAs from a hot spring environment. Chimeras were detected by analysis of changes in patterns of sequence similarity between the suspected chimeras and other known SSU rRNA sequences throughout different regions of the molecule.

A sample of Octopus Spring cyanobacterial mat (50 to 55°C; top 4 mm in the effluent channel) was collected, frozen with liquid N_2, transported to the laboratory on dry ice, and stored at −70°C until it was resuspended, mixed with 1 M NaCl, 5 mM $MgCl_2$, and 10 mM Tris (pH 8.0) in a tissue homogenizer, and passed through a French pressure cell at 10,000 lb/in^2. The sample was diluted to 0.2 M NaCl and 1 mM $MgCl_2$ in 10 mM Tris (pH 8.0)–1 mM EDTA–1% sodium dodecyl sulfate. Proteinase K (60 μg/ml) was added, and the mixture was incubated for 1 h at 50°C. Nucleic acids were extracted and ethanol precipitated. RNA was removed by using DNase-free RNase A. SSU rRNA genes were amplified by PCR, using primers with polylinker tails, 8 FPL and 1492 RPL (1). Each 100-μl amplification reaction mixture contained 50 ng of DNA; 10 μl of 10× reaction buffer (100 mM Tris-HCl [pH 8.3], 500 mM KCl, 15 mM $MgCl_2$); 5 μl of 1% Nonidet P-40; 10 μl of 50% acetamide; 200 μM (each) dATP, dGTP, dTTP, and dCTP; 2 μg of each primer; and 1 U of Ampli*Taq* DNA polymerase. The reaction mixtures were overlaid with mineral oil and incubated in a thermal cycler for 4 min at 94°C followed by 30 cycles of 92°C (1.5 min), 50°C (1.5 min), and 72°C (2 min). After the final cycle, the samples were incubated for 7 min at 72°C and then held at 4°C. The PCR products were cloned as *Pst*I-*Not*I fragments (16) in pBluescript KS− (Stratagene) and sequenced by using Sequenase (United States Biochemicals) with primers complementary to conserved regions of SSU rRNA (21).

Sequences were aligned to other SSU rRNA sequences derived from our Octopus Spring mat data base and the Ribosomal Database Project on the anonymous ftp server at Argonne National Laboratory in Argonne, Ill., updated 17 February 1993 (12). Two clones of nine which were analyzed were revealed as chimeras. Sequence data (approximately *Escherichia coli* positions 1090 to 1350) were first obtained for all nine clones by using one primer, and a preliminary similarity analysis was made, as previously described (20), to determine phylogenetic placement. When more sequence was determined closer to the 5′ end of the molecule, the phylogenetic placement of the two clones changed dramatically, implying a possible chimera. As near full-length sequences were obtained, searching for chimeric junctions was done by comparing short sequence fragments of the clones with their closest relatives. These fragments were initially about 100 bp at the 3′ and 5′ extremes but increased in size until the chimeric junctions were identified.

Clone OPI-8 is a hybrid between a cyanobacterial sequence previously recovered from Octopus Spring mat, OS type I (20), and another SSU ribosomal DNA sequence, OPI-2, from this PCR library (Fig. 1a). The junction of the chimera was found to be in a semiconserved helix between *E. coli* positions 1219 and 1233. Between these two positions the OS type I and OPI-2 sequences are exactly alike, so that the specific point of chimera formation cannot be determined. The 5′ end of the molecule is nearly identical to the OS type I sequence, whereas the 3′ end is nearly identical to the OPI-2 sequence. No abnormal base pairs were observed in helical regions formed by the two phylogenetically distinct fragments of the chimera (Fig. 1b).

Clone OPI-6 is a hybrid between one sequence type previously recovered from Octopus Spring mat material and a sequence which is phylogenetically related, but not identical, to another Octopus Spring sequence type (Fig. 2a). Clone OPI-6 is nearly identical to OS type K, a spirochete-like sequence (20), between the 5′ end and position 369, after which similarity to the OS type K sequence decreases. The 3′ end is 95.7% similar to OS type E, a sequence type (19) most similar to the sequences of green sulfur bacteria (3). The only abnormality in

* Corresponding author. Mailing address: Department of Microbiology, Montana State University, Bozeman, MT 59717. Phone: (406) 994-3401. Fax: (406) 994-4926. Electronic mail address: DMW@WARD.OSCS.MONTANA.EDU.

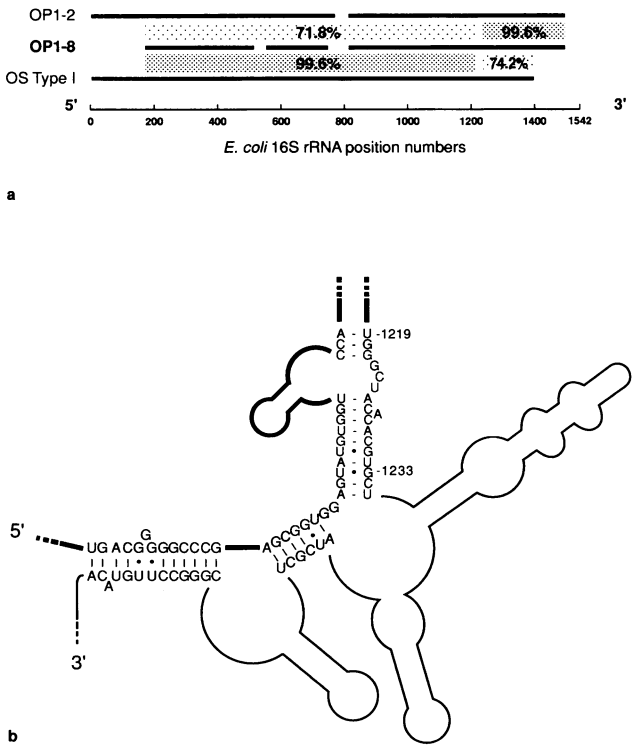

FIG. 1. (a) Percent similarity between fragments of chimeric clone OPI-8 and sequence types OS type I and OPI-2. (b) Partial secondary-structure diagram of chimeric clone OPI-8; the OPI-2-like portion is indicated by thin lines, and the OS type I-like portion is indicated by thick lines. Bases are shown for helical regions formed by the two fragments of the chimera.

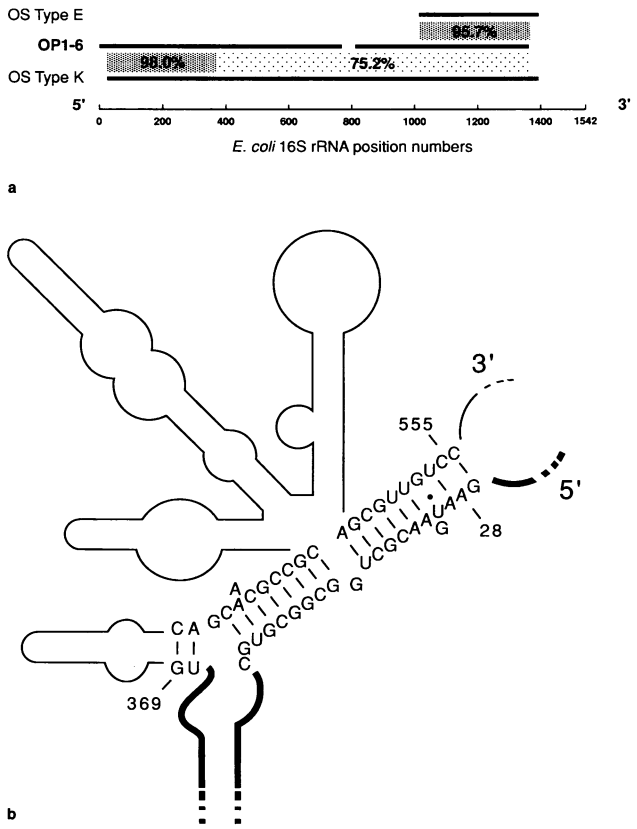

FIG. 2. (a) Percent similarity between fragments of chimeric clone OPI-6 and sequence types OS type K and OS type E. (b) Partial secondary-structure diagram of chimeric clone OPI-6; the OS type K-like portion is indicated by thick lines, and the OS type E-like portion is indicated by thin lines. Bases are shown for helical regions formed by the two fragments of the chimera.

base pairing detected in the helical regions formed by the two phylogenetically distinct fragments of the chimera was a single-base-pair mismatch at *E. coli* positions 28 and 555 (Fig. 2b).

SSU rRNA chimeras are thought to arise from the interaction of different SSU rRNA genes during PCR (10). The tendency toward chimera formation is thought to be increased when PCR is performed on fragmented DNA, such as that obtained by French press lysis methods like the one we used. However, chimeras have also been detected at similar frequency (i.e., 10 to 30%) in SSU rRNA libraries generated by PCR amplification of DNA obtained by more gentle detergent (10) or enzymatic (2, 14) lysis. Size fractionation after any lysis method may help to decrease chimera generation (2), but it is apparent that more must be learned about chimera prevention.

It is in any case crucial to be able to test SSU rRNA sequences for their possible chimeric nature. Detection of chimeric SSU rRNA sequences by secondary structural abnormalities is not fail-safe. One of the chimeras that we observed could not have been detected by any secondary structural abnormality; the other had only a single abnormality in the helices involving different fragments, which could have been misinterpreted as a single base sequencing error. Detection of such chimeras thus depends on independent analysis of different domains within the same SSU rRNA sequence. In cases when the parental SSU rRNA sequences contributing to the chimera are known, detection can be based on identity between domains of the parental SSU rRNA sequences and the chimera (10). As rRNA methods are applied to the analysis of naturally occurring microorganisms, the detection of chimeras will not be so straightforward, because natural environments have been shown to contain mainly uncultivated species with novel SSU rRNA sequences (6, 7, 9, 19). The detection of the chimeras described in this study was aided by the fact that each contained fragments that were nearly identical to sequences of uncultivated species that had been detected in this mat before. Chimeras formed from uncultivated species whose SSU rRNA sequences have not yet been discovered can be recognized only by establishing different phylogenetic affiliations (as opposed to identity) for separate sequence domains.

An automated procedure for detection of chimeric SSU rRNA sequences through similarity analysis (CHECK_CHIMERA) has recently become available through the Ribosomal Database Project (8), and we recommend that it be used. This program succeeded in detecting both chimeras that we report herein, as well as most of those cited above (2, 14). However, we caution that this service may be unable to detect all chimeras, since our preliminary investigation of a similar method suggests that the ability to detect chimeric SSU rRNAs decreases as the parental sequences become more similar (15). We are currently attempting to establish the limitations of both methods. For novel SSU rRNA sequences which have no close relatives in the data base, it may also be difficult to prove through similarity analysis that different domains have the same phylogenetic affiliation. The statistical power of methods

that rely on tree analysis of short sequence domains (6, 10) should also be evaluated.

Nucleotide sequence accession numbers. Sequence data are available from GenBank under the accession numbers L22045 (clone OPI-2), L22046 (clone OPI-6), L22047 (clone OPI-8), L04709 (OS type I), and L04711 (OS type K).

This work was supported by grants from the U.S. National Science Foundation (BSR-9209677) and the U.S. National Aeronautics and Space Administration (NAGW-2764).

We thank the U.S. National Park Service for permission to conduct research in Yellowstone National Park.

REFERENCES

1. **Angert, E. R., K. D. Clements, and N. R. Pace.** 1993. The largest bacterium. Nature (London) **362:**239–241.
2. **Barns, S., and N. R. Pace (Indiana University).** 1993. Personal communication.
3. **Bateson, M. M., E. D. Kopczynski, and D. M. Ward.** Unpublished results.
4. **Brakenhoff, R. H., J. G. G. Schoenmakers, and N. H. Lubsen.** 1991. Chimeric cDNA clones: a novel PCR artifact. Nucleic Acids Res. **19:**1949.
5. **Fuhrman, J. A., K. McCallum, and A. A. Davis.** 1992. Novel major archaebacterial group from marine plankton. Nature (London) **356:**148–149.
6. **Fuhrman, J. A., K. McCallum, and A. A. Davis.** 1993. Phylogenetic diversity of subsurface marine microbial communities from the Atlantic and Pacific Oceans. Appl. Environ. Microbiol. **59:**1294–1302.
7. **Giovannoni, S. J., T. B. Britschgi, C. L. Moyer, and K. G. Field.** 1990. Genetic diversity in Sargasso Sea bacterioplankton. Nature (London) **345:**60–63.
8. **Larsen, N., G. J. Olsen, B. L. Maidak, M. J. McCaughey, R. Overbeek, T. J. Macke, T. L. Marsh, and C. R. Woese.** 1993. The ribosomal database project. Nucleic Acids Res. **21:**3021–3023.
9. **Liesack, W., and E. Stackebrandt.** 1992. Occurrence of novel groups of the domain *Bacteria* as revealed by analysis of genetic material isolated from an Australian terrestrial environment. J. Bacteriol. **174:**5072–5078.
10. **Liesack, W., H. Weyland, and E. Stackebrandt.** 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:**191–198.
11. **Meyerhans, A., J. Yartanian, and S. Wain-Hobson.** 1990. DNA recombination during PCR. Nucleic Acids Res. **18:**1687–1691.
12. **Olsen, G. J., R. Overbeek, N. Larsen, T. L. Marsh, M. J. McCaughey, M. A. Maciukenas, W. M. Kaun, T. J. Macke, Y. Xing, and C. R. Woese.** 1992. The Ribosomal Database Project. Nucleic Acids Res. **20**(Suppl.)**:**2199–2200.
13. **Paabo, S., D. M. Irwin, and A. C. Wilson.** 1990. DNA damage promotes jumping between templates during enzymatic amplification. J. Biol. Chem. **265:**4718–4721.
14. **Paster, B. (Forsyth Dental Center).** 1993. Personal communication.
15. **Robison-Cox, J., M. M. Bateson, and D. M. Ward.** Unpublished data.
16. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
17. **Shuldiner, A. R., A. Nirula, and J. Roth.** 1989. Hybrid DNA artifact from PCR of closely related target sequences. Nucleic Acids Res. **17:**4409.
18. **Ward, D. M., M. M. Bateson, R. Weller, and A. L. Ruff-Roberts.** 1992. Ribosomal RNA analysis of microorganisms as they occur in nature. Microb. Ecol. **12:**219–286.
19. **Ward, D. M., R. Weller, and M. M. Bateson.** 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature (London) **344:**63–65.
20. **Weller, R., M. M. Bateson, B. K. Heimbuch, E. D. Kopczynski, and D. M. Ward.** 1992. Uncultivated cyanobacteria, *Chloroflexus*-like and spirochete-like inhabitants of a hot spring microbial mat. Appl. Environ. Microbiol. **58:**3964–3969.
21. **Weller, R., J. W. Weller, and D. M. Ward.** 1991. 16S rRNA sequences of uncultivated hot spring cyanobacterial mat inhabitants retrieved as randomly primed cDNA. Appl. Environ. Microbiol. **57:**1146–1151.