

# The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species

Grace C.-Y. Wang and Yue Wang

Author for correspondence: Yue Wang. Tel: +65 7726861. Fax: +65 7791117.  
e-mail: mcbwangy@leonis.nus.sg

Microbial Collection and Screen Laboratory, Institute of Molecular and Cell Biology, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260

**Our understanding of microbial diversity is greatly hampered by the inability to culture as much as 99% of the microbial community in the biosphere. Development of methods for identification and determining microbial phylogenies based on gene sequences, and for recovering genes directly from diverse environmental samples has made it possible to study microbes without the need for cultivation. PCR techniques have revolutionized retrieval of conserved gene sequences. However, it is well known that co-amplification of homologous genes may generate chimeric sequences leading to descriptions of non-existent species. To quantify the frequency of chimeric sequences in PCR amplification of 16S rRNA genes, we chose several 16S rDNAs with known sequences and mixed them for PCR amplifications under various conditions. Using this model system, we detected 30% occurrence of chimeric sequences after 30 cycles of co-amplification of two nearly identical 16S rRNA genes. The frequency of chimera formation decreased to 12.9% and 14.7% for templates with 82% and 86% similarity, respectively. We also examined effects of the number of amplification cycles, length of elongation periods and presence of damaged DNA on chimera formation. The results should provide useful information for microbiologists who use PCR-based strategies to retrieve conserved genes from mixed genomes.**

**Keywords:** PCR, 16S rRNA gene, chimeric molecules, microbial diversity

## INTRODUCTION

There is consensus among microbiologists that only a small fraction of natural micro-organisms are culturable under laboratory conditions (Amann *et al.*, 1995; Barns *et al.*, 1994; Giovannoni *et al.*, 1990; Ward *et al.*, 1990). Culture-dependent studies of microbial communities may generate highly biased representations of the compositions of these communities. Nucleic acid sequence analyses of conserved genes in establishing phylogenetic relationships of organisms have made it possible to identify known species and predict novel organisms without laboratory cultivation (Amann *et al.*, 1995; Woese *et al.*, 1985; Woese, 1987). Nucleic acids of mixtures of organisms can be extracted from environmental samples and sequences of appropriate marker genes, such as 16S rRNA genes, are obtained by a series of molecular manipulations (Giovannoni *et al.*, 1990; Tsai & Olson, 1992; Ward *et al.*, 1990; Weller & Ward, 1989). The origins of such sequences are determined by comparison with known sequences in databases. Such strategies have

proven powerful in the investigation of microbial diversity and resolution of microbial community structures (Bruce *et al.*, 1992; Liesack & Stackebrandt, 1992; Schuppler *et al.*, 1995; Ward *et al.*, 1990; Weller & Ward, 1989).

PCR-based strategies have been the most powerful because they allow highly specific recovery and convenient cloning of complete or selected regions of 16S rRNA genes (Barns *et al.*, 1994; Liesack & Stackebrandt, 1992; Reysenbach *et al.*, 1994; Weller *et al.*, 1991). However, this strategy has two serious drawbacks. Firstly, some genes may be preferentially amplified so that the frequency of a sequence occurring in a 16S rDNA library prepared from an environmental sample does not reflect its relative abundance in the microbial community (Reysenbach *et al.*, 1992). Secondly, when the 16S rRNA genes of more than one species is PCR-amplified in a single reaction, chimeric or recombinant molecules may be generated which consist of mixtures of sequences from different 16S rRNA genes (Kopczynski *et al.*, 1994;

Liesack *et al.*, 1991). Thus, without careful examination of the derived sequences, artifactual sequences can lead to description of non-existent organisms. Normally, it is possible to identify chimeric molecules by examining the base complementarity within the helical regions of rRNA, or by using the program CHECK-CHIMERA provided by the Ribosomal Database Project (Larsen *et al.*, 1993). However, if a chimeric sequence is a result of fusion between closely related species, it may not be possible to identify the component sequences (Kopczynski *et al.*, 1994; Robison-Cox *et al.*, 1995).

In previous studies, the frequency of chimeric molecules in PCR products has been estimated to range from 4% to 20% (Robison-Cox *et al.*, 1995). In an effort to quantify the PCR-generated artifacts, Meyerhans *et al.* (1990) demonstrated that when two 275 bp DNA fragments of HIV1 *tat* genes with 85% identity were mixed for co-amplification under standard conditions, 5.4% of the amplified molecules were chimeric after 30 cycles. It is reasonable to assume that the frequency of chimera formation in PCR is a function of the length and the sequence similarities of the amplified DNA molecules. Therefore, PCR co-amplification of the 1550 bp 16S rRNA genes of different bacterial species would be expected to generate a much higher proportion of chimeric molecules. Furthermore, previous studies (Meyerhans *et al.*, 1990; Liesack *et al.*, 1991; Kopczynski *et al.*, 1994) underestimated the actual frequency of recombination events, as they did not consider recombinations between identical molecules which would not generate new sequences.

The aims of this study were to quantify the frequencies of chimeric molecule formation during PCR co-amplification of different 16S rRNA genes, and to examine the effects on chimera formation of various parameters such as sequence similarities of the input genes, the number of PCR cycles, the presence of damaged DNA and the length of the PCR elongation period.

## METHODS

**16S rRNA gene sequences used in this study.** The 16S rRNA gene sequences and cloned 16S rRNA genes of *Streptosporangium nondiasticum* (IFO13990), *Streptosporangium pseudovulgare* (IFO13991), *Promicromonospora sukumoe* (IFO14650) and *Micromonospora chalcea* (ATCC 12452) were obtained from our Actinomycete 16S rRNA Gene Database Project (sequences available upon request), at the Institute of Molecular and Cell Biology, National University of Singapore.

**Oligonucleotides for PCR and hybridization.** The two universal primers used for PCR amplification of 16S rRNA genes were: 5' TTA CCT GAT AGCGGCCCGC AGA GTT TGA TCC TGG CTC 3' (nt 8–25, *Escherichia coli* numbering, Brosius *et al.*, 1981) and 5' TAC AGG ATC CGCGGCCCGC TAC GG(CT) TAC CTT GTT ACG ACT T 3' (nt 1492–1513). The underlined nucleotides indicate *NotI* sites for cloning. Non-specific nucleotides were added at the 5' end of each primer to facilitate efficient cleavage by the restriction enzyme. The oligonucleotides used for mutagenesis (see below) were: 5' CGGGATCCA GAG TTT GAT CCT GGC TCT CGA CGA

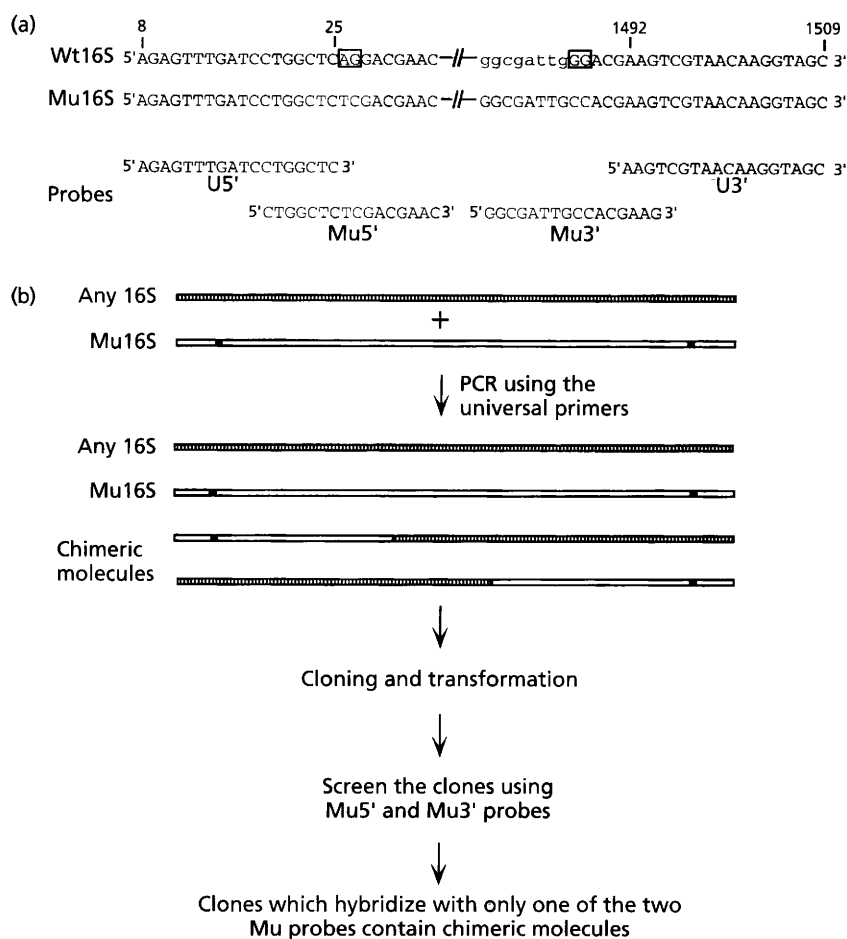
ACG CT 3' (nt 8–37, *E. coli* numbering) and 5' CGGGATCCT ACG GTT ACC TTG TTA CGA CTT CGT GGC AAT CGC C 3' (nt 1479–1513). The underlined nucleotides are *Bam*HI sites for cloning. The two universal probes U5' and U3' and the two Mu16S-specific probes Mu5' and Mu3' are shown in Fig. 1. The oligonucleotides were synthesized by Oligos Etc, Inc.

**Site-specific mutagenesis.** Site-specific mutagenesis was carried out using PCR. Two oligonucleotides containing the mutations (see Fig. 1 for the sites of mutagenesis and above for the sequences) were used in PCR to amplify a mutated product from a cloned 16S rRNA gene of *S. nondiasticum*. The PCR product was cleaved with *Bam*HI, ligated into plasmid vector pUC18 and transformed into *E. coli* DH5 $\alpha$  cells (Gibco BRL). Three white clones were picked and each grown in 2 ml LB medium containing 100  $\mu$ g ampicillin ml<sup>-1</sup> overnight. Plasmid DNA was prepared using a Promega Wizard Minipreps DNA purification system. The mutations were confirmed by sequence analysis.

**PCR amplification of 16S rRNA genes.** The 16S rRNA gene fragments of different species were cut from respective plasmids, purified by electrophoresis using low-melting-point agarose gels (Gibco BRL) and quantified using the Gibco BRL FastCheck Nucleic Acid Quantification System. The purified DNAs were used as templates for PCR. A PCR reaction contained 0.1 ng 16S rDNA as template, 20 pmol each of the two universal primers, 200  $\mu$ M each of dATP, dGTP, dCTP and dTTP, 2.5 U *Taq* DNA polymerase (Amersham) and 1  $\times$  *Taq* buffer provided by the supplier. The final volume was 100  $\mu$ l. The PCR program was 95 °C for 40 s, 52 °C for 30 s and 72 °C for 2 min or 5 min. The number of cycles varied with different experiments. The PCR products were purified using a QIAquick Spin PCR Purification Kit (Qiagen).

**Cloning and transformation.** The purified PCR products from 16S rRNA genes were cut with *NotI* (Amersham) and cloned into pBluescript SK (Stratagene). *E. coli* DH5 $\alpha$  cells (Gibco BRL) were prepared and transformed as described by Maniatis *et al.* (1982).

**Detection of chimeric PCR molecules.** *E. coli* colonies (white) containing cloned DNA fragments were picked and grown in 50  $\mu$ l LB with 100  $\mu$ g ampicillin ml<sup>-1</sup> in 96-well microtitre plates for 16 h. Cells were lysed and DNA denatured in the wells by addition of an equal volume of a solution containing 1% (w/v) SDS, 3 M NaCl and 1 M NaOH. Each lysate (20  $\mu$ l) was blotted onto Hybond-N membrane (Amersham) using a slot-blot apparatus (Hoefer Scientific Instruments). When a large number of clones needed to be screened, the colonies were directly transferred to a membrane by lifting from agar plates and a duplicate membrane was prepared as described by Maniatis *et al.* (1982). The membrane was then baked in an oven at 80 °C for 2 h to immobilize the nucleic acids. Oligonucleotide probes were labelled in a 20  $\mu$ l reaction containing 10 pmol of the oligonucleotide, 10 U T4 DNA polynucleotide kinase (New England Biolabs), 5  $\mu$ l [ $\gamma$ -<sup>32</sup>P]ATP (5000 Ci mmol<sup>-1</sup>, Amersham) and 1  $\times$  buffer (supplied with the kinase) at 37 °C for 30 min. Prehybridization, hybridization and washes were carried out as described by Maniatis *et al.* (1982) in a HB 400 hybridization oven (Hoefer Scientific Instruments). The  $T_m$  (°C) of the four probes U5', U3', Mu5' and Mu3' (Fig. 1) are 54, 52, 52 and 52 °C, respectively as determined by the formula:  $T_m$  (°C) = 4(G + C) + 2(A + T) (Itakura *et al.*, 1984). The hybridization was carried out at 20 °C below the  $T_m$  of each probe overnight in 20 ml hybridization solution containing 6  $\times$  SSC (1  $\times$  SSC is 0.15 M NaCl, 0.015 M sodium citrate), 2  $\times$  Denhardt's solution, 0.25% SDS and 100  $\mu$ g salmon testis DNA ml<sup>-1</sup>. After hybridization, the membrane was washed first



**Fig. 1.** A schematic illustration of the strategy to detect chimeric molecules generated during PCR co-amplification of different 16S rRNA genes. (a) Wt16S is the original sequence of the 16S rRNA gene of *S. nondiasticum* and Mu16S is the mutated sequence. The positions of mutagenesis are boxed. U5' and U3' are universal probes of 16S rRNA genes. Mu5' and Mu3' probes recognize the mutated sites at the 5' end and the 3' end of the Mu16S, respectively. (b) Flow diagram illustrating the procedure used in this study.

at 22 °C in 112 × SSC/0.05 % SDS for 1 h and then in the same solution at 10 °C below the  $T_m$  of the probe for 15 min. The membrane was exposed to X-ray film with intensifying screens for 4 h at -80 °C. For repeated hybridizations with different probes, the membrane was stripped in 2 × SSC containing 50 % formamide and 0.05 % SDS at 68 °C for 30 min.

## RESULTS

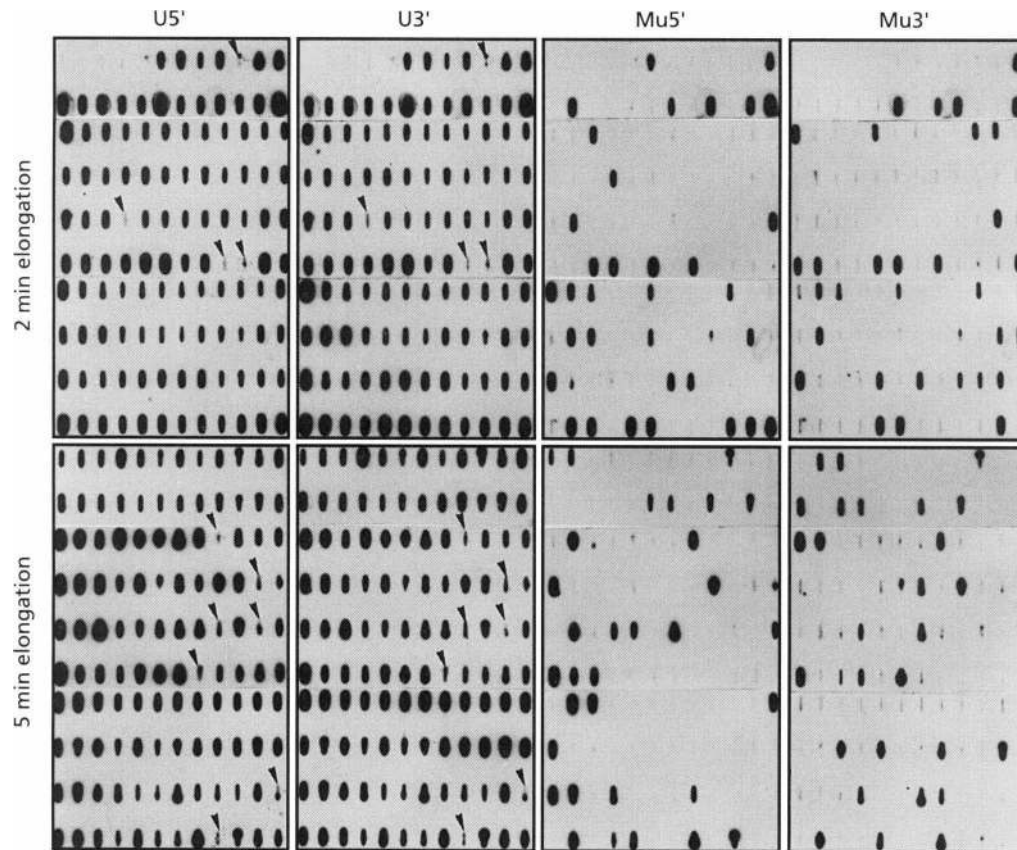
### Strategy

Fig. 1 outlines the strategy for identification of chimeric molecules after PCR co-amplification of 16S rRNA sequences. Two nucleotides near each end of a near full-length 16S rDNA fragment of *S. nondiasticum*, which are invariant among 16S rRNA genes of different bacteria (Liesack & Stackebrandt, 1992), were mutated to generate specific markers. The mutated nucleotides are immediately downstream from the binding sites of the two universal PCR primers (see Methods). The mutated 16S rDNA (named Mu16S) was mixed with different 16S rDNA species in PCR amplifications using the two universal primers. Where recombination occurred during the PCR between the Mu16S molecules and other 16S rDNA molecules, the resultant chimeric molecules would carry only one of the two markers. After cloning of the PCR products and transformation, clones harboring the chimeric molecules were identified by hybridization using

the two specific probes Mu5' and Mu3'. Since one internal mismatch between a short oligonucleotide probe and its target sequence lowers the  $T_m$  by 5.0–7.5 °C (Ikuta *et al.*, 1987), the  $T_m$  of either of the two Mu probes was at least 10 °C higher for its target than for the corresponding sequences in any 16S rRNA genes. The two universal probes, U5' and U3', were used to check the presence of full length 16S rDNA inserts in all clones. Recombination between identical molecules could not be detected directly but can be estimated from the observed frequency of chimeric molecules. A small fraction of molecules which hybridized with both universal or both Mu probes may have been products of multiple recombination events but they could not be distinguished by this strategy. Therefore, the detected frequency of chimeric molecules is an underestimate.

### Chimera formation between nearly identical genes and effect of elongation time

To avoid any effects of genomic DNA on PCR co-amplification of 16S rDNA sequences, we used purified 16S rDNA fragments in our model system. The Mu16S DNA was mixed at a 1:1 ratio with the 99.3 % identical counterpart of *S. pseudovulgare*. A total of 0.1 ng of the mixed 16S rDNAs, roughly equivalent to the amount of 16S rRNA genes in 100 ng genomic DNA, was used as



**Fig. 2.** Autoradiograph of slot-blot DNA hybridizations using different oligonucleotide probes. The presence of complete 16S rDNA in the clones was first examined by separate hybridizations with U5' and U3' probes, then the membranes were probed with Mu5' or Mu3' to detect clones containing chimeric 16S rDNAs. The specificities of the two Mu probes were first tested by hybridizing to Mu16S (results not shown). The membranes were stripped after each hybridization and reprobed. The same procedure was followed for all the experiments summarized in Table 1 and Fig. 4. The length of PCR elongation periods are shown on the left and the names of probes are given along the top. The first five slots in the top lane contained cell lysates from blue colonies which served as negative controls. The arrowheads indicate the clones that do not contain inserts.

template for each PCR reaction. The PCR elongation period was set at 2 or 5 min in two separate reactions to examine the possible effect of elongation time on recombination frequency. The PCR reactions were carried out for 30 cycles and followed by cloning the PCR products. Transformed clones were examined by slot-hybridizations using different oligonucleotide probes (Fig. 2) and the results are summarized in Table 1 (rows C and G). As many as 30% of the clones contained chimeric 16S rDNA. Increasing elongation time of PCR from 2 to 5 min slightly reduced the frequency of chimera formation. This experiment and the experiments described below were carried out in duplicate. In subsequent PCRs only 2 min elongation periods were used.

#### Effect of the number of amplification cycles on chimera formation

To determine the effect of the number of PCR cycles on the occurrence of chimeric sequences, the above experiment was repeated in two PCR reactions with 10 and

20 cycles. Hybridization analyses detected 4.8% (mean of two results) and 20.9% chimeric clones from the 10-cycle and the 20-cycle PCRs (Table 1, rows A and B), respectively. A positive correlation was found between the generation of chimeric molecules and the number of PCR cycles.

#### Effect of sequence similarity on chimera formation

To investigate the relationship between the frequency of chimera formation and sequence similarities, two other 16S rRNA genes of *P. sukumoe* and *M. chalcea*, which have 82% and 86% similarities with the Mu16S sequence, respectively, were used as templates. Mu16S was mixed with the other 16S rDNA at a ratio of 1:1 for 30 cycles of PCR. The frequencies of chimeric molecules detected were 12.9% between Mu16S and 16S rDNA of *P. sukumoe* (Table 1, row H) and 14.7% between Mu16S and 16S rDNA of *M. chalcea* (Table 1, row I). These frequencies were only about half of that observed resulting from the two sequences with 99.3% similarity (Table 1, row C).

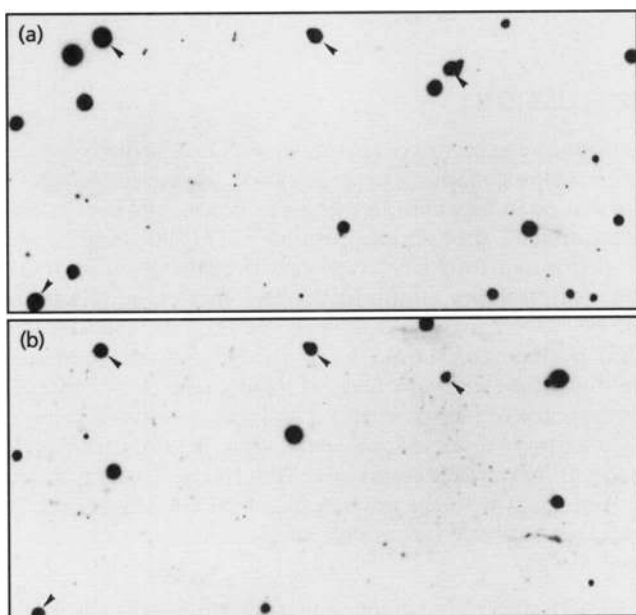
**Table 1.** Occurrence of chimeric 16S rDNA sequences under different PCR conditions

PCR conditions			No. of clones hybridizing with oligonucleotide probes*					Observed frequency of chimeras (%)†
Templates‡	Elongation time (min)	Cycles	U5' & U3'	Mu5' & Mu3'	Mu5' only	Mu3' only	Neither Mu5' nor Mu3'	
A Mu16S+S16S	2	10	92 (94)	19 (23)	2 (2)	2 (3)	69 (66)	4.37 (5.3)
B Mu16S+S16S	2	20	93 (94)	18 (21)	12 (7)	8 (11)	54 (55)	21.5 (20.3)
C Mu16S+S16S	2	30	111 (90)	14 (11)	18 (13)	16 (15)	63 (51)	30.6 (31.1)
D Mu16S+S16S + sheared DNA	2	10	95 (93)	24 (21)	2 (2)	2 (3)	67 (67)	4.20 (5.4)
E Mu16S+S16S + sheared DNA	2	20	94 (93)	17 (15)	11 (8)	9 (10)	57 (60)	21.3 (19.4)
F Mu16S+S16S + sheared DNA	2	30	93 (94)	14 (12)	14 (15)	15 (13)	50 (54)	31.2 (29.8)
G Mu16S+S16S	5	30	120 (93)	16 (14)	18 (11)	14 (15)	75 (53)	26.6 (27.9)
H Mu16S+P16S	2	30	125 (116)	12 (15)	8 (9)	7 (7)	98 (85)	12.0 (13.8)
I Mu16S+M16S	2	30	92 (91)	2 (5)	7 (7)	7 (5)	76 (74)	15.2 (14.2)

\* Figures in parentheses are results from a duplicate set of experiments.

† Observed frequency = (Mu5' only + Mu3' only)/U5' & U3'.

‡ S16S, 16S rDNA of *S. pseudovulgare*; P16S, 16S rDNA of *P. sukumoe*; M16S, 16S rDNA of *M. chalcone*.



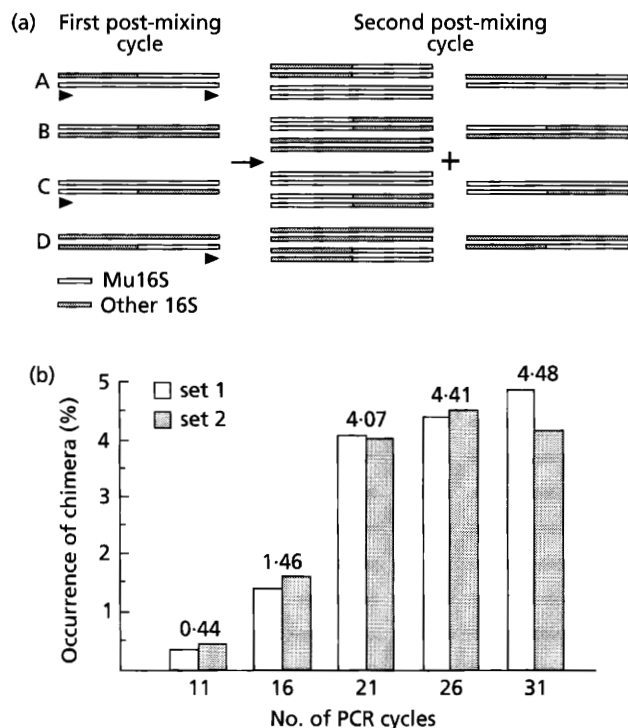
**Fig. 3.** Autoradiograph of colony hybridizations using the two Mu probes, Mu5' (a) and Mu3' (b). The arrowheads point to colonies that hybridize with both Mu probes. Since only a small fraction of clones would hybridize with the Mu probes, for us to obtain a significant number of positive clones hundreds of colonies were transferred and duplicated on nylon membranes by lifting from large agar plates (20 × 20 cm), followed by hybridization with each of the two Mu probes. Because the universal probes generated too many signals of different size on the X-ray films and it was difficult to distinguish true and false signals, the positive clones revealed by the Mu probes were isolated and cultured and then confirmed for the presence of complete inserts using different probes (results not shown).

### Effect of abundance of 16S rDNA species on chimera formation

The above observations demonstrated up to 30% recombination events during PCR co-amplification of highly related sequences. This prompted us to speculate that if a 16S rDNA species was present as a minor fraction, after 30 or more cycles of amplification the minor species might disappear as a result of chimera formation. To test this hypothesis, we mixed the Mu16S with 16S rDNA of *S. pseudovulgare*, at ratios of 1:5 and 1:10 for 30 cycles of PCR. Fig. 3 is an autoradiograph of the result of the PCR where the two templates were mixed 1:5. Out of 274 clones, 25 clones were identified as positive because of their hybridization with one or both Mu probes. Only four clones hybridized with both Mu probes while the rest hybridized with only one of the Mu probes. Among a total of 392 clones from the product of the PCR in which templates were mixed 1:10, only one clone was found to hybridize with both Mu probes while 25 clones hybridized with only one of the Mu probes (results not shown).

### Effect of sheared DNA on chimera formation

The formation of chimeric molecules has been attributed to the presence of damaged DNA (Liesack *et al.*, 1991; Paabo *et al.*, 1992; Shuldiner *et al.*, 1989). To examine the effect of mechanically sheared DNA on the formation of chimeric molecules, we added to the PCR reactions an equal amount of 16S rDNA that had been fragmented by sonication for 10, 20 and 30 cycles of amplification. The power and duration of the sonication was chosen such that most of the 1.5 kb DNA fragment was converted to a smear as observed on an agarose gel (results not shown).



**Fig. 4.** (a) A diagram illustrating the generation of chimeric DNAs in the first and second cycles of PCR amplification after mixing two individually amplified 16S rDNAs. In the first post-mixing PCR cycle, a partial DNA strand hybridizes to a homologous complementary template to prime DNA synthesis leading to a partially heteroduplex chimeric molecule. Four possible configurations of the partial heteroduplexes are shown. Since both Mu probes are complementary to the same strand of the 16S rDNA (the small arrows denote the locations where the Mu probes can hybridize), they will recognize configurations C and D as chimeric, but not configurations A and B. The second post-mixing cycle converts all the heteroduplexes to homoduplexes. Partial heteroduplexes are also generated in the second PCR cycle. With the assumption that the second cycle would produce partially heteroduplex DNA at a similar frequency to the first, the frequency of chimeric sequence per cycle was derived by subtracting the frequency observed after one post-mixing cycle from that observed after two post-mixing cycles. (b) Percentage occurrence of chimeric sequences per PCR cycle for *Streptosporangium* 16S rRNA genes. For each selected cycle, 96 clones of PCR products were examined as described in the legend to Fig. 2. The frequency per cycle was calculated as described in the text. Results were obtained from two independent sets of experiments and the mean percentages are shown on the top of the bars.

We did not observe any appreciable increase in the occurrence of chimeric sequences (Table 1, rows D, E and F) by comparison with the results of PCR without the damaged DNA (Table 1, rows A, B and C).

#### Frequency of chimera formation per PCR cycle

In theory, the amount of target DNA for PCR amplification doubles with each cycle (Innis & Gelfand, 1990; Saiki, 1989). However, in practice, the concentration of amplified DNA will initially increase exponentially and

later gradually enters a plateau when prematurely terminated strands are generated due to depletion of reagents. Based on this known phenomenon, we hypothesized such prematurely terminated DNA strands could anneal with homologous templates in subsequent PCR cycles to prime DNA synthesis and cause formation of chimeric molecules. If our hypothesis was correct, we expected to observe an increase of the frequency of chimeric molecules per cycle with the progress of PCR.

To determine the frequency of chimera formation at a particular cycle of PCR, we separately amplified 0.1 ng of each of the Mu6S and the 16S rDNA of *S. pseudovulgare* for 10, 15, 20, 25 and 30 cycles, and then mixed the two PCR reactions for one or two more cycles of co-amplification. As illustrated in Fig. 4(a), the first post-mixing cycle generated partially heteroduplex DNA molecules and the Mu probes would not correctly recognize all the heteroduplex configurations. A second cycle was needed to convert the heteroduplexes to homoduplexes. By doing this, we could also minimize effect of repair of the mismatches by the host cells after transformation. A heteroduplex DNA formed in the first cycle could be involved again in recombination in the second cycle, but the frequency should be very low and therefore was not considered here. Fig. 4(b) shows that the frequency of chimeric sequences was low in early cycles and increased with the progress of PCR, reaching a plateau after 20 cycles.

#### DISCUSSION

Although the results of this study may not be directly used to extrapolate the frequencies of chimeric sequence formation in other studies where experimental parameters are different, they should provide useful information for improvement of PCR-based experiments. In our experiments involving amplification of mixed templates of related DNA sequences, misintegration of nucleotides by *Taq* polymerase should have a negligible effect on the frequency of chimera formation, because the chance of occurrence of errors within the target regions of primers is extremely low. Asymmetric amplification from two ends of a template could affect the frequency, but it was not obvious with the few templates we used (see Table 1, roughly equivalent numbers of clones detected by Mu5' and Mu3').

Chimeric 16S rRNA gene sequences are usually identified by either checking the base complementarity within paired regions (Giovannoni *et al.*, 1990) or by using software programs such as CHECK-CHIMERA available through the Ribosomal Database Project (Barns *et al.*, 1994; Choi *et al.*, 1994; Kopczynski *et al.*, 1994; Moyer *et al.*, 1995). The frequency of chimeric sequences has been reported to range from 4 to 20%, but both methods underestimate the frequency. Checking base pair mismatches has been proven not to be fail-safe because some chimeras do not exhibit such abnormalities (Kopczynski *et al.*, 1994). Robison-Cox *et al.* (1995) carried out a mathematical evaluation of the CHECK-CHIMERA program and found that the confidence in detection of chimeras by this

method decreased from 95 to 50% as the estimated similarity between template DNA parental sequences increased from 82 to 96%.

By using oligonucleotide probes to detect exchanges of fragments between parental DNA molecules, we were able to detect chimeras formed during co-amplification of almost identical sequences. The observed frequencies of 12.9 and 30% between sequences of 82 and 99.7% identity, respectively, are in agreement with the reported range. Meyerhans *et al.* (1990) reported a maximum frequency of 5.4% between two 275 bp DNA fragments with 85% identity, and Choi *et al.* (1994) reported a frequency of 8.6% when a 500 bp region was amplified from DNA extracted from a subgingival plaque sample of a patient with severe periodontitis. These seemingly low frequencies can be explained by the much shorter DNA templates used for amplification.

When using two or a small number of templates for PCR to quantify the frequency of chimera formation, the observed frequency represents an underestimate of actual frequency because recombination frequently happens between identical molecules that does not lead to new sequences and will not be detected. When many homologous sequences are amplified, such as in DNA samples extracted from soil where even abundant bacterial species may account for only small fractions of the microbial population (Torsvik *et al.*, 1990), prematurely terminated, partial-length DNA molecules will rarely anneal with molecules of the same species and the recombination events could be maximally expressed as chimeric molecules. This hypothesis is supported by our observation that when two 16S rDNA fragments showing 99.3% similarity were mixed 1:10, nearly all the molecules of the less abundant species recombined because the partial DNA strands of this species had only 10% chance to reanneal with its own species. Therefore, higher frequencies of chimera formation should be expected when DNA with high complexity is used for PCR.

Using our model system, we did not detect any significant effect of sheared DNA on the frequency of chimera formation. This observation seemed to be contrary to some previously published reports (Liesack *et al.*, 1991, Paabo *et al.*, 1992; Shuldiner *et al.*, 1989); however, the results of these experiments may not be directly compared because DNA templates of different nature were used. We cannot rule out the possibility that damaged DNA could play a major role in chimera formation when a large fraction of the templates is damaged. However, when a DNA sample contains a large fraction of intact templates, prematurely terminated DNA strands generated during PCR are more likely to be the major cause of chimera formation for the following reasons. Firstly, as we observed, the frequency of chimera formation was higher in later cycles than in early cycles. Secondly, prolonged elongation periods, limiting premature termination, could reduce the frequency (Meyerhans *et al.*, 1990 and this work). Thirdly, in the early cycles of PCR the molar concentration of damaged DNA is negligible in comparison with those of the primers. The damaged DNA

would therefore not compete with the primers for annealing to the template and its relative amount in the total DNA will drop rapidly with the exponential increase of the amplified DNA.

### Concluding remarks

PCR has been and is likely to remain the most powerful tool in retrieving DNA sequences of culturable and unculturable micro-organisms from environmental samples. Awareness of the intrinsic problems of the technique, elucidation of the possible mechanisms and investigation of critical parameters will lead to improved protocols minimizing artifacts of the PCR-based strategies. The results of this study suggest it is advisable to use the least possible number of amplification cycles, longer elongation times, and take greater caution in analysis of sequences PCR-amplified from complex genomes.

### ACKNOWLEDGEMENTS

This work was supported by the Institute of Molecular and Cell Biology (IMCB), National University of Singapore. We thank Professor Julian Davies, Department of Microbiology, University of British Columbia, Drs Catherine J. Pallen and Hans Ulrich Bernard of the IMCB for critical reading of the manuscript.

### REFERENCES

- Amann, R., Ludwig, W. & Scheifer, K. H. (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59**, 143–169.
- Barns, S. M., Fundyga, R. E., Jeffries, M. W. & Pace, N. R. (1994). Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. *Proc Natl Acad Sci USA* **91**, 1609–1613.
- Brosius, J., Dull, T. J., Sleeter, D. D. & Noller, H. F. (1981). Gene organization and primary structure of a ribosomal RNA operon from *Escherichia coli*. *J Mol Biol* **148**, 107–127.
- Bruce K. D., Hiorns, W. D., Hobman, J. L., Osborn, A. M., Strike, P. & Ritchie, D. A. (1992). Amplification of DNA from native populations of soil bacteria by using the PCR. *Appl Environ Microbiol* **58**, 3413–3416.
- Choi, B. K., Paster, B. J., Dewhirst, F. E. & Gobel, U. B. (1994). Diversity of cultivable and uncultivable oral Spirochetes from a patient with severe destructive periodontitis. *Infect Immun* **62**, 1889–1895.
- Giovannoni, S. J., Britschgi, T. B., Moyer, C. L. & Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**, 60–63.
- Ikuta, S., Takagi, K., Wallace, R. B. & Itakura, K. (1987). Dissociation kinetics of 19 base paired oligonucleotide–DNA duplexes containing different single mismatched base pairs. *Nucleic Acids Res* **15**, 797–805.
- Innis, M. A. & Gelfand, D. H. (1990). Optimization of PCRs. In *PCR Protocols: a Guide to Methods and Applications*, pp. 3–12. Edited by M. A. Innis, D. H. Gelfand, J. J. Sninski & T. J. White. San Diego: Academic Press.

- Itakura, K., Rossi, J. J. & Wallace, R. B. (1984).** Synthesis and use of synthetic oligonucleotides. *Annu Rev Biochem* **53**, 323–340.
- Kopczynski, E. D., Bateson, M. M. & Ward, D. M. (1994).** Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. *Appl Environ Microbiol* **60**, 746–748.
- Larsen, N., Olsen, G. J., Maidak, B. L., McCaughey, M. J., Overbeek, R., Macke, T. J., Marsh, T. L. & Woese, C. R. (1993).** The Ribosomal Database Project. *Nucleic Acids Res* **21**, 3021–3023.
- Liesack, W. & Stackebrandt, E. (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.
- Liesack, W., Weyland, H. & Stackebrandt, E. (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.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982).** *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Meyerhans, A., Vatanian, J. P. & Wain-Hobson, S. (1990).** DNA recombination during PCR. *Nucleic Acids Res* **18**, 1687–1690.
- Moyer, C. L., Dobbs, F. C. & Carl, D. M. (1995).** Phylogenetic diversity of the bacterial community from a microbial mat at the active, hydrothermal vent system, Loihi Seamount, Hawaii. *Appl Environ Microbiol* **61**, 1555–1562.
- Paabo S., Irwin, D. M. & Wilson, A. C. (1992).** DNA damage promotes jumping between templates during enzymatic amplification. *J Biol Chem* **265**, 4718–4721.
- Reysenbach, A. L., Giver, L. J., Wickham, G. S. & Pace, N. R. (1992).** Differential amplification of rRNA genes by polymerase chain reaction. *Appl Environ Microbiol* **58**, 3417–3418.
- Reysenbach, A. L., Wickham, G. S. & Pace, N. R. (1994).** Phylogenetic analysis of the hyperthermophilic pink filament community in Octopus Spring, Yellowstone National Park. *Appl Environ Microbiol* **60**, 2113–2119.
- Robison-Cox, J. F., Bateson, M. M. & Ward, D. M. (1995).** Evaluation of nearest-neighbor methods for detection of chimeric small-subunit rRNA sequences. *Appl Environ Microbiol* **61**, 1240–1245.
- Saiki, R. K. (1989).** The design and optimization of the PCR. In *PCR Technology: Principles and Applications for DNA Amplifications*, pp. 7–16. Edited by H. A. Erlich. New York: Stockton Press.
- Schuppler, M., Mertens, F., Schon, G. & Gobel, U. B. (1995).** Molecular characterization of nocardioform actinomycetes in activated sludge by 16S rRNA analysis. *Microbiology* **141**, 513–521.
- Shuldiner, A. R., Nirula, A. & Roth, J. (1989).** Hybrid DNA artifact from PCR of closely related target sequences. *Nucleic Acids Res* **17**, 4409.
- Torsvik V., Goksoyt, J. & Daae, F. L. (1990).** High diversity in DNA of soil bacteria. *Appl Environ Microbiol* **56**, 782–787.
- Tsai, Y. L. & Olson, B. H. (1992).** Rapid method for separation of bacterial DNA from humic substances in sediments for PCR. *Appl Environ Microbiol* **58**, 2292–2295.
- Ward, D. M., Weller, R. & Bateson, M. M. (1990).** 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* **345**, 63–65.
- Weller, R. & Ward, D. M. (1989).** Selective recovery of 16S rRNA sequences from natural microbial communities in the form of cDNA. *Appl Environ Microbiol* **55**, 1818–1822.
- Weller, R., Weller, J. W. & Ward, D. M. (1991).** 16S rRNA sequences of uncultivated hot spring cyanobacterial mat inhabitants retrieved as randomly primed cDNA. *Appl Environ Microbiol* **57**, 1146–1151.
- Woese, C. R. (1987).** Bacterial evolution. *Microbiol Rev* **51**, 221–271.
- Woese, C. R., Stackebrandt, E., Macke, T. J. & Fox, G. E. (1985).** A phylogenetic definition of the major eubacterial taxa. *Syst Appl Microbiol* **6**, 143–151.

Received 11 September 1995; revised 10 January 1996; accepted 17 January 1996.