Evaluation of primers and PCR conditions for the analysis of 16S rRNA genes from a natural environment

Yuichi Hongoh a,*, Hiroe Yuzawa a, Moriya Ohkuma b,c, Toshiaki Kudo a,c,d

a International Cooperative Research Project, Japan Science and Technology Corporation (JST-ICORP), Saitama, Japan
b Molecular Microbial Ecology Division, Bioscience Technology Center, RIKEN (The Institute of Physical and Chemical Research), Wako, Saitama 351-0198, Japan
c Precursory Research for Embryonic Science and Technology, Japan Science and Technology Corporation (JST-PRESTO), Saitama, Japan
d Division of Environmental Molecular Biology, Graduate School of Integrated Science, Yokohama City University, Kanagawa 230-0045, Japan

Received 3 February 2003; received in revised form 10 March 2003; accepted 11 March 2003

First published online 3 April 2003

Abstract

We investigated biases occurring in the polymerase chain reaction (PCR) amplification of 16S rRNA genes from an environmental sample, by comparing the clone libraries that we had previously prepared from the gut homogenate of the termite Reticulitermes speratus. We detected a significant increase in the expected number of phylotypes by lowering the annealing temperature, and a significant decrease in the proportion of clones belonging to the predominant group by raising the number of PCR cycles. We also found that the Bacteria-universal primer, 63F, introduced a seriously biased amplification caused by primer mismatches, in contrast to a previous report. These results, together with suggestions from previous studies using simplified model samples, will help us to recognize the limitations of PCR-based analysis.

Keywords: 16S rDNA; Phylogenetic diversity; Termite; Gut bacterium; Rarefaction

1. Introduction

Polymerase chain reaction (PCR)-based analysis of 16S rRNA genes is a powerful and essential tool for studies of bacterial diversity, community structure, evolution and taxonomy. It has enabled us to detect and identify as-yet unculturable bacteria, and in recent years has led to an enormous increase in our knowledge of bacterial ecology and taxonomy. However, the PCR procedure intrinsically results in biases and errors, attributable to complex factors such as preferential annealing between primers and templates, self-annealing between PCR products, different copy numbers of the targeted genes and the formation of chimeric sequences [1–7]. For these reasons, we need to interpret PCR experimental data with great caution.

In order to estimate the effect of the parameters involved in PCR, such as the sequence of a primer, the annealing temperature and the number of PCR cycles, some studies have used a simplified model system with heterogeneous templates comprised of only 2–10 species [1,3–6]. These studies suggested that the annealing temperature should be low enough to suppress the amplification bias caused by primer mismatches, and that the number of PCR cycles should be reduced to minimize the effects of primer mismatches and self-annealing between PCR products. In addition, Marchesi et al. [8,9] proposed the new Bacteria-universal primer set, 63F-1387R (later modified to 1389R by Osborn et al. [10]), as an alternative to the generally used 27F-1392R and 27F-1492R primers. Whereas this new primer set contains mismatches against some bacterial groups, they observed that it greatly improved the efficiency of amplification, and amplified targets from a wider diversity of monocultured bacterial species, including mismatched targets, over 27F-1392R and 27F-1492R primers. Therefore, Marchesi et al. predicted that the new set would minimize amplification bias derived from the preferential annealing of primers. While all of these
observational suggestions seem reasonable, it may be naive to assume that these predictions hold in a study using environmental samples, which contain hundreds of bacterial species, without first testing for amplification bias. However, there have yet to be any reports, except for a study using length heterogeneity analysis by PCR [11], on the effect of these factors when using environmental samples, as opposed to a simplified model system.

In our previous study [12], we investigated the phylogenetic diversity of gut bacteria from the termite Reticulitermes speratus, and accumulated data from 16S rDNA clone libraries, prepared with various PCR primer sets and conditions, in order to evaluate the effect of the primer sequences, annealing temperature and number of PCR cycles. Here, we report our evaluation of several Bacteria-universal primers for 16S rRNA genes, annealing temperatures and numbers of PCR cycles, using environmental samples.

2. Materials and methods

2.1. 16S rDNA clone libraries

In our previous study [12], we established 14 clone libraries of 16S rDNA from the gut homogenate of the termite R. speratus (Isoptera, family Rhinotermitidae), and sequenced 96 clones from each of the libraries (Table 1). In the present study, we also prepared an additional two 16S rDNA clone libraries from the same DNA sample used in the previous study [12]. The PCR was performed under the following conditions: 2 min initial denaturation at 95°C; 24 cycles of denaturation (30 s at 95°C), annealing (1 min at 55°C), and extension (2 min at 72°C); a final extension at 72°C for 10 min. The primer sets 27F-1389R and T63F(5'-CAGGCCTAACACATGCAAGTT-3')-1389R were used to amplify the region corresponding to 64–1388 in Escherichia coli (J01695). The other conditions and procedures were as described previously [12]. We randomly chose 420 and 159 clones from the new libraries, prepared using 63F-1389R and T63F-1389R, respectively, and analyzed them by restriction fragment length polymorphism (RFLP) using the restriction enzymes HhaI and MboI (Takara). The restriction fragments were electrophoresed on 2.5% agarose gels (classic type; Nacalai) and sorted according to their RFLP patterns. A representative clone of each RFLP type was sequenced and sorted into a phylotype using the criterion of 97% sequence identity as described previously [12]. All of the sequences were subjected to the Check-Chimera program on the Ribosomal Database Project (RDP) web site [13]. After identifying the sequences, using FASTA 3.0 [14], phylotypes that we did not detect in our previous study were analyzed phylogenetically as described previously [12]. The taxonomic affiliation was conducted using the criterion of 80% bootstrap confidence level.

The sequence data of the new phylotypes will appear in the DDBJ, EMBL and GenBank nucleotide sequence databases under accession numbers, AB100456–99.

2.2. Statistical analysis

The diversity of the clone libraries was estimated by rarefaction analysis [15], using the software Analytic Rarefaction 1.3 (Steven M. Holland, University of Georgia), which is freely distributed at the web site http://www.uga.edu/strata/software/. Analysis was conducted according to the software’s manual. The coverage [16] was calculated by the formula \(1 - (n/N)\), where \(n\) is the number of phylotypes represented by only one clone and \(N\) is the total number of clones. The calculation was conducted using the program LIBSHUFF [17] as described previously [12].

Table 1

| Clone library | Primer set | Annealing temperature (°C) | Number of PCR cycles | Concentration of PCR products (mM) | Number of phylotypes |
|---------------|------------|-----------------------------|----------------------|-----------------------------------|----------------------|
| A             | 27F-1389R  | 50                          | 12                   | 1.0                               | 51                   |
| B             | 27F-1389R  | 50                          | 18                   | 4.2                               | 57                   |
| C             | 27F-1389R  | 55                          | 12                   | 1.0                               | 47                   |
| D             | 27F-1389R  | 55                          | 18                   | 4.0                               | 53                   |
| E             | 27F-1429R  | 50                          | 12                   | no data                           | 50                   |
| F             | 39F-1389R  | 50                          | 18                   | 6.6                               | 48                   |
| G             | 39F-1389R  | 55                          | 18                   | 5.0                               | 48                   |
| H             | 41F-1389R  | 50                          | 18                   | 4.7                               | 51                   |
| J             | 41F-1389R  | 55                          | 18                   | 4.7                               | 49                   |
| K             | 63F-1389R  | 45                          | 18                   | no data                           | 66                   |
| M             | 63F-1389R  | 50                          | 18                   | 5.5                               | 58                   |
| N             | 63F-1389R  | 55                          | 18                   | 5.3                               | 50                   |
| P             | 64F-1389R  | 50                          | 18                   | 5.4                               | 52                   |
| Q             | 64F-1389R  | 55                          | 18                   | 4.7                               | 51                   |

* Ninety-six clones were analyzed from each of the clone libraries [12].
* The sequences and references were described previously [12].
* Concentration was calculated from \(A_{260}\) measured with a DU530 spectrophotometer (Beckman). The measurement was done after purification using the QIAquick PCR Purification kit (QIAGEN).
3. Results and discussion

3.1. Evaluation of Bacteria-universal primers

In concordance with the findings of Marchesi et al. [8], 63F-1389R produced PCR products more efficiently than the primer 27F (Table 1). However, in contrast to Marchesi et al. [8], we found almost no spirochetal clones from libraries K, M and N, prepared using 63F-1389R, whereas they were the most abundant clones in the other libraries (Fig. 1). Even the predominant phylotype, Treponema Rs-D73 (AB088857) [12], was rarely found in libraries K, M and N. Since there were no or only a few mismatches against the dominant groups (spirochetes, Bacteroides, clostridia and Termite Group I (TG-I)), with primers 27F, 39F and 64F, it is unlikely that these primers amplified spirochetes preferentially as the predominant clones. Therefore, it is probable that this was caused by preferential annealing, due to six to eight mismatches between 63F and the corresponding region of the spirochetal sequences (Table 2). Considering the fact that 63F-1387R was able to amplify spirochetes well from monocultures [8], it should be noted that confirmation of a PCR primer set amplifying a target from a monoculture does not ensure unbiased amplification from heterogeneous templates.

From this point of view, a mismatch at the 3' end between a PCR primer and its target seems more critical, even though Marchesi et al. [8] reported that 63F-1387R retained its competence when a mismatch at the 3' end existed. In our study, no clones of Rs-D17 (AB089048) [12], one of the dominant TG-I phylotypes, were found from libraries K, M or N, whereas Rs-D17 clones were found from all of the libraries prepared using the other primer sets. Since the other dominant TG-I phylotypes, Rs-D95 (AB089049) and Rs-D43 (AB089050) [12], were found abundantly from K, M and N, the lack of Rs-D17 is solely attributable to the mismatch at the 3' end, that is, between cytosine in 63F and thymine in Rs-D17 (Table 2). In addition, when we investigated 420 clones from the new clone library, prepared using 63F-1389R, we found no Rs-D17 clones again. On the other hand, when we used T63F of which the 3' end was converted from cytosine in 63F to thymine, we found that 88 of the 159 analyzed clones were those Rs-D17. The result indicates that a single mismatch at the 3' end of a PCR primer can give a totally different figure of the bacterial community structure.

A possible occurrence of preferential annealing due to mismatches was also found from the other primer sets. In libraries H and J, prepared using 41F-1389R, the number of clones related to Bacteroides was limited to as low as seven and three, respectively (Fig. 1). This is probably due to adenine at the fourth position from the 3' end of 41F mismatching against guanine at the corresponding posi-

![Fig. 1. Relative abundance of 16S rDNA clones in the 14 clone libraries A–Q (modified from Fig. 1 in our previous paper [12]). Ninety-six clones were analyzed for each clone library. The clones were affiliated with the taxonomic groups by phylogenetic analysis in our previous study [12].](https://academic.oup.com/femsle/article-abstract/221/2/299/630719)
tion in Bacteroides (Table 2). In the 480 clones from the five libraries A–E (Table 1), prepared using 27F-1389R, we found no clones of the 8-proteobacterial phylotype Rs-N31 (AB089104), even though it was one of the dominant phylotypes in most of the other clone libraries [12] (Fig. 1). It is possible that this was also caused by preferential annealing of 27F, although we cannot discuss it here due to a lack of information on the sequences in the corresponding region.

These biased amplifications seem to be caused by primer mismatches of the forward primers, because no obvious difference was found between the libraries prepared using the reverse primers 1389R and 1492R (Fig. 1). Furthermore, the biased amplification with 63F or 41F was improved when the degenerate primers 64F and 39F were used, which were designed to match spirochetal and Bacteroides sequences (Fig. 1 and Table 2). These results suggest that we should not underestimate the effect of primer mismatches.

In spite of the inferiority of 63F-1389R as a Bacteria-universal primer set, it is still useful as long as we realize its characteristics. In the case of our study, the use of 63F-1389R enabled us to amplify a great diversity of non-spirochetal clones, which would not have been acquired using the other primers [12]. In the present study, we found a further 44 non-spirochetal clones from the new clone libraries. We obtained the new phylotypes of 34 clostridia (AB100461–94), one Enterococcus (AB100496), one Desulfosporosinus (AB100497), five Bacteroides (AB100456–60), one α-Proteobacterium (AB100498), one TM7 bacterium (AB100499) and one unidentified bacterium (AB100495) clustered with Rs-A23 (AB089057) and Rs-J96 (AB089068). Therefore, the total number of 16S rDNA phylotypes from the gut of R. speratus now becomes 312 (44+268). In addition, as reported by Marchesi et al. [8], 63F-1389R is a very effective primer set for amplifying monocultured bacteria; we often experienced that it worked very well even for samples where 27F-1392R or 1389R at 45°C and 55°C, respectively, a significant difference in the expected diversity of detectable phylotypes was demonstrated by rarefaction analysis (Fig. 2). The number of phylotypes found from library K was 66, much greater than the 50 in library N (Table 1). This is concordant with a report using a model system [4], which found that a lower annealing temperature tended to allow annealing at a mismatched site. However, the effect of the 10°C-shutdown of annealing temperature, from 55°C to 45°C, was limited to the increase of diversity within the range.

Table 2
Alignment of sequences in the region corresponding to the PCR primers 39F, 41F, 63F and 64F.

| Sequence name (Acc. No.) | Sequence |
|-------------------------|----------|
| 39F primer | TGGCTCAGRWYGAACGCTRG |
| 41F primer | --GCTCAGATGGAAACGCTGGC |
| 63F primer | --GCTCAGATTGAACGCTGGC |
| 64F primer | --GCTCAGATTGAACGCTGGG |

E. coli (J01695) | TGGCTCAGATGGAAACGCTGGC |
Treponema pallidum (M88726) | TGGCTCAGACGAAACGCTGGCCG |
Treponema Rs-D73 (AB088857) | --ACGAAACGCTGGCCG |
Treponema Rs-G65 (AB088858) | --CAGAAACGCTGGCCG |
Treponema Rs-A48 (AB088859) | --ACGAAACGCTGGCCG |
Bacteroides fragilis (M61006) | TGGCTCAGATGGAAACGCTGGC |
Bacteroides Rs-B92 (AB088922) | --ATGGAAACGCTGGAGC |
Bacteroides Rs-D66 (AB088923) | --ATGGAAACGCTGGAGC |
TG-I* Rs-D95 (AB089049) | --ATGGAAACGCTGGAGC |
TG-I Rs-D17 (AB089048) | --ATGGAAACGCTGGAGC |

*TG-I: the candidate division Termite Group I.

Fig. 2. Rarefaction plots for clone libraries K, M and N, prepared using primers 63F and 1389R. Each clone library was generated from the PCR products, using 45, 50 or 55°C as the annealing temperature [12]. The expected numbers of phylotypes are plotted with 95% confidence intervals. The expected numbers of phylotypes were significantly different (P < 0.05) between K and N, in the range 50 < (number of analyzed clones) < 96.

3.2. Effect of annealing temperature and number of PCR cycles

For all of the primer sets, the number of phylotypes found from the clone libraries was the same or greater when the annealing temperature was lowered (Table 1). When comparing libraries K and N, prepared using 63F-1389R at 45°C and 55°C, respectively, a significant difference in the expected diversity of detectable phylotypes was demonstrated by rarefaction analysis (Fig. 2). The number of phylotypes found from library K was 66, much greater than the 50 in library N (Table 1). This is concordant with a report using a model system [4], which found that a lower annealing temperature tended to allow annealing at a mismatched site. However, the effect of the 10°C-shutdown of annealing temperature, from 55°C to 45°C, was limited to the increase of diversity within the range.
less than a 0.15 nucleotide difference per site (Fig. 3). This means that the 10°C-shiftdown of annealing temperature did not allow an increase of division-level diversity as, indeed, shown as the absence of spirochetal clones in libraries K, M and N (Fig. 1). This implies that, when primer mismatches are severe, a considerable shiftdown of annealing temperature cannot compensate the mismatches fully.

We also found an effect of reducing the PCR cycle number from 18 (libraries B and D) to 12 (libraries A and C) (Table 1). The abundance of spirochetal clones, the predominant group in these libraries, increased more in libraries A and C than in B and D (Fig. 1). The tendency was most obvious in the comparison between A and B, both prepared using the lower annealing temperature of 50°C, where there was a significant increase from 42 to 57% in spirochetal abundance (Fisher’s exact probability test, \(P = 0.0430\)), when the number of cycles was reduced. Suzuki and Giovannoni [1,11] reported that biased amplification occurred when the concentration of PCR products exceeded 2 mM, probably due to self-annealing between the PCR products. The effect was more pronounced as the concentration increased, resulting in a substantial decrease in the proportion of predominant products. Since the concentration of the PCR products in our study for libraries A and C was 1.0 mM after 12 cycles, and 4.2 and 4.0 mM for libraries B and D, after 18 cycles, respectively (Table 1), the difference in abundance of spirochetal clones is probably for the same reason. In this case, the frequency of clones observed in library A should be more representative of the original composition than those in library B. The smaller difference in spirochete fragments between C and D, compared to between A and B, may be attributed to the lower probability of self-annealing between heterogeneous spirochetal fragments at the higher annealing temperature.

The decrease in spirochetal abundance in libraries B and D resulted in an increase in phylogenetic diversity of the clones. The numbers of phylotypes were 57 and 53 in libraries B and D, greater than 51 and 47 in libraries A and C, respectively (Table 1). The rarefaction plots indicated the greater expected diversity of the clones in the libraries prepared using 18 cycles than in those prepared using 12 cycles (Fig. 4), although the differences were significant (\(P < 0.05\)) only in the range from 90 to 96 analyzed clones. This tendency was parallel with the reports of Suzuki and Giovannoni, which found that the increase of PCR cycle number resulted in an overestimation of phylogenetic diversity in the samples [1,11].

These results suggest that a primer best matched with diverse bacterial groups, a lower annealing temperature and a decreased number of PCR cycles should be used to minimize amplification bias, as implied by the previous studies using model samples [1,4]. Although the PCR-based analysis of a microbial community cannot be free from biases and errors completely, the method is still essential to the studies of microbes. It is important that we continue to optimize PCR conditions to obtain the highest quality data.

Acknowledgements

We thank Ms. Chutima Disyen, Ms. Khajarin Sirihongsuwan and Ms. Chanpen Boontong for assisting with sequencing. This work was partially supported by grants for the Bioarchitect Research Program and the Eco Molecular Science Research Program from RIKEN.

References

[1] Suzuki, M.T. and Giovannoni, S.J. (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol. 62, 625–630.
[2] Wang, G.C. and Wang, Y. (1996) The frequency of chimeric molecules as a consequence of PCR co-amplification of 16S rRNA genes from different bacterial species. Microbiology 142, 1107–1114.
[3] von Wintzingerode, F., Gobel, U.B. and Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. FEMS Microbiol. Rev. 21, 213–229.
[4] Ishii, K. and Fukui, M. (2001) Optimization of annealing temperature to reduce bias caused by a primer mismatch in multitemplate PCR. Appl. Environ. Microbiol. 67, 3753–3755.
[5] Qiu, X., Wu, L., Huang, H., McDonel, P.E., Palumbo, A.V., Tiedje, J.M. and Zhou, J. (2001) Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Appl. Environ. Microbiol. 67, 880–887.
[6] Schmalenberger, A., Schwieger, F. and Tebbe, C.C. (2001) Effect of primers hybridizing to different evolutionarily conserved regions of the small-subunit rRNA gene in PCR-based microbial community analyses and genetic profiling. Appl. Environ. Microbiol. 67, 3557–3563.
[7] Speksnijder, A.G., Kowalchuk, G.A., De Jong, S., Kline, E., Stephen, J.R. and Laanbroek, H.J. (2001) Microvariation artifacts introduced by PCR and cloning of closely related 16S rRNA gene sequences. Appl. Environ. Microbiol. 67, 469–472.
[8] Marchesi, J.R., Sato, T., Weightman, A.J., Martin, T.A., Fry, J.C., Hiom, S.J., Dymock, D. and Wade, W.G. (1998) Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. Appl. Environ. Microbiol. 64, 795–799.
[9] Marchesi, J.R., Weightman, A.J., Cragg, B.A., Parkes, R.J. and Fry, J.C. (2001) Methanogen and bacterial diversity and distribution in deep gas hydrate sediments from the Cascadia Margin as revealed by 16S rRNA molecular analysis. FEMS Microbiol. Ecol. 34, 221–228.
[10] Osborn, A.M., Moore, E.R. and Timmis, K.N. (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.
[11] Suzuki, M., Rappe, M.S. and Giovannoni, S.J. (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.
[12] Hongoh, Y., Ohkuma, M. and Kudo, T. (2003) Molecular analysis of bacterial microbiota in the gut of the termite Reticulitermes speratus (Isoptera; Rhinotermitidae). FEMS Microbiol. Ecol., in press.
[13] Maidak, B.L., Cole, J.R., Lilburn, T.G., Parker, C.T.J., Saxman, P.R., Farris, R.J., Garrity, G.M., Olsen, G.J., Schmidt, T.M. and Tiedje, J.M. (2001) The RDP-II (Ribosomal Database Project). Nucl. Acids Res. Res. 29, 173–174.
[14] Lipman, D.J. and Person, W.R. (1985) Rapid and sensitive protein similarity searches. Science 227, 1435–1441.
[15] Heck, K.L., VanBelle Jr., G. and Simberloff, D. (1975) Explicit calculation of the rarefaction diversity measurement and determination of sufficient sample size. Ecology 56, 1459–1461.
[16] Good, I.J. (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40, 237–264.