The Journal of Microbiology (2012) Vol. 50, No. 6, pp. 1071–1074
DOI 10.1007/s12275-012-2642-z

NOTE

Effects of PCR Cycle Number and DNA Polymerase Type on the 16S rRNA Gene Pyrosequencing Analysis of Bacterial Communities[§]

Jae-Hyung Ahn, Byung-Yong Kim, Jaekyeong Song, and Hang-Yeon Weon*

Agricultural Microbiology Division, National Academy of Agricultural Science, Rural Development Administration, Suwon 441-707, Republic of Korea

(Received November 19, 2012 / Accepted December 11, 2012)

The effects of PCR cycle number and DNA polymerase type on 16S rRNA gene pyrosequencing analysis were investigated using an artificially prepared bacterial community (mock community). The bacterial richness was overestimated at increased PCR cycle number mostly due to the occurence of chimeric sequences, and this was more serious with a DNA polymerase having proofreading activity than with *Taq* DNA polymerase. These results suggest that PCR cycle number must be kept as low as possible for accurate estimation of bacterial richness and that particular care must be taken when a DNA polymerase having proofreading activity is used.

Keywords: pyrosequencing, 16S rRNA gene, PCR artifacts, PCR cycle number, DNA polymerase

Next-generation sequencing (NGS) technologies have enabled us to explore the enormous diversity of microorganisms in various environments at an unprecedented scale. Two types of NGS applications have been used to investigate microbial communities: 16S rRNA amplicon sequencing (Roesch *et al.*, 2007; Galand *et al.*, 2009; Bates *et al.*, 2011; Caporaso *et al.*, 2011; Nam *et al.*, 2011) and whole-genome shotgun sequencing (Qin *et al.*, 2010; Delmont *et al.*, 2012). Whole-genome shotgun sequencing is advantageous in that it provides clues regarding the functional capacity of a microbial community as well as its phylogenetic composition, and is also free from PCR artifacts. However, 16S rRNA amplicon sequencing still has a valuable role to play in understanding the structures and variations of microbial communities (Tringe and Hugenholtz, 2011).

Many studies have, however, reported that 16S rRNA amplicon sequencing using NGS technologies can skew the true community structure through biases (misrepresentation of the relative abundances of microbial populations) and errors (misrepresentation of an actual sequence) (Hamp *et al.*, 2009; Kunin *et al.*, 2010; Lee *et al.*, 2012a; Pinto and Raskin, 2012), which can occur during DNA extraction, PCR amplification, or sequencing. Although it is believed that these artifacts do not have significant effects on the high-level taxonomic distribution and comparison of community structure between different samples (Weinstock, 2011; Pinto and Raskin, 2012), depicting an accurate community structure (the total number of species and the relative abundance of each species) is also important for monitoring the variation of specific microbial groups within a community.

Although numerous data processing measures to remove PCR errors have been proposed (Huse *et al.*, 2007, 2010; Quince *et al.*, 2009; Sun *et al.*, 2009; Edgar *et al.*, 2011; Schloss *et al.*, 2011; Schloss and Westcott, 2011), to our knowledge, there has been no entirely satisfactory one. Moreover, it is important to prevent the PCR errors from occurring. In this study, we investigated the effects of PCR cycle number and DNA polymerase type on the pyrosequencing analysis of a known bacterial community to reduce PCR errors by controlling these parameters.

The 15 bacterial strains (Supplementary data Table S1) were cultured as recommended by the Korean Agricultural Culture Collection (KACC, Korea) and the genomic DNAs were isolated using a Wizard genomic DNA purification kit (Promega, USA). The 16S rRNA genes were PCR-amplified as previously described (Lee *et al.*, 2012b), gel-purified using a QIAquick gel extraction kit (Qiagen, Germany), cloned into the pCR 2.1-TOPO vector, and transformed into *Escherichia coli* TOP10 (Invitrogen, USA). The plasmid DNA was extracted using an Accuprep nano-plus plasmid mini extraction kit (Bioneer, Korea), checked by sequencing, and quantified using a micro-volume spectrophotometer ASP-2680 (ACTgene, USA). The 15 plasmids containing the 16S rRNA genes were combined in the final concentrations presented in Supplementary data Table S1 and the combined mixture (mock community) was used as a template for PCR. Bacterial 16S rRNA genes were amplified using the primers V1-9F (5′-X-AC-GAGTTTGATCMTGGCTCAG-3′) and V3-541R (5′-X-AC-WTTACCGCGGCTGCTGG-3′) as previously described (Ahn *et al.*, 2012) except for the DNA polymerase and PCR program. PCR was performed using High-fidelity Phusion DNA polymerase (Phusion DP, New England Biolabs, UK) or *Taq* DNA polymerase (*Taq* DP, Roche, Germany). The

*For correspondence. E-mail: why@korea.kr; Tel: +82-31-290-8478; Fax: +82-31-290-8488
[§]Supplemental material for this article may be found at http://www.springer.com/content/120956.

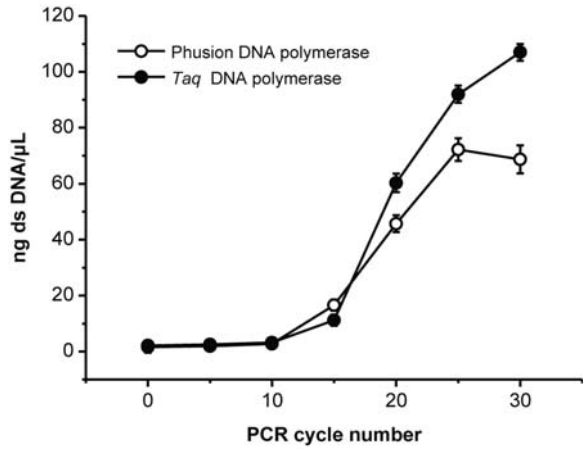

Fig. 1. PCR amplification curves when the mock community was amplified with Phusion DNA polymerase or *Taq* DNA polymerase. Error bars indicates standard deviations of two measurements.

PCR program consisted of an initial denaturation step (5 min at 94°C) followed by 30 cycles of denaturation (94°C for 30 sec), annealing (55°C for 30 sec), and extension (72°C for 60 sec) and a final extension step (72°C for 5 min). The PCR amplification curves (Fig. 1) were constructed by sacrificing the replicate PCR tubes at every 5 cycles and quantifying them using a Qubit dsDNA BR assay kit (Invitrogen, USA). The concentration of PCR products amplified with Phusion DP reached to the saturation level at earlier cycles than with *Taq* DP. The amount of the PCR product was even slightly reduced at 30 cycles compared with that at 25 cycles with Phusion DP. This was consistently observed in the independent experiments irrespective of DNA template and attributed to the 3′→5′ exonuclease activity of Phusion DP based on our communication with the manufacturer. To examine the effect of PCR cycle number, the PCR products of each 15 and 30 cycle reaction with two types of polymerase were selected for subsequent pyrosequencing analysis. The PCR products were gel-purified with the QIAquick Gel extraction kit (Qiagen) and subjected to pyrosequencing, which was performed by National Instrumentation Center for Environmental Management (Seoul, Korea) using a 454 GS FLX Titanium Sequencing System (Roche), according to the manufacturer's instructions.

The analysis of pyrosequecing data was performed using the Mothur software package (version 1.27.0) (Schloss *et al.*, 2009) and a pipeline based on previous studies (Schloss *et al.*, 2011; Schloss and Westcott, 2011) with some modifications. Pre-filtered flowgrams of the pyrosequencing reads were clustered using the PyroNoise algorithm (Quince *et al.*, 2011) and chimeric sequences were removed using UCHIME (Edgar *et al.*, 2011). UCHIME was performed in *de novo* mode and reference database mode in which, the mock community sequences were used as the reference database. The pair-wise distances were calculated using the "pairwise.seqs" command and the sequences were then clustered into operational taxonomic units (OTUs) at distances from 0.00 to 0.05 using the average-neighbor algorithm. The mock community sequences were included in the clustering step to examine how many pyrosequencing reads were clustered together with the mock community sequences. The raw pyrosequencing data are available in the NCBI Sequence Read Archive under the accession number SRA059986.

Table 1. The numbers of pyrosequencing reads before and after chimera removal

Chimera removal	PCR cycle number			
	Phusion		*Taq*	
	15	30	15	30
Before	10,839	6,118	15,293	6,763
After	10,771	3,927	15,257	4,594
Chimeras identified (%)	0.6	35.8	0.3	32.1

Table 2. The proportion of pyrosequencing reads (%) clustered with the mock community sequences at various distances

Distance	PCR cycle number			
	Phusion		*Taq*	
	15	30	15	30
0.00	38.0	32.5	36.2	21.6
0.01	99.1	95.0	99.1	95.2
0.02	99.9	97.2	99.8	98.6
0.03	100.0	98.3	99.9	99.4
0.04	100.0	98.8	100.0	99.7
0.05	100.0	99.2	100.0	99.7

The number of PCR cycles significantly affected the proportion of identified chimeric sequences (Table 1), which were reduced from 32 or 36% to below 1% by decreasing the PCR cycle number from 30 to 15. This observation is consistent with previous studies (Judo *et al.*, 1998; Acinas *et al.*, 2005).

We found that the 16S rRNA gene sequence of *E. coli* was recovered from all four datasets. It was believed that the genomic DNA of the competent *E. coli* was co-extracted with the plasmid DNA. Thus, we added the *E. coli* 16S rRNA gene sequence to the list of the mock community sequences as the 16th member. The pair-wise similarities between the mock community sequences ranged from 70.2% to 95.4% in the amplified region. Thus, ideally, the pyrosequencing reads produced from the mock community should have yielded 16 OTUs up to a distance of 0.04.

The proportions of pyrosequencing reads clustered with the mock community sequences at a distance of 0.00 were low for both types of DNA polymerase, ranging from 21.6%

Table 3. The number of OTUs calculated at various distances

Distance	PCR cycle number							
	Before normalization				After normalization[a]			
	Phusion		*Taq*		Phusion		*Taq*	
	15	30	15	30	15	30	15	30
0.00	211	329	1,230	666	104	327	427	587
0.01	44	139	99	146	30	139	43	130
0.02	24	89	36	63	19	88	22	57
0.03	19	58	21	35	17	58	18	33
0.04	17	46	19	28	16	46	17	27
0.05	14	31	17	23	14	31	16	22

[a] The number of pyroseqencing reads was normalized to 3,900.

to 38.0% while it were increased to above 95.0% at distances above 0.01 (Table 2). The small portion (<5.0%) of the sequences not clustered with the mock community sequences at distances above 0.01, however, had a significant effect on the number of OTUs (Table 3, before normalization). At a distance of 0.01, the numbers of OTUs were 44 and 139 at cycle numbers of 15 and 30, respectively, with Phusion DP and they were 99 and 146 with *Taq* DP. At a distance of 0.03, which was the most frequently used distance in bacterial community analysis using 16S rRNA gene sequences, the numbers of OTUs decreased to 19 and 58 at cycle numbers 15 and 30, respectively, with Phusion DP, and 21 and 35 with *Taq* DP. This is consistent with previous observations in which the species richness was significantly overestimated by minor erroneous sequences (so called "rare biospheres") (Reeder and Knight, 2009; Kunin *et al.*, 2010). Visual inspection of the pyrosequencing reads that were not clustered with the mock community sequences at 30 cycles and a distance of 0.03 showed that most of the sequences (93% and 95% for Phusion DP and *Taq* DP, respectively) were the chimeric sequences, resulting from recombination between the mock community sequences (Supplementary data Table S2). These sequences were not filtered although we used UCHIME, which is a chimera detection program that showed the best performance in several studies (Edgar *et al.*, 2011; Schloss *et al.*, 2011; Wright *et al.*, 2012). At 15 cycles, the numbers of OTUs observed at a distance of 0.03 were slightly higher than the expected one: they were 19 and 21 with Phusion DP and *Taq* DP, respectively. Visual inspection of the sequences which were not clustered with the mock community sequences at 15 cycles revealed that most of them were the mock community sequences with error rates >3% (Supplementary data Table S2). This result indicates that some of pyrosequencing reads can have error rates >3% and thus can be recorded as distinct OTUs at a distance of 0.03 even at reduced PCR cycle numbers.

Because the number of qualified pyrosequencing reads was markedly different among the samples (Table 1), it was normalized to the smallest number of pyrosequencing reads among the samples: 3,900 of the pyrosequencing reads were re-sampled 1,000 times without replacement and the numbers of OTUs observed were averaged (Table 3, after normalization). The results showed that PCR cycle number had a significant effect on the number of OTUs. When PCR cycle number increased from 15 to 30, the number of OTUs increased from 17 to 58 with Phusion DP and from 18 to 33 with *Taq* DP at a distance of 0.03, indicating that PCR cycle number must be kept low for accurate estimation of bacterial richness. An increase in PCR errors in higher PCR cycle numbers was observed in many studies (Judo *et al.*, 1998; Qiu *et al.*, 2001; Acinas *et al.*, 2005). At 15 cycles, the numbers of OTUs were lower with Phusion DP than with *Taq* DP but the difference was small at distances above 0.03. This indicates that the proofreading activity of the high-fidelity DNA polymerase has no significant effect on 16S rRNA gene pyrosequencing analysis if the pyrosequencing reads are clustered at a distance above 0.03. On the contrary, the numbers of OTUs were higher with Phusion DP than with *Taq* DP at 30 cycles except for at a distance of 0.00. The proportion of the identified chimeric sequences at 30 cycles was higher with Phusion DP than with *Taq* DP (Table 1) and so was the proportion of the OTUs which consisted of the undetected chimeric sequences (39/58=67% vs. 18/35=51%, Supplementary data Table S2). We think that this is due to the earlier occurrence of PCR saturation with Phusion DP than with *Taq* DP (Fig. 1) because it is believed that the occurrence of chimerc sequences will be more frequent under the condition where PCR components are limited. Two studies showed that the DNA polymerases with proofreading activity enhance the formation of chimeric sequences compared to *Taq* DNA polymerase (Judo *et al.*, 1998; Zylstra *et al.*, 1998). Judo *et al.* (1998) attributed this to the earlier saturation of the PCR with the DNA polymerases with proofreading activity than with *Taq* DNA polymerase, which is consistent with our hypothesis.

The present study showed that bacterial richness was overestimated at increased PCR cycle number mostly due to the occurrence of chimeric sequences, and this was more serious with a DNA polymerase having proofreading activity than with *Taq* DNA polymerase. Our results also suggested that even at reduced PCR cycle numbers, the pyrosequencing reads with error rate >3% can occur and thus inflate bacterial richness when sequences are clustered at a distance of 0.03. Although the decrease in PCR cycle number will also reduce the amount of PCR product, it is recommended that PCR cycle number should be kept as low as possible for more accurate estimation of bacterial richness.

This study was made possible by the support of the "Research Program for Agricultural Science & Technology Development (Project No. PJ907127032012)" of the National Academy of Agricultural Science, Rural Development Administration, Republic of Korea.

References

Acinas, S.G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., and Polz, M.F. 2005. PCR-induced sequence artifacts and bias: insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl. Environ. Microbiol.* **71**, 8966–8969.

Ahn, J.-H., Hong, I.-P., Bok, J.-I., Kim, B.-Y., Song, J., and Weon, H.-Y. 2012. Pyrosequencing analysis of the bacterial communities in the guts of honey bees *Apis cerana* and *Apis mellifera* in Korea. *J. Microbiol.* **50**, 735–745.

Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R., and Fierer, N. 2011. Examining the global distribution of dominant archaeal populations in soil. *ISME J.* **5**, 908–917.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., and Knight, R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* **108**, 4516–4522.

Delmont, T.O., Prestat, E., Keegan, K.P., Faubladier, M., Robe, P., Clark, I.M., Pelletier, E., Hirsch, P.R., Meyer, F., Gilbert, J.A., and *et al.* 2012. Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J.* **6**, 1677–1687.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. 2011. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194–2200.

Galand, P.E., Casamayor, E.O., Kirchman, D.L., and Lovejoy, C. 2009. Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc. Natl. Acad. Sci. USA* **106**, 22427–22432.

Hamp, T.J., Jones, W.J., and Fodor, A.A. 2009. Effects of experimental choices and analysis noise on surveys of the "rare biosphere". *Appl. Environ. Microbiol.* **75**, 3263–3270.

Huse, S.M., Huber, J.A., Morrison, H.G., Sogin, M.L., and Welch, D.M. 2007. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* **8**, R143.

Huse, S.M., Welch, D.M., Morrison, H.G., and Sogin, M.L. 2010. Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.* **12**, 1889–1898.

Judo, M.S.B., Wedel, A.B., and Wilson, C. 1998. Stimulation and suppression of PCR-mediated recombination. *Nucleic Acids Res.* **26**, 1819–1825.

Kunin, V., Engelbrektson, A., Ochman, H., and Hugenholtz, P. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* **12**, 118–123.

Lee, C.K., Herbold, C.W., Polson, S.W., Wommack, K.E., Williamson, S.J., McDonald, I.R., and Cary, S.C. 2012a. Groundtruthing next-gen sequencing for microbial ecology–biases and errors in community structure estimates from PCR amplicon pyrosequencing. *PLoS ONE* **7**, e44224.

Lee, S., Weon, H.-Y., Han, K., and Ahn, T.-Y. 2012b. *Flavobacterium dankookense* sp. nov., isolated from a freshwater reservoir, and emended descriptions of *Flavobacterium cheonanense*, *F. chungnamense*, *F. koreense* and *F. aquatile*. *Int. J. Sys. Evol. Microbiol.* **62**, 2378–2382.

Nam, Y.D., Jung, M.J., Roh, S.W., Kim, M.S., and Bae, J.W. 2011. Comparative analysis of Korean human gut microbiota by barcoded pyrosequencing. *PLoS ONE* **6**, e22109.

Pinto, A.J. and Raskin, L. 2012. PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS ONE* **7**, e43093.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., and *et al.* 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65.

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.

Quince, C., Lanzén, A., Curtis, T.P., Davenport, R.J., Hall, N., Head, I.M., Read, L.F., and Sloan, W.T. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nature Methods* **6**, 639–641.

Quince, C., Lanzén, A., Davenport, R., and Turnbaugh, P. 2011. Removing noise from pyrosequenced amplicons. *BMC Bioinformatics* **12**, 38.

Reeder, J. and Knight, R. 2009. The 'rare biosphere': a reality check. *Nature Methods* **6**, 636–637.

Roesch, L.F.W., Fulthorpe, R.R., Riva, A., Casella, G., Hadwin, A.K.M., Kent, A.D., Daroub, S.H., Camargo, F.A.O., Farmerie, W.G., and Triplett, E.W. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* **1**, 283–290.

Schloss, P.D., Gevers, D., and Westcott, S.L. 2011. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE* **6**, e27310.

Schloss, P.D. and Westcott, S.L. 2011. Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl. Environ. Microbiol.* **77**, 3219–3226.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., and *et al.* 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **75**, 7537–7541.

Sun, Y., Cai, Y., Liu, L., Yu, F., Farrell, M.L., McKendree, W., and Farmerie, W. 2009. ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. *Nucleic Acids Res.* **37**, e76.

Tringe, S.G. and Hugenholtz, P. 2011. The enduring legacy of small subunit rRNA in microbiology, pp. 123–128. *In* de Bruijn, F.J. (ed.), Handbook of molecular microbial ecology, John Wiley & Sons, Inc., New Jersey, USA.

Weinstock, G.M. 2011. The impact of next-generation sequencing technologies on metagenomics, pp. 143–147. *In* De Bruijn, F.J. (ed.), Handbook of molecular microbial ecology, John Wiley & Sons, Inc., New Jersey, USA.

Wright, E.S., Yilmaz, L.S., and Noguera, D.R. 2012. DECIPHER, a search-based approach to chimera identification for 16S rRNA sequences. *Appl. Environ. Microbiol.* **78**, 717–725.

Zylstra, P., Rothenfluh, H.S., Weiller, G.F., Blanden, R.V., and Steele, E.J. 1998. PCR amplification of murine immunoglobulin germline V genes: strategies for minimization of recombination artefacts. *Immunol. Cell Biol.* **76**, 395–405.