ARTICLE

Rectification of artificial molecular recombination with the use of high fidelity enzyme in the amplification of 16S rDNA sequences from Stool sample.

Vijay Nema

All Res. J. Biol., **2012**, 3, 6-9

All your results are good results!

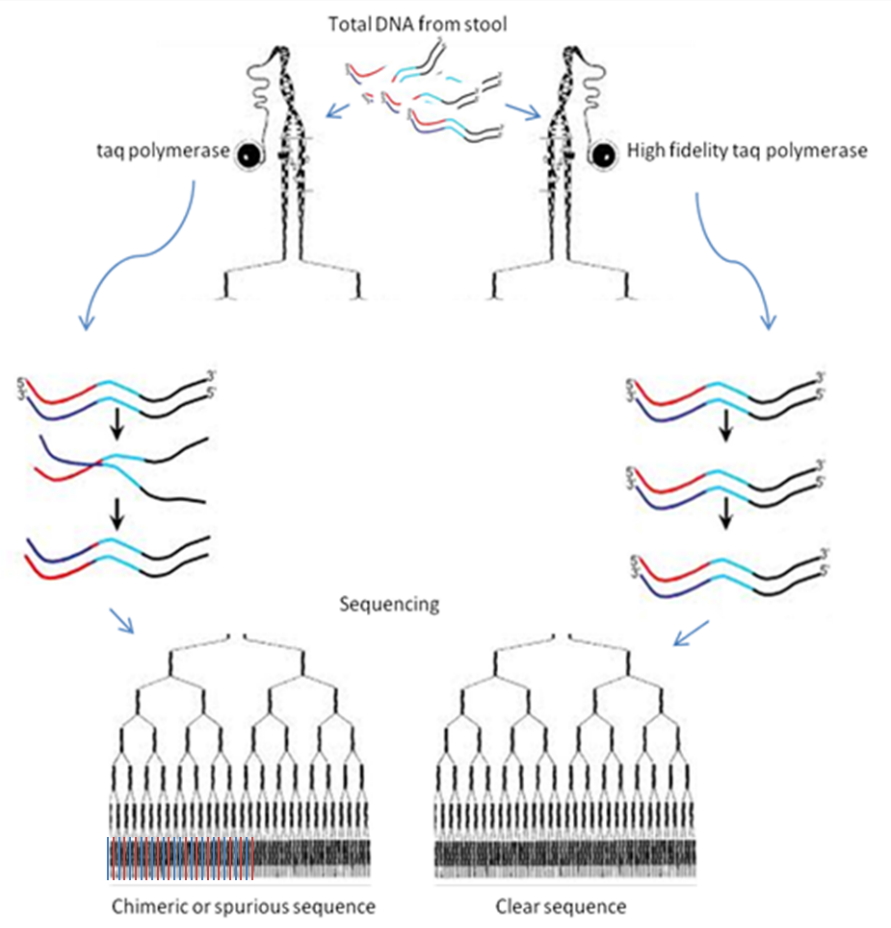
Issue 1, Vol. 3, 2012, 6-9

Rectification of artificial molecular recombination with the use of high fidelity enzyme in the amplification of 16S rDNA sequences from Stool sample.

Vijay Nema [1]

1) National AIDS Research Institute, 73 G MIDC Bhosari, Pune, Maharashtra India 411026. E-mail: vnema@nariindia.org

Graphical Abstract :



Abstract: Reliance on routinely used taq polymerases for amplification may generate spurious sequences, especially in metagenomic studies consuming complex mixtures of various DNA templates. This study reports one such incident wherein initial amplification of 16SrRNA gene from purified DNA isolated from a stool sample, generated spurious sequences when amplified utilizing the taq polymerase used in routine amplifications. This was rectified when the same gene was reamplified using high-fidelity taq polymerase. Use of high fidelity enzymes and verification of the sequences using various software tools before submission to the databases ensures better quality and confidence.

Keywords: Metagenomics; 16S rDNA; Sequencing; Chimera; High fidelity taq polymerase; Pintail.

Introduction

With the legendary work of Woese and his co-workers,[1] a new kingdom Archaea was defined using 16S ribosomal DNA sequences and since then thousands of sequences have been submitted to ribosomal databases defining new strains and species. Amplification of the 16S rDNA sequences from various environmental and biological niches has become a general practice now.[2-4] These PCR-based methods, which are otherwise very robust, have some significant drawbacks. Chimera formation is one such problem that becomes more likely with the homology that exists between the different DNA templates present in the samples from complex habitats. All 16S rDNA sequences share a degree of homology due to the highly conserved nature of the 16S rRNA gene. A chimera is a spurious gene sequence derived from two microorganisms that forms during PCR amplification.[5,6] The formation of chimeras can be attributed to several phenomena like DNA damage,[7] high numbers of PCR amplification cycles[8] and errors by non proof-reading enzymes during amplification.[9] DNA in stool samples may suffer these snags during amplification as it is always exposed to various kinds of deleterious conditions. This work reports one such incident where the use of high fidelity enzymes could rectify the problem of chimera formation through the use of Pintail software to determine chimera like sequences.

Materials and Methods

Two frozen stool samples were obtained from stocks of microbiology division with the available background information about their origin for an exploratory study. Sample 1, was obtained from a HIV positive individual with a diarrheal episode. Sample 2 was obtained from a non HIV infected person with diarrhea. DNA was isolated from these samples using QIAmp DNA Stool Mini Kit (QIAgen). Isolated DNA was amplified using universal primers viz. 27F and 1492R (Table 1). The reaction mixture contained Taq polymerase (Bangalore Genei 105914) with 10X buffer containing 100 mM Tris (pH 9.0), 500 mM KCl and 0.1% Gelatin with final concentration of 200µM of dNTPs and 20pM of each primer in a solution of 50µl adjusted with DNase free water. The reaction was repeated for 30 cycles with annealing temperature of 54°C for 1.5 minutes. The amplicons were purified by cutting the band from 1.5% low melting agarose gel after electrophoresis and eluting using QIAquick Gel Extraction Kit. Purified amplicons were cloned into StrataClone™ PCR Cloning Vector pSC-A that uses ligation of PCR product through A-U base-pairing followed by topoisomerase I-mediated strand ligation. Clones obtained after transformation were picked and analyzed for insertion by gel electrophoresis. Positive clones were subjected to plasmid isolation using QIAGEN Plasmid Mini Kit. Purified plasmids were used for sequencing the inserted gene using M13 forward and M13 reverse primers and pD primer as an intermediate primer to cover the complete sequence (Table 1).

Table 1. Primers and their sequences used for amplification and sequencing.

Primer Name	Nucleotide Sequences 5´ to 3´	Reference
27 Forward	GAG TTT GAT CMT GGC TCA G	11
1492 Reverse	GGY TAC CTT GTT ACG ACT T	11
M 13 Forward	GTA AAA CGA CGG CCA G	Universal sequencing primer
M 13 Reverse	CAG GAA ACA GCT ATG AC	Universal sequencing primer
pD	CAG CAG CCG CGG TAA TAC	12

The sequencing of plasmid DNA was carried out using Big Dye Terminator v. 3.1 cycle sequencing kit and an ABI 3730xl DNA analyzer, according to the manufacturer's protocol. Nucleotide sequence analysis and curing were done using Applied Biosystem SeqScape v.2.5. The cured sequences were blasted to analyze sequence similarities with the sequences present in nucleotide database using BLASTN 2.2.25 from NCBI.

Results and Discussion

Different sequences showed similarities with different genera in the cloning products from both the samples. Ribosomal RNAs are generally highly conserved and thus are expected to obtain similarities to other rRNA sequences over the entire length. Keeping this in consideration, the sequences were checked with Pintail software[10] for chimera formation or any other sequence anomalies. All the clones from sample 2 were found to be non-anomalous. In sample 1 all (except one) sequences were found to be non-anomalous. A sequence (1442bp long) from one of the clones from sample 1 showed the highest sequence similarity of 94% with 28 gaps to Lactobacillus gasseri JV-V03 Sca04-Contig11, whole genome shotgun sequence (accession no. ACGO02000005.1). The analysis was done using Pintail software with the sequence in question as query (1442 nt) and subject was the most identical sequence as obtained with BLAST. The in-built parameters of the software determined the sequence variability with a 300 base window, moving 25 bases at a time along the sequence length. The software depicted the probability of two non-anomalous sequences producing a DE (deviation from expectation) of 3.51, when they differ by 6.01 % overall, to be $0.25 > P > 0.05$. Hence despite an acceptable DE value, the lower confidence indicated a caution (Figure 1). To locate the segment of variation the sequence was run again in Pintail by removing 100bp at a time from beginning to end of the query sequence and vice versa. It was observed that region around 750 to 950 bp according to the base position of 16S rRNA gene gave the highest anomaly and could have been the region of chimera formation. The clone was again sequenced for 16SrRNA gene using the purified plasmid and exactly the same sequences were obtained. This indicated the problem with the insert used for cloning and not with the clone itself. This

insert was the amplicon obtained after the amplification of 16SrRNA gene from purified DNA of sample 1.

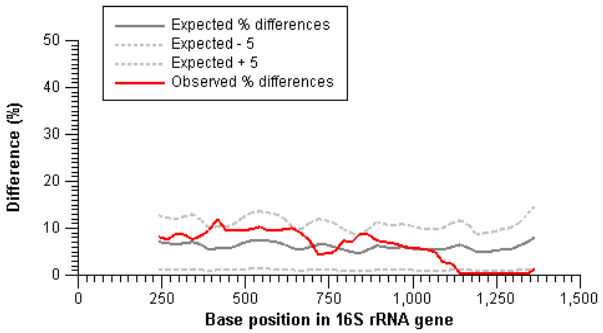


Figure 1. Variation in % difference between clone sequence (query) and ACGO02000005.1 sequence (subject).

To confirm if the sequence has an artificial event of recombination or is a naturally occurring chimera-like sequence, the experiment was performed from the beginning by repeating the PCR amplification using the original DNA isolated from the sample 1. This time, a high fidelity Taq polymerase (Platinum® Taq DNA Polymerase High Fidelity from Invitrogen) was used for the amplification. A compatible 10X buffer was used as provided by the enzyme manufacturer. Other reagents like dNTPs and primers were all kept constant like previous experiments. Also, other steps of gel elution, cloning, sequencing and BLAST analysis remained the same as before. Apart from other sequences showing similarity with other organisms, a few clone sequences were found to be 99-100% identical with *Lactobacillus gasseri* gene for 16S rRNA, complete sequence (accession no. AB517146.1). All these sequences when compared with the gene sequence mentioned above using Pintail, gave no anomaly and were also observed to be practically identical. No other clone gave any such indication for chimera formation using Pintail, when compared with the closest match as per the BLAST analysis. This indicates that the use of high fidelity enzymes for amplification of 16S rDNA sequences from complex DNA samples can help in obtaining reliable sequences. A few reagents and reaction conditions used in both the experiments were different but this did not produce any variation in other clones when the results from both experiments were compared. Also within one experiment, when different clones were analysed for anomalies, only the experiment with Taq polymerase without proofreading activity produced a clone with spurious sequences. This observation substantiated the outcomes of an earlier study that discussed the use of processivity-enhanced polymerases for the reduction in PCR mediated recombination.[9] Also in this study, amplification with high fidelity enzyme has generated 40 different positive clones identified after sequencing as compared to 30 different positive clones been generated with the use of routinely used Taq polymerase when all other procedures were kept constant. Later analysis of the results showed that sample 1 contained a diverse set of genus whereas sample 2 was limited in genus diversity (data not shown). Hence the complexity of sample 1 but not sample 2 might have made the phenomenon of chimera formation possible. As metagenomic approaches are taking the front seat in the analysis of environmental samples, practices involving amplification of various genes, sequencing and sequence submission to various databases are becoming a routine. With the advanced knowledge and tools, we now have sufficient evidence that whatever is being amplified may not show the real picture. Hence care must be taken at every step to ensure the quality of data being generated and made available as standards for future analysis in the databases.

Acknowledgements: Author is thankful to Director NARI for infrastructural and moral support. Also the help of Dr. Jayanta Bhattacharya, Head, Molecular Virology Division at NARI and his team during experimentation and constructive suggestions of the GenBank submissions staff for sequence analysis is highly acknowledged.

Competing interest: The author declare no competing interests

References:

1. Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A 74, 5088-5090.
2. Giovannoni, S. J., Britschgi, T. B., Moyer, C. L., and Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. Nature 345, 60-63.
3. Ward, D. M., Weller, R., and Bateson, M. M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature 345, 63-65.
4. Weller, R., and 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.
5. Meyerhans, A., Vartanian, J. P., and Wain-Hobson, S. (1990). DNA recombination during PCR. Nucleic Acids Res 18, 1687-1691.
6. Shuldiner, A. R., Nirula, A., and Roth, J. (1989). Hybrid DNA artifact from PCR of closely related target sequences. Nucleic Acids Res 17, 4409.
7. Paabo, S., Irwin, D. M., and Wilson, A. C. (1990). DNA damage promotes jumping between templates during enzymatic amplification. J Biol Chem 265, 4718-4721.
8. 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 (Pt 5), 1107-1114.
9. Lahr, D. J., and Katz, L. A. (2009). Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. Biotechniques 47, 857-866.

10. Ashelford, K. E., Chuzhanova, N. A., Fry, J. C., Jones, A. J., and Weightman, A. J. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl Environ Microbiol 71, 7724-7736.

11. Frank, J. A., Reich, C. I., Sharma, S., Weisbaum, J. S., Wilson, B. A., and Olsen, G. J. (2008). Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. Appl Environ Microbiol 74, 2461-2470.

12. Weisburg, W. G., Barns, S. M., Pelletier, D. A., and Lane, D. J. (1991). 16S ribosomal DNA amplification for phylogenetic study. J Bacteriol 173, 697-703.