A NOVEL METHOD FOR ASSESSMENT OF 16S RRNA GENE COPY NUMBER IN BACTERIAL GENOMES BY PULSED-FIELD GEL ELECTROPHORESIS AND PCR AMPLIFICATION

YONGHUI ZENG and NIANZHI JIAO[1]

State Key Laboratory of Marine Environmental Science
Xiamen University
Xiamen 361005, China

Accepted for Publication June 23, 2008

ABSTRACT

*16S rRNA gene (*rrn*) copy number in bacterial genomes is indicative of ecological strategies of bacteria and is critical for quantification of bacterial abundance in mixed populations using* rrn*-based approaches. For accurate assessment of* rrn *copies, a novel technical strategy by means of pulsed-field gel electrophoresis and polymerase chain reaction amplification analysis was introduced. Experimental and* in silico *analysis on a test bacterial culture* Caulobacter crescentus *proved it to be simple, effective, accurate and a good alternative to traditional time-consuming methods.*

PRATICAL APPLICATIONS

This method can be used for routine determination of gene copy number in most bacteria whose full genome sequences are not available. Moreover, the pulsed-field gel electrophoresis bands containing a target gene fragment can be determined and therefore constructing an expected fragments oriented genomic library is possible.

INTRODUCTION

Most bacterial genomes contain multiple rRNA operons. The number of 16S rRNA genes (*rrn*) significantly correlates with the rate at which bacteria respond to resource availability and thus could serve as an indicator of

[1]Corresponding author. TEL/FAX: +0086-592-2187869; EMAIL: jiao@xmu.edu.cn

ecological strategies of bacteria (Klappenbach *et al.* 2000). In microbial ecology studies, real-time polymerase chain reaction (PCR) technique is frequently used to determinate relative or absolute abundance of bacterial groups or even a specific species in a community, where accurate *rrn* copy number need to be known prior to quantification.

Currently, several empirical values for bacterial *rrn* copy number have been suggested, e.g., 3.8 ± 2.9 for general applications (Fogel *et al.* 1999) or 1.9 for marine oligotrophic waters (Campbell *et al.* 2008), both of which were calculated based on bacterial or environmental genome data. However, such applications may introduce several fold errors. Theoretically, *rrn* copy numbers of all bacterial species need to be determined for reducing the bias. Unfortunately, only a few techniques are available for this purpose. The most commonly used one is the Southern blotting technique (e.g., see Maslunka *et al.* 2006), the whole procedure of which is extremely fussy, time-consuming and rich-experience needed. Real-time PCR can also be used for this purpose, but large uncertainty remains (Smith *et al.* 2006). Cell lysis efficiency and DNA quality may greatly affect the accuracy of this technique. Temperature gradient gel electrophoresis (TGGE) has been used to assess the 16S rRNA gene heterogeneity (Nübel *et al.* 1996) and has the potential for counting 16S rRNA gene copy numbers. However, TGGE (or DGGE) could not discriminate identical 16S rRNA gene fragments and therefore their bands are only the lowest estimate of 16S rRNA gene copy number in bacterial genomes.

Here we proposed a novel strategy for determination of *rrn* copy number in bacterial genomes by combining pulsed-field gel electrophoresis (PFGE) and PCR amplification techniques (Fig. 1). Compared with traditional methods such as that based on Southern blotting, this novel method could possibly save half the time required (see Fig. 1 for comparison). Meanwhile, it could be easily carried out after PFGE, the initial aim of which may be to determine genome size or fingerprint. In such case, no additional efforts such as preparing an over 2 days blotting experiment are needed.

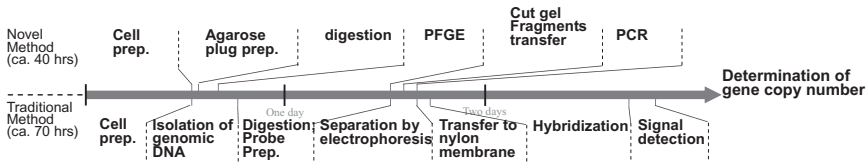

FIG. 1. SCHEMATIC DIAGRAM OF THE NOVEL PROCEDURE (UPPER PART) FOR ASSESSMENT OF *RRN* COPY NUMBER IN BACTERIAL GENOMES AND ITS COMPARISON WITH A TRADITIONAL METHOD BASED ON SOUTHERN BLOTTING TECHNIQUE (LOWER PART)

MATERIALS AND METHODS

An aquatic bacterium *Caulobacter crescentus* ATCC 19089 was used to test this strategy, the whole genome of which has been sequenced in 2001 (Nierman *et al.* 2001). For PFGE analysis, single colony was inoculated to medium (ATCC 36) and cultured for 14–18 h at 30C. A total of 400 μL resuspension ($OD_{610} \approx 1.35$) in cell suspension buffer (CBS: 100 mM Tris–HCl : 100 mM EDTA, pH 8.0) was treated with 20 μL proteinase K (conc. 20 mg/mL) and then mixed with 400 μL melting 1% SDS : 1% Seakem Gold agarose (Cambrex Co., Rockland, ME) for plug preparation. Plugs were then digested at 54C for ca. 2 h within cell lysis buffer (CLB, pH 8.0: 50 mM Tris, 50 mM EDTA, 1% Sarkosyl) with final proteinase K concentration of 0.1 mg/mL. After repeated washing with MilliQ water and TE buffer, plugs containing naked and consecutive genomes were prepared. After digest with AseI (New England Biolabs, Beverly, MA) for ca. 2 h, plugs were loaded for fingerprinting analysis on a PFGE system (Rotahpor 6.0, Biometra Co., Germany). Running parameters were voltage 6V, angle 120°, pulse time 2.2–63.8 s, and electrophoresis time ca. 18 h. Finally, gel was stained with EB and subjected to gel image analysis.

For 16S rRNA gene amplification analysis, all PFGE bands were cut out and electro-eluted in 100 μL TE. A total of 5 μL DNA-dissolved TE served as template for PCR analysis using universal bacterial primers 27F (5′-AGA GTT TGA TCT GGC TCA G-3′) and 1492R (5′-GGT TAC CTT GTT ACG ACT T-3′). A routine amplification protocol was used as described previously by Sekiguchi *et al.* (2002).

RESULTS AND DISCUSSIONS

After PFGE analysis of the genomic DNA of *Caulobacter crescentus* ATCC 19089, 11 distinct bands were obtained with the fragment size ranging from 20 kb to1 Mb (Fig. 2A). Agarose electrophoresis analysis showed that two samples (No. 2 and 3) had amplification product of 16S rRNA gene with expected band size (Fig. 2B), suggesting that this genome contains two *rrn* copies.

To verify this result, we searched the genome of *C. crescentus* ATCC 19089 on NCBI website (http://www.ncbi.nih.gov). The result showed that *C. crescentus* has exactly two *rrn* copies, located in the nucleotide position 2840149-2841639 and 3770180-3771670. To further check the potential influence or bias by enzyme choice, whole *C. crescentus* genome sequence was downloaded from NCBI database (GenBank accession number: AE005673.1) and then subjected to *in silico* digestion by the molecular program suite Vector

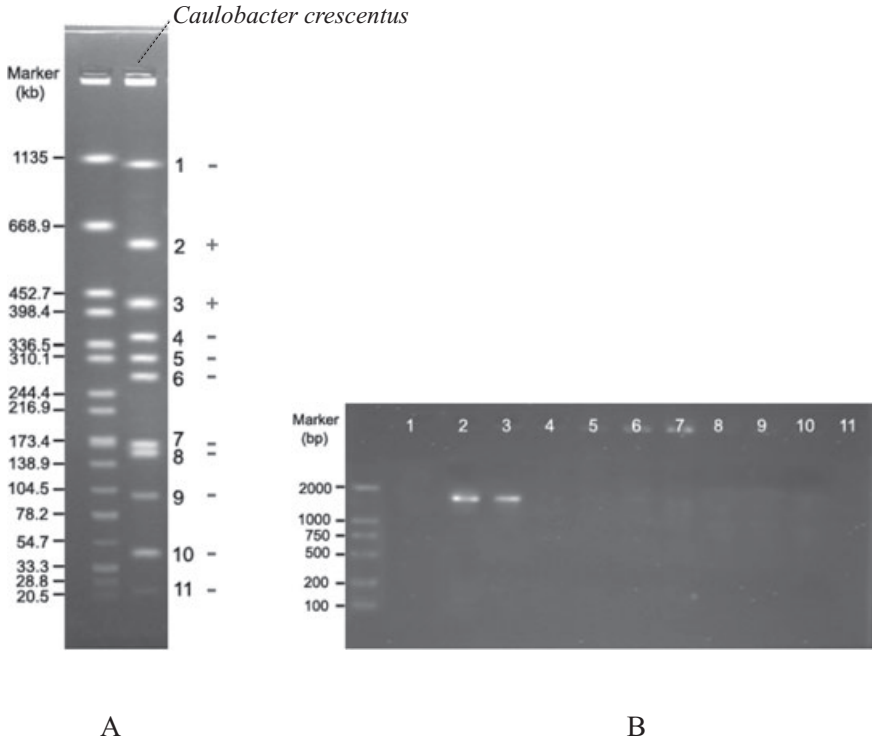

FIG. 2. PULSED-FIELD GEL ELECTROPHORESIS (PFGE) ANALYSIS OF ASEI DIGESTS OF *CAULOBACTER CRESCENTUS* CULTURES (A), AND 16S RRNA GENE PCR AMPLIFICATION RESULT (B)
Positive amplification is marked with "+", while negative results are marked with "–".

NTI 9.0 (Informax, Inc., Bethesda, MD). Simulation results showed that five enzymes were able to cut the genome into 10–30 fragments, the range of which could be best resolved on PFGE gel in our experience. They are AhlI (29), AseI (15), BssNAI (30), BstSNI (23) and PsiI (28) with total number of digested fragments listed in parentheses. All five enzymes have cutting sites flanking two 16S rRNA genes and, for each enzyme, two fragments containing full-length 16S rRNA gene are distributed into different-length digests: AhlI (53 kb/333 kb), AseI (408 kb/657 kb), BssNAI (79 kb/325 kb), BstSNI (50 kb/275 kb) and PsiI (234 kb/229 kb). The two digests from PsiI have only ca. 5 kb difference in length and therefore PFGE parameters need to be optimized to reach best resolution. The other four enzymes can be successfully used in routine PFGE analysis for accurate determination of *rrn* copy number, suggesting that criterion for enzyme choice is loose for this method.

In summary, a novel strategy was proposed for assessing *rrn* copy number in bacterial genome. Experimental and *in silico* analysis on a test bacterial culture proved it to be simple, effective and accurate. We suggest that this method be used for routine determination of gene copy number in most bacteria whose full genome sequences are not available. When applying it to an unidentified bacterial culture, the enzyme selection can refer to the genome sequences of its relatives in public databases, such as the NCBI genome database (http://www.ncbi.nlm.nih.gov/Genomes), where over 600 microbial genomes are available and the number is increasing. In general, for different bacterial species, different enzymes need to be used since bacterial genome composition varied highly among bacteria. Currently, we are developing an enzyme database for this method by statistical analysis of all available bacterial genome sequences. The common enzymes applicable to all bacterial genomes are not yet found. Here we propose several fundamental criterions, following which one can easily choose out suitable enzymes for a specific species: (1) choose a reference genome whose phylogenetic affiliation is closely related to your target bacterial species. Then, search the enzymes that only cut in the non-*rrn* regions to determine the most appropriate restriction endonucleases; and (2) try more enzymes to verify results. Double enzyme digest can definitely increase the accuracy. On the sequenced bacterial genomes we pre-screened, there are enough enzymes to reach our purpose because 16S rRNA gene copy number is low in most bacterial genomes (mostly 3–8 copies per genome, for review, see Tourova 2003) and therefore inter-ribosomal operon sequences are long enough to search out common enzymes for cutting them. In addition to the 16S rRNA gene, this novel technical strategy may also be used to determine the copy number of functional genes. Moreover, PFGE bands containing a target gene fragment can easily be identified and therefore constructing an expected fragments oriented genomic library is possible.

ACKNOWLEDGMENTS

We are grateful to Dr. Liyun Liang (Huayue Co., China) for her help on PFGE and PCR analysis. This work was supported by the NSFC project (40632013), the MOST projects (2007CB815900 and 2006BAC11B04), the AOMRRD's 11th 5-year program (DYXM-115-02-4-3), the MOE project (704029) and the SOA project (200805068).

REFERENCES

CAMPBELL, B.J., WAIDNER, L.A., COTTRELL, M.T. and KIRCHMAN, D.L. 2008. Abundant proteorhodopsin genes in the North Atlantic ocean. Environ. Microbiol. *10*, 99–109.

FOGEL, G.B., COLLINS, C.R., LI, J. and BRUNK, C.F. 1999. Prokaryotic genome size and SSU rDNA copy number: Estimation of microbial relative abundance from a mixed population. Microb. Ecol. *38*, 93–113.

KLAPPENBACH, J.A., DUNBAR, J.M. and SCHMIDT, T.M. 2000. rRNA operon copy number reflects ecological strategies of bacteria. Appl. Environ. Microbiol. *66*, 1328–1333.

MASLUNKA, C., CARRA, E., GÜRTLER, V., KÄMPFER, P. and SEVIOUR, R. 2006. Estimation of ribosomal RNA operon (rrn) copy number in Acinetobacter isolates and potential of patterns of rrn operon-containing fragments for typing strains of members of this genus. Syst. Appl. Microbiol. *29*, 216–228.

NIERMAN, W.C., FELDBLYUM, T.V., LAUB, M.T., PAULSEN, I.T., NELSON, K.E., EISEN, J.A., HEIDELBERG, J.F., ALLEY, M.R., OHTA, N. and MADDOCK, J.R. 2001. Complete genome sequence of Caulobacter crescentus. Proc. Natl. Acad. Sci. USA *98*, 4136–4141.

NÜBEL, U., ENGELEN, B., FELSKE, A., SNAIDR, J., WIESHUBER, A., AMANN, R.I., LUDWIG, W. and BACKHAUS, H. 1996. Sequence heterogeneities of genes encoding 16S rRNAs in Paenibacillus polymyxa detected by temperature gradient gel electrophoresis. J. Bacteriol. *178*, 5636–5643.

SEKIGUCHI, H., KOSHIKAWA, H., HIROKI, M., MURAKAMI, S., XU, K., WATANABE, M., NAKAHARA, T., ZHU, M. and UCHIYAMA, H. 2002. Bacterial distribution and phylogenetic diversity in the Changjiang estuary before the construction of the Three Gorges Dam. Microbial. Ecol. *43*, 82–91.

SMITH, C.J., NEDWELL, D.B., DONG, L.F. and OSBORN, A.M. 2006. Evaluation of quantitative polymerase chain reaction based approaches for determining gene copy and gene transcript numbers in environmental samples. Environ. Microbiol. *8*, 804–815.

TOUROVA, T.P. 2003. Copy number of ribosomal operons in prokaryotes and its effect on phylogenetic analyses. Microbiology *72*, 389–402.