3778–3783 *Nucleic Acids Research, 1996, Vol. 24, No. 19*

Use of tagged random hexamer amplification (TRHA) to clone and sequence minute quantities of DNA—application to a 180 kb plasmid isolated from *Sphingomonas* F199

Kwong-Kwok Wong*, Lisa C. Stillwell, Christopher A. Dockery and Jeffrey D. Saffer

Molecular Biosciences Department, P7-56, Pacific Northwest National Laboratory, PO Box 999, Richland, WA 99352, USA

Received June 6, 1996; Revised and Accepted August 22, 1996

ABSTRACT

We have developed a novel method to clone and sequence minute quantities of DNA. The method was applied to sequence a 180 kb plasmid pNL1. The first step was the production of a size distributed population of DNA molecules that were derived from the 180 kb plasmid pNL1. The first step was accomplished by a random synthesis reaction using Klenow fragment and random hexamers tagged with a T7 primer at the primer 5′-end (T7-dN_6, 5′-GTAATACGACTCACTATAGGGCNNN-NNN-3′). In the second step, Klenow-synthesized molecules were amplified by PCR using T7 primer (5′-GTAATACGACTCACTATAGGGC-3′). With a hundred nanograms starting plasmid DNA from pNL1, we were able to generate Klenow-synthesized molecules with sizes ranging from 28 bp to >23 kb which were detectable on an agarose gel. The Klenow-synthesized molecules were then used as templates for standard PCR with T7 primer. PCR products of sizes ranging from 0.3 to 1.3 kb were obtained for cloning and sequencing. From the same Klenow-synthesized molecules, we were also able to generate PCR products with sizes up to 23 kb by long range PCR. A total 232.5 kb sequences were obtained from 593 plasmid clones and over twenty putative genes were identified. Sequences from these 593 clones were assembled into 62 contigs and 99 individual sequence fragments with a total unique sequence of 86.3 kb.

INTRODUCTION

Sphingomonas F199 was isolated from sediments at a depth of 407 m (1). This bacterium has the ability to use toluene, all isomers of xylene, *p*-cresol, naphthalene, salicylate and benzoate as sole carbon and energy sources (1). It harbors two megaplasmids of 180 kb (pNL1) and 475 kb (pNL2) (2). An initial catabolic screening study with a cosmid library generated from pNL1 has located catechol 2,3-dioxygenase activity on pNL1 (2). To explore the potential of using pNL1 for subsurface bioremediation, we initiated a project to sequence pNL1 completely.

Partial restriction digestion or mechanical shearing of DNA templates to obtain DNA fragments with sizes ~1 kb are the currently-preferred methods for the construction of plasmid libraries for sequencing. However, such methods often require a large quantity (50–100 μg) of starting DNA material and are labor intensive. Two methods which are the whole genome PCR (3) and primer-extension preamplification (PEP) (4) have been developed to amplify the whole genome DNA from a small quantity of DNA material. Amplified materials were used for the isolation of specific DNA sequences and for genetic analysis (3,4). The whole genome PCR method still requires restriction digestion or sonication to generate small fragments, which are then ligated to a linker for PCR amplification. The PEP approach uses a mixture of 15-base oligonucleotides as primers for PCR amplification. However, the efficiency of the amplification is very low and PEP approach does not amplify enough material to make a plasmid library that is adequate for sequencing.

To increase the amplification efficiency of PEP, a tagged random primers PCR (T-PCR) method (5) was developed to amplify efficiently from small quantities of DNA samples with sizes ranging from 400 bp to 1.6 kb. This method involves two PCR reactions with tagged random primers containing nine to 15 random bases at the 3′-end and a constant 17 bp at the 5′-end. In the first PCR step, the tagged random primer is used to generate products with tagged primer sequences at both ends, which are achieved by using a low annealing temperature in the PCR cycles. Excess tagged primers are then removed. In the second PCR step, the primer with the constant 17 bp sequence is used to amplify PCR products from the first PCR step. Since tagged primers with 12 or more random bases will generate non-specific products, presumably resulting from primer–primer extensions or less efficient elimination of these longer primers during the filtration step (5), it would be advantageous to use a tagged random primer with shorter random bases. In this report, random synthesis was achieved by using Klenow fragment and random hexamer tagged with T7 primer at the primer 5′-end. Klenow-synthesized molecules were then amplified with T7 primer. As far as we

* To whom correspondence should be addressed

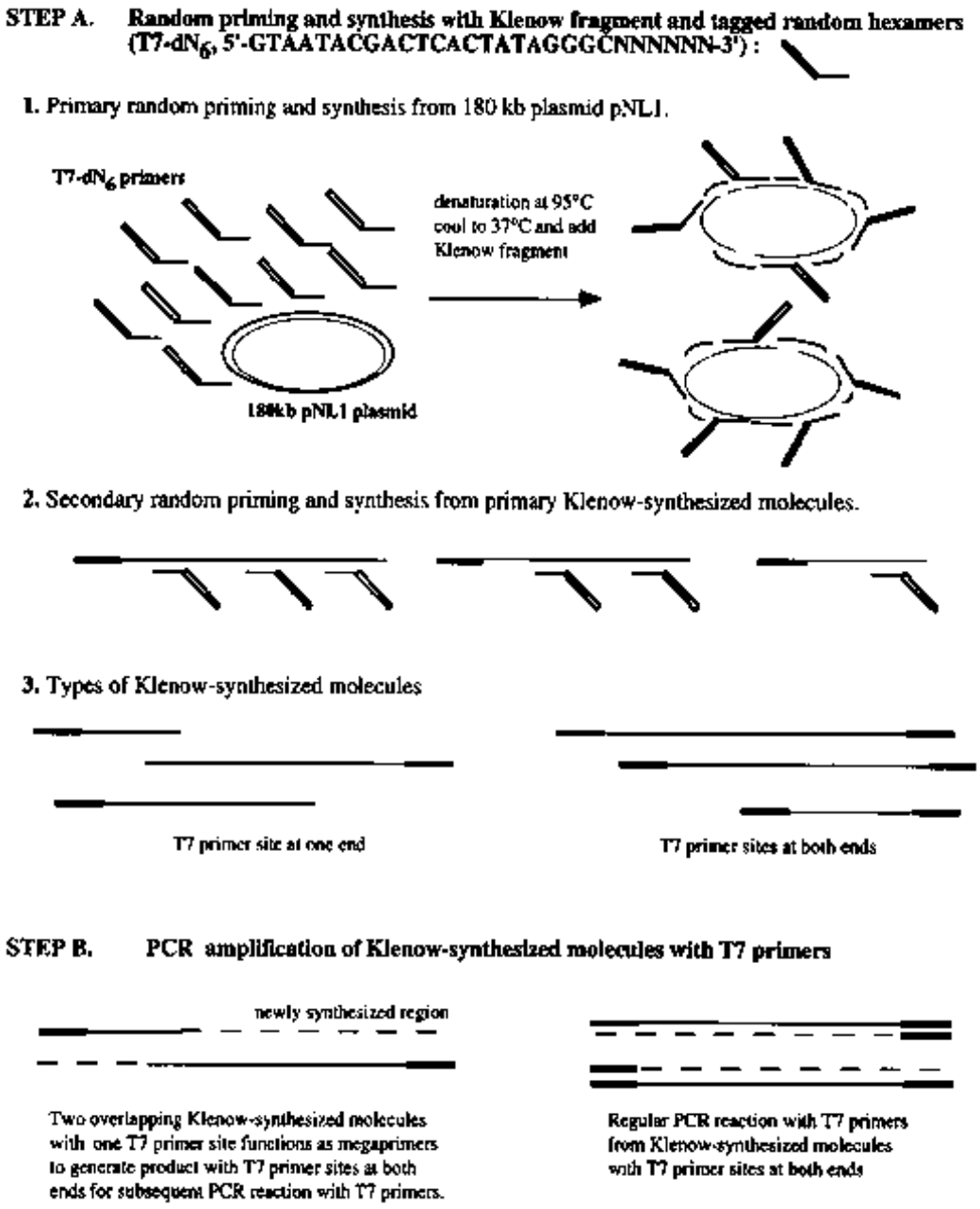

Figure 1. Tagged random hexamer amplification (TRHA) strategy.

know, this is the first report of the cloning and sequencing of randomly amplified PCR products to assess its coverage of the original template.

MATERIALS AND METHODS

Plasmid and primers

pNL1 is a 180 kb plasmid isolated from a subsurface bacterium *Sphingomonas* F199, which can utilize a variety of aromatic compounds as sole carbon sources (6). Preparation of the plasmid pNL1 was described previously (2). T7 primer (5′-GTAATACGACTCACTATAGGGC-3′) and tagged random hexamer, T7 primer-dN6 (5′-GTAATACGACTCACTATAGGGCNNNNNN-3′) were synthesized using an Applied Biosystems RNA/DNA synthesizer 392 (Perkin Elmer, Foster City, CA).

Tagged random hexamer amplification (TRHA) (Fig. 1)

A. Random synthesis with Klenow fragment and tagged random hexamer. Two 100 µl reactions were prepared, using 0.5× universal buffer (50 mM KOAc, 12.5 mM Tris-acetate, pH 7.6, 5 mM MgOAc, 0.25 mM β-mercaptoethanol and 5 µg/ml BSA), 0.2 mM dNTP, 0.9, 1.8 or 2.7 µM T7-dN_6 primer and 0.1 µg pNL1 plasmid in each case. The reactions were heated at 100°C to denature the DNA for 5 min and cooled to room temperature. Klenow fragment (10 U) was added to each mixture before incubating them at 37°C for 2 h. Thereafter, to one reaction, the

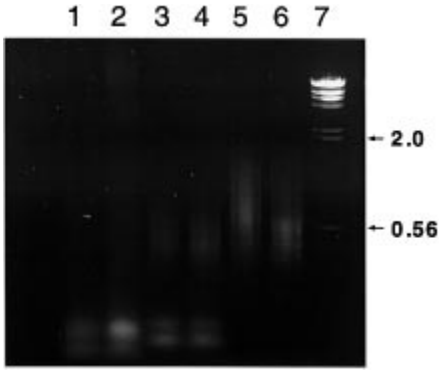

Figure 2. Tagged random hexamer amplification of the 180 kb pNL1 plasmid using T7-dN_6 primer. Lane 1, one cycle of random synthesis with Klenow fragment and T7-dN_6 primer. Lane 2, two cycles of random synthesis with Klenow fragment and T7-dN_6 primer. Lanes 3 and 4, PCR amplifications with T7 primer using Klenow-synthesized molecules as shown in lanes 1 and 2 respectively. Lanes 5 and 6, similar PCR amplifications as lanes 3 and 4 but the excess T7-dN_6 primers have been removed from the Klenow-synthesized molecules. Lane 7, *Hin*dIII-digested lambda DNA was used as size markers. DNA sizes in kb are shown to the right.

reaction mix was again heated at 100°C for 5 min and another 10 U Klenow was added for a second round of Klenow random synthesis. The second reaction had no additions. Excess T7-dN_6 primers in the Klenow-synthesized molecules were subsequently removed through Centricon-100 by washing with 2 ml sterile water twice.

B. PCR amplification of Klenow-synthesized molecules. In a 100 µl reaction mix, 1.5 mM $MgCl_2$, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 0.2 mM dNTP, 0.5 µM T7 primers, 1 µl Klenow-synthesized molecules and 2.5 U ampliTaq (LD) (Perkin Elmer, Foster City, CA) were added. The reactions were heated to 72°C for 10 min and followed by 30 cycles of 95°C for 1 min, 60°C for 1 min and 72°C for 2 min. Here, we referred the PCR amplified products as TRHA products.

Library construction and screening

TRHA products were blunt-end ligated into the pCRscript cloning vector (Stratagene, La Jolla, CA) using standard cloning procedures (7). Ligated products were used to transform XL1-Blue supercompetent cells (Stratagene, La Jolla, CA) and transformed cells were then plated onto LB agar plates containing 200 µg/ml ampicillin, X-gal and IPTG (X-gal plates). White colonies which appeared on X-gal plates were picked individually with pipette tips into a 96-well microtiter plates containing 2YT medium (16 g/l tryptone, 10 g/l yeast extract and 5 g/l NaCl). The cells were allowed to grow in the microtiter plates at 37°C without shaking for 4 h. A 2 µl cell culture was then directly transferred to a 50 µl PCR reaction mix which contained 1.5 mM $MgCl_2$, 10 mM Tris–HCl (pH 8.3), 50 mM KCl, 0.2 mM dNTP, 0.2 µM RSP (5′-GGAAACAGCTATGACCATGA-3′) and SP (5′-GTAAAACGACGGCCAGT-3′) primers, and 1.25 U ampliTaq (LD) (Perkin Elmer, Foster City, CA). The PCR reactions were first heated to 72°C for 10 min, then 30 cycles of the following temperature profile: 95°C for 1 min, 60°C for 1 min and 72°C for 2 min. PCR products were analyzed on a 1.2% agarose gel.

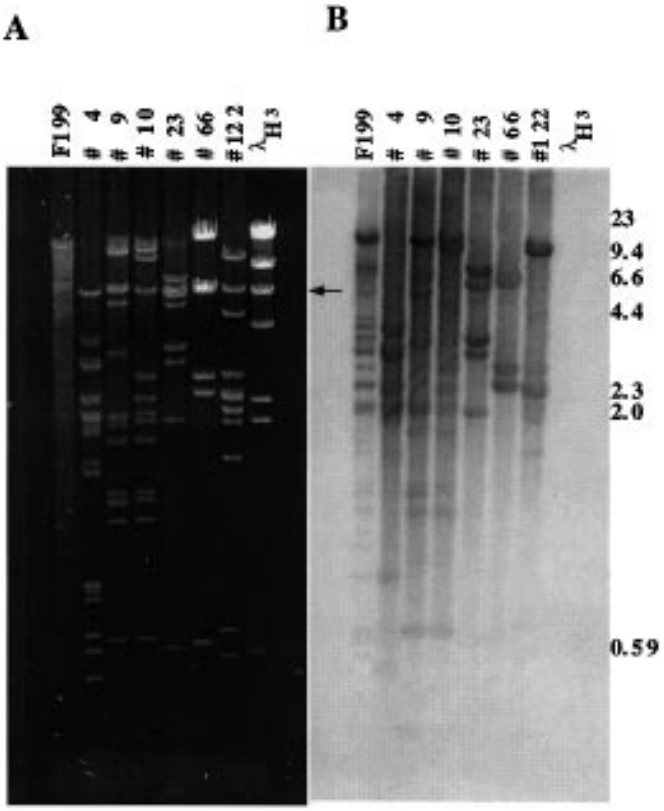

Figure 3. Southern blot of miniset cosmid clones of pNL1 plasmid and hybridization with fluoresceinated DNA probe generated from PCR products as shown in Figure 2, lane 5. (**A**) Agarose gel electrophoresis of total genomic DNA from *Sphingomonas* F199 and DNAs from six cosmid clones of pNL1 plasmid (#4, #9, #10, #23, #66 and #122), which have been completely digested with restriction enzyme *Eco*RI. (**B**) Southern blot of the gel shown in Figure 3A following hybridization with fluoresceinated DNA probe generated from PCR products as shown in Figure 2, lane 5. Hybridization signal was detected with FluoImager SI (Molecular Dynamics, Sunnyvale, CA). λH_3 was molecular weight markers generated from *Hin*dIII digested lambda DNA and sizes are shown to the right. The arrow indicated the location of sCos1 vector fragment which has a size of 7.6 kb.

Southern blot and hybridization

Cosmid clones of pNL1 were digested with *Eco*RI and the resulting fragments were separated by electrophoresis on a 0.8% agarose gel. The gel was then blotted onto Duralon membrane and UV cross-linked. Fluoresceinated DNA probes were prepared with Prime-It Fluor Fluorescence Labelling Kit (Stratagene, La Jolla, CA) according to manufacturer's instruction. Hybridization was done at 60°C in 15 ml QuickHyb solution (Stratagene, La Jolla, CA) for 3 h. The hybridization signal was detected with Illuminator Non-radioactive Detection System (Stratagene, La Jolla, CA). Fluorescent substrate ATTOPHOS (JBL Scientific INC., San Luis Obispo, CA) was used in place of CSPD chemiluminescent substrate and hybridization signals were detected with FluorImager SI (Molecular Dynamics, Sunnyvale, CA).

Long range PCR reaction

Klenow-synthesized molecules were used as template for long range PCR using TaqPlus DNA polymerase (Stratagene, La Jolla,

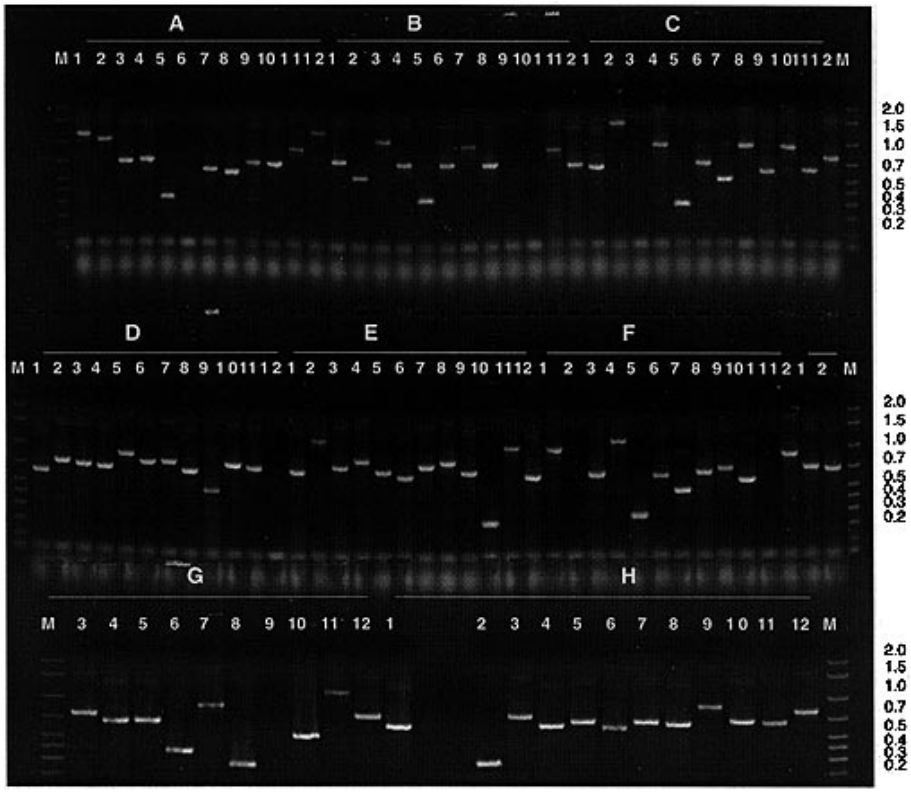

Figure 4. PCR analysis of insert sizes of 96 putative clones. PCR products as shown in Figure 2, lane 5 were blunt-end cloned into*Srf*I restriction site of a plasmid vector pCR-script Amp SK(+) (Stratagene, La Jolla, CA). The insert sizes of 96 putative clones were determined by PCR amplification using vector primers (RSP, 5′-GGAAACAGCTATGACCATG-3′ and SP, 5′-GTAAAACGACGGCCAGT-3′). The 96 clones were named as A1 to H12. M was DNA size marker using AmpliSize Standard (BioRad, Hercules, CA). DNA sizes in kb are shown to the right.

CA). Two different reactions were set up, one in a 100 µl low salt buffer (20 mM Tris–HCl, pH 8.75, 10 mM KCl, 10 mM $(NH_4)SO_4$, 2 mM $MgSO_4$), the other in high salt buffer (20 mM Tris–HCl, pH 9.2, 60 mM KCl, 2 mM $MgCl_2$), which in addition, contained 0.2 mM dNTP and 0.5 µM T7 primer. The PCR reactions were heated to 72°C for 10 min and 5 U TaqPlus DNA polymerase was then added. The reactions were then subjected to 30 cycles of PCR reaction with the following temperature profile: 94°C for 30 s, 60°C for 30 s and 72°C for 15 min.

RESULTS AND DISCUSSION

Tagged random hexamer amplification (TRHA)

Reaction products after the Klenow (Klenow-synthesized molecules) and PCR (TRHA products) steps were analyzed by agarose gel electrophoresis (Fig. 2). A 10 µl volume of Klenow-synthesized molecules from pNL1 were separated on a 1% agarose gel (Fig. 2, lanes 1 and 2). The DNA quantities were amplified during the Klenow random synthesis step. Klenow-synthesized molecules appeared as a smear with sizes ranging from that of the primer to >23 kb. The amount of products were significantly increased by two rounds of random synthesis with Klenow fragment. Klenow-synthesized molecules with or without excess T7-dN_6 removed were then used as templates for subsequent PCR amplification. The relative size and yield of TRHA products increased when the excess T7-dN_6 primer was removed (Fig. 2, lanes 5 and 6). This is probably due to primer dimer formation during PCR amplification. Similarly, we also found that the yield of Klenow-synthesized molecules decreased when the concentration of T7-dN_6 primer increased from 0.9 to 2.7 µM (data not shown).

We were able to obtain PCR products in the range from 0.3 to 1.3 kb from Klenow-synthesized molecules derived from either one or two cycles of Klenow random synthesis (Fig. 2, lanes 3–6). This result implies that two possible events might have taken place during the one cycle of Klenow random synthesis. Firstly, Klenow fragment is able to use the first synthesized DNA strand as template to generate a second strand DNA without requiring a second denaturing step. Secondly, the template, being denatured, provides sites for first-strand annealing in both directions and products with 3′-ends complementary to each other will be present. In the former case, the second strand DNA will have two T7 primer sites in opposite orientation at both ends for PCR reaction. In the second case, Klenow-synthesized molecules with 3′ complementary to each other will be extended in the PCR step. As a result, products with T7 sequences at both ends will be available for subsequent PCR amplification. Furthermore, the average size of PCR products using template from one cycle of Klenow random synthesis (Fig. 2, lane 5) was comparatively larger than that from two cycles of Klenow random synthesis (Fig. 2, lane 6). This size difference indicated that two cycles of Klenow random synthesis may have biased towards smaller fragments because smaller molecules with T7 primer sites at both ends are preferentially-synthesized.

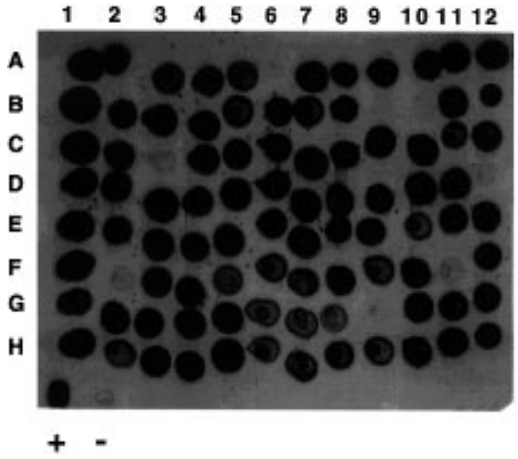

Figure 5. Dot blot analysis of the cloned PCR products as shown in Figure 4. PCR amplified products as shown in Figure 4 were denatured, dot-blotted and UV cross-linked onto a Duralon membrane (Stratagene, La Jolla, CA). The membrane was then hybridized with fluoresceinated DNA probe generated from the 180 kb pNL1 plasmid. Hybridization signal was detected with FluorImager SI (Molecular Dynamics, Sunnyvale, CA). +, positive control with *Sphingomonas* F199 genomic DNA. –, negative control with pCR-script Amp SK(+) plasmid DNA.

Figure 6. Long range PCR of T7-dN_6 Klenow-synthesized molecules. Lane 1, PCR with TaqPlus using low salt buffer; lane 2, *Hin*dIII digested lambda DNA; lane 3, PCR with TaqPlus using high salt buffer. Sizes in kb are shown to the right.

TRHA products were derived from pNL1 plasmid

The coverage of the TRHA products was tested using a miniset of cosmid clones from pNL1 plasmid (1). Six cosmid clones with an average insert size of 40 kb that provide overlapping coverage of the 180 kb plasmid were completely digested with *Eco*RI and run on a 0.8% agarose gel (Fig. 3A). The gel was Southern blotted and hybridized with fluoresceinated DNA probes generated from the TRHA products (Fig. 2, lane 5). Hybridization results (Fig. 3B) showed that all the *Eco*RI fragments including those of ~500 bp were represented, but some fragments had low hybridization signals. We speculate that those fragments with low intensity may represent the region of the plasmid with non-random sequences such as repetitive sequences, GC or AT rich regions.

TRHA products (Fig. 2, lane 5) were blunt-end ligated into pCRscript plasmid vector. The insert sizes of putative clones were determined by direct PCR amplification using bacterial cells carrying the recombinant plasmid as described in the Methods section. It was found that the insert sizes of the clones range from 0.3 to 1.3 kb with an average insert size of 0.5 kb (Fig. 4). To further confirm that the insert is derived from pNL1, PCR amplified products as shown in Figure 4 were dot-blotted onto a Nylon membrane and hybridized with a fluoresceinated DNA probe generated from pNL1 plasmid DNA. Only clones without an insert as revealed by PCR amplification did not have a hybridization signal (Fig. 5). Thus, the results indicated that all the inserts were derived from pNL1.

Randomness of the TRHA products

A total 593 clones were sequenced with T3 primers using reagents from ABI PRISM Dye terminator cycle sequencing ready reaction kit with ampliTaq DNA polymerase FS and analyzed with an ABI377 sequencer (Perkin Elmer, Foster City, CA). After trimming away the vector sequences, we obtained a total sequence of 232.5 kb and over twenty putative genes were identified by searching non-redundant protein and nucleic acid databases at the National Center for Biotechnology Information using Blastn and Blastx via the Internet. Using sequence analysis software AssemblyLIGN (Oxford Molecular, Campbell ,CA), sequences from these 593 clones were assembled into 62 contigs and 99 individual sequence fragments with a total unique sequence of 86.3 kb. Among these 593 clones, 17 clones were concatemers which were resulted from the annealing of T7 primer sites at the ends of two or more smaller Klenow-synthesized molecules. However, we did not detect any clones that align to more than one contig.

To analyze the randomness of the TRHA product, the following statistical analysis was performed. Assuming every fragment of the 180 kb pNL1 plasmid was equally amplified, we can apply the Poisson distribution equation to calculate the probability that a base is not sequenced. That is $P = e^{-m}$, where m is the sequence coverage (total sequence information obtained/size of the original template) (8). In our case, m is equal to 1.29 (i.e. 232.5/180). Thus, the expected P value will be 0.28, which means 28% of the pNL1 plasmid will not be sequenced. However, the experimental result for the P value is 0.52 (i.e. 1–86.3/180) since we have obtained 86.3 kb unique sequence out of the 180 kb pNL1 plasmid. This discrepancy shows that the TRHA products are not fully random. The hybridization result (Fig. 3) also indicated that some regions of the plasmid pNL1 were under-represented in the TRHA products. To calculate the fraction (f) of pNL1 preferentially amplified as the TRHA products, we applied the following formula, $P = e^{-m} + (1 - f)$. When m approaches infinity, P will equal $(1 - f)$, which is the fraction of sequence under-represented in the TRHA products. By substituting the numbers from the experimental result ($P = 0.52$ and $m = 1.29$), f is 0.76, which suggests that 76% of the total pNL1 plasmid sequence were preferentially amplified and represented in the TRHA products. In contrast, a mathematical model (9) suggested that only 37% of the DNA template are preferentially amplified in T-PCR (5). The difference is probably due to the Klenow step used in this protocol which generates longer primary products for subsequent PCR amplification. The sizes of the Klenow-synthesized molecules range from ~28 bp to >23 kb (Fig. 2, lane 2). Using a long range PCR amplification protocol, we were able

to amplify PCR products up to 23 kb from the Klenow-amplified products (Fig. 6). Such longer products are likely to cover regions of pNL1 not represented in the TRHA products and will be useful for making a lambda library for gap closure. Currently we are designing primers from the ends of each contigs and trying to get the rest of the sequence of pNL1 from cosmid clones by primer walkings. Alternatively, we are also trying to design tagged random hexamer with a bias towards GC or AT so as to enrich regions which were under-represented in the TRHA products in this study.

Conclusion

A simple, fast and efficient method was described for making a plasmid library from 100 ng of starting material. After sequencing 593 clones and obtaining 232.5 kb total sequences, we estimated that 76% of the original DNA template were preferentially amplified and represented in the TRHA products. This method can be applied to any situation where the amount of RNA or DNA is limited. For example, only a small quantity of nucleic acid materials can be obtained from (i) soil and subsurface environment with low biomass, (ii) restriction DNA fragments purified from agarose gels or YAC chromosomal DNA purified from pulsed-field gel electrophoresis, or (iii) a single cell. Large sequencing of BAC (bacterial artificial chromosome), PAC (P1 clones) and cosmids are taking place as the human genome project is going into its sequencing phase. Nebulization or sonication are currently used to make plasmid libraries from cosmid clones, PAC clones or BAC clones for sequencing. TRHA may be a timely method for making plasmid subclone libraries for initial random sequencing. Since only a small quantity of DNA is needed, miniprep of BAC, PAC or cosmid will be sufficient for library construction. After initial sequencing, the rest of the sequence may be obtained by PCR amplification of sequence gaps and primer walkings.

ACKNOWLEDGEMENTS

We thank Susan Varnum, Sarah Thruston, Toyoko Tsukuda and Rita Cheng for discussions and reading the manuscript. We also thank Ellen Sisk for making the tagged random hexamer. This work was supported by DOE contract DE-AC06-76RLO 1830.

REFERENCES

1 Fredrickson,J.K., Brockman ,F.J., Workman,D.J., Li,S.W. and Stevens,T.O. (1991) *Appl. Environ. Microbiol.*, **57**, 796–803.
2 Stillwell,L.C., Schneider,R.P., Thurston,S.J., Romine,M.F., Fredrickson,J.K. and Saffer,J.D. (1995) *J. Bacteriol.*, **177**, 4537–4539.
3 Kinzler,K. and Vogelstein,B. (1989) *Nucleic Acids Res.*, **17**, 3645–3653.
4 Zhang,L., Cui,X., Schmitt,K., Hubert,R., Navidi,W. and Arnheim,N. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 5847–5851.
5 Grothues,D., Cantor,C.R. and Smith,C.L. (1993) *Nucleic Acids Res.*, **21**, 1321–1322.
6 Fredrickson,J.K., Balkwill,D.L., Drake,G.R., Romine,M.F., Ringlber,D.B. and White,D.C. (1995) *Appl. Environ. Microbial.*, **61**, 1917–1922.
7 Maniatis,T., Fritsch,E.F. and Sambrook,J. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor University Press, Cold Spring Harbor, NY.
8 FleischmannR.D., Adams,M.D., White,O. *et al.* (1995) *Science*, **269**, 496–512.
9 Sun,F., Arnheim,N. and Waterman,M.S. (1995) *Nucleic Acids Res.*, **23**, 3034–3040.