Production of Shotgun Libraries Using Random Amplification

Article *in* BioTechniques · August 2001

DOI: 10.2144/01311rr02 · Source: PubMed

5 authors, including:

Victor Seguritan

37 PUBLICATIONS **4,053** CITATIONS

SEE PROFILE

Dong Han Choi

Korean Institute of Ocean Science and Technology

89 PUBLICATIONS **1,700** CITATIONS

SEE PROFILE

Anca M Segall

San Diego State University

89 PUBLICATIONS **3,355** CITATIONS

SEE PROFILE

Research Report

Production of Shotgun Libraries Using Random Amplification

BioTechniques 31:108-118 (July 2001)

F. Rohwer, V. Seguritan[1], D.H. Choi[2], A.M. Segall[1], and F. Azam
Scripps Institution of Oceanography, San Diego, [1]San Diego State University, San Diego, CA, USA, and [2]Seoul National University, Seoul, Korea

ABSTRACT

In the following report, thermal cycling coupled with random 10-mers as primers was used to construct randomly amplified shotgun libraries (RASLs). This approach allowed shotgun libraries to be constructed from nanogram quantities of input DNA. RASLs contained inserts from throughout a target genome in an unbiased fashion and did not appear to contain chimeric sequences. This protocol should be useful for shotgun sequencing the genomes of unculturable organisms and rapidly producing shotgun libraries from cosmids, fosmids, yeast artificial chromosomes (YACs), and bacterial artificial chromosomes (BACs).

INTRODUCTION

Whole genome sequencing has revolutionized biology (4). Shotgun sequencing, the preferred method for producing genomic sequences, requires high-quality libraries that (*i*) contain fragments completely covering the genome in an unbiased fashion and (*ii*) have inserts of a standard size, usually about 1 kb, so that bidirectional sequencing completely covers the fragment (reviewed in Reference 3). Shotgun libraries are traditionally produced by shearing the target DNA with enzymes (e.g., partial digestion with DNase I in the presence of Mn^{2+}) or with physical force (e.g., with a nebulizer or sonicator). The resulting fragments are then size-selected on a gel, blunt-ended using enzymes (e.g., Klenow, T4 DNA polymerase, mung bean nuclease), and cloned into an appropriate vector. In an alternative approach, the genomic DNA is partially digested with a frequently cutting restriction enzyme (e.g., *Sau*3AI) and then cloned into a complementary site within the vector. Both of these approaches suffer from certain limitations. First, relatively large quantities of genomic DNA are required (>1 μg). Second, DNA from certain genomes (e.g., bacteriophage) often contain modified nucleotides that cannot be directly cloned into *E. coli*. Third, both blunt-ending and ligation of blunt-ends are troublesome (2), and problems during these steps frequently lead to low numbers of clones relative to the amount of starting DNA. Finally, though partial digestion with a restriction enzyme avoids the need for blunt-ending, this procedure often produces biased libraries because of preferential cutting by the restriction enzyme.

Traditional methods of producing shotgun libraries are appropriate as long as the DNA to be sequenced can be isolated in relatively large amounts. However, DNA availability is a major limitation for the majority of microbes because they are unculturable (5). For example, marine Archaea are major components of the microbial food web in both the deep sea and in surface waters (7). Nonetheless, not one marine Archaea has been successfully cultivated. Similarly, it is impossible to isolate sufficient quantities of DNA for traditional shotgun libraries from plasmid, transposons, and bacteriophage that are associated with the unculturable microorganisms. One novel approach to this problem is to immortalize large fragments from uncultured microbes in cosmids, fosmids (9), yeast artificial chromosomes (YACs), or bacterial artificial chromosomes (BACs). Such libraries can then be probed for specific 16S rDNA sequences. Once a clone containing the desired DNA is identified, the ends of the insert can be sequenced, and probes can be made to identify overlapping clones. By repeating this process, it should be possible to assemble a complete genome of sequential clones that can then be sequenced [i.e., sequence-tagged-connector strategy (2)]. Three problems complicate this approach: (*i*) it is time consuming to

construct the large insert libraries, screen them, and then sequence the inserts; (*ii*) large DNA inserts are notoriously unstable; and (*iii*) no library can be made if the target organism has modified DNA that is unclonable.

In the following report, we describe the production of randomly amplified shotgun libraries (RASLs) produced by (*i*) amplifying the target DNA using Vent DNA polymerase (New England Biolabs, Beverly, MA, USA) in conjunction with random 10-mer primers, (*ii*) gel isolating the desired fragments size, and (*iii*) cloning. This protocol uses nanogram quantities of DNA to produce plasmid libraries suitable for shotgun sequencing. This procedure is also extremely rapid and requires very little manipulation of the DNA. Additionally, because this method initially amplifies the DNA in vitro, complications arising from modified nucleotides are avoided. This protocol should be especially useful for shotgun sequencing the genomes of unculturable organisms and rapidly producing shotgun libraries from cosmids, fosmids, YACs, and BACs.

MATERIALS AND METHODS

Standard Random Amplification

Coliphage λ genomic DNA (18 ng) was subjected to 25 thermal cycles (94°C for 1 min, 35°C for 1 min, and 72°C for 3 min). A "standard reaction" contained the target DNA, 5 mM $MgSO_4$ (New England Biolabs), 1 mM dNTPs (Apex BioResearch Products, Research Triangle Park, NC, USA), 34 μM random 10-mers [San Diego State University Microchemical Core Facility (SDSU-MCF), San Diego, CA, USA], 2 U Deep Vent™ DNA polymerase, and 1× ThermoPol Buffer (both from New England Biolabs) in a final volume of 50 μL. Where indicated, different concentrations of input DNA, dNTPs, and random 10-mers were used. After thermal cycling, 10 μL of the products were analyzed on a TAE agarose gel. The products in the rest of the reaction were purified using the UltraClean™ PCR Clean-up kit (Mo Bio Laboratories, Solana Beach, CA, USA) and quantified by spectrophotometry (i.e., A_{260}).

Random Amplification with Various Thermally Stable DNA Polymerases

Coliphage λ genomic DNA (16 ng) was randomly amplified in reactions containing 2 U Deep Vent, Vent (exo$^-$), *REDTaq* (Sigma, St. Louis, MO, USA), *KlenTaq* (AB Peptides, St. Louis, MO, USA), or *Pfu* DNA polymerases (Stratagene, La Jolla, CA, USA). In each case, the buffer used was that supplied by the manufacturer. Each reaction was supplemented with 5 mM Mg^{2+} in the form of $MgSO_4$ for Deep Vent and Vent (exo$^-$) DNA polymerases, and $MgCl_2$ for *REDTaq*, *KlenTaq*, and *Pfu* DNA polymerases.

Production of Coliphage λ RASL and Sequencing

A standard reaction was assembled with 18 ng coliphage λ DNA and split into five equal parts. After thermal cycling, the five aliquots were pooled, and the resulting products were separated on an agarose gel. Fragments between 1000 and 2000 bp were isolated and purified (UltraClean GelSpin Gel Purification Kit; Mo Bio Laboratories). The DNA was subsequently cloned into pCR4Blunt-Topo (Invitrogen, Carlsbad, CA, USA) as recommended by the manufacturer and introduced into TOP10 cells by heat shock transformation. Fifty-eight clones were picked, plasmids purified (6 Minute Plasmid Miniprep Kit; Mo Bio Laboratories), and sequenced [ABI PRISM® BigDye™ Terminators on an ABI 377XL sequencer (Applied Biosystems, Foster City, CA, USA) at the SDSU-MCF]. The first 500 bp of the resulting sequence were analyzed using Basic Local Alignment Search Tool (BLAST) (1).

Southern Blots

Lambda genomic DNA was digested with *Sau*3AI, *Rsa*I, *Hha*I, or *Msp*I (New England Biolabs), the resulting fragments separated on a TAE agarose gel, photographed, and the DNA was denatured, neutralized, and transferred to MagnaCharge™ membranes (Osmonics, Westborough, MA, USA) as suggested by the manufacturer. A random amplification probe was made by adding ^{32}P-dCTP (Amersham Pharma-

cia Biotech, Piscataway, NJ, USA) to a scaled-down (i.e., 20 µL) standard random amplification reaction. After thermal cycling, the probe was denatured at 96°C for 10 min, snap-cooled on ice, and then added to the hybridization solution. After washing, the blot was exposed to film, stripped, and then hybridized to a second probe that was made with Klenow and random hexamers (Random Primed DNA Labeling Kit; Roche Molecular Biochemicals, Indianapolis, IN, USA).

Production of Vibriophage Φ16 Libraries

Two separate standard random amplification reactions were performed on Vibriophage Φ16 genomic DNA (100 ng). In the first reaction, the whole mixture was cycled together in the same tube (Single-RASL). In the second reaction, the mixture was split into 10-µL aliquots that were thermal-cycled in separate tubes and then pooled for the rest of the library construction (Split-RASL). A second aliquot of Vibriophage Φ16 genomic DNA (5 µg) was physically sheared with a nebulizer, a third aliquot (5 µg) was partially digested with DNase I (8), and a fourth aliquot (5 µg) was partially digested with *Sau*3AI. All four samples were loaded onto a TAE agarose gel, and the 500–2000-bp fragments were gel-purified. The randomly amplified, DNase I-digested, and nebulized samples were then blunt-ended by treating with T4 DNA polymerase and Klenow fragment, extracted with phenol:chloroform (pH 7.0; 1:1), ethanol-precipitated, washed, dried, "tailed" with *REDTaq* DNA polymerase (fragments, 1× *REDTaq* Buffer, 1 mM dNTP, and 1 U *REDTaq*), and then cloned into pCR4-Topo (Invitrogen). The *Sau*3AI fragments were cloned into pZErO-2 (Invitrogen) linearized with *Bam*HI (New England Biolabs).

Stacking Characteristics of the Libraries

The stacking characteristics of approximately 200 fragments from the Vibriophage Φ16 Single- and Split-RASLs, the *Sau*3AI, and a previously constructed DNase I library [made from the marine bacteriophage Roseophage SIO1 (8)] were derived by first using Gene Codes Sequencher™ 3.0.6 software (Ann Arbor, MI, USA) to align sequences into contigs. The stacking was determined by summing the length of cloned fragments within the resulting contigs and dividing by the length of the contig (see legend for Figure 4B). This reflects an average number of clones used to contribute any base within the consensus sequence of the contig. Stacking calculations were performed on 20 contigs from each library.

Unless otherwise noted, reagents were purchased from Sigma.

RESULTS AND DISCUSSION

To determine if repeating cycles of random amplification could be used to produce quantities of DNA useful for shotgun cloning (e.g., >1 µg) from nanogram quantities of starting material, 18 ng coliphage λ genomic DNA were subjected to 25 thermal cycles using the standard reaction conditions, except the concentration of the $MgSO_4$ was manipulated from 0, 2.5, and 5 mM. This procedure yielded large quantities of DNA in the 5 mM $MgSO_4$ reaction (Figure 1). Using spectrophotometry (A_{260}), it was shown that 18 ng coliphage λ genomic DNA resulted in 1.67 µg of products, an increase of 93×. Interestingly, the majority of random amplification products were confined to the 500–2000 bp size range.

Vent DNA polymerase was initially used because of its low error rate and lack of terminal transferase activity. To determine how other thermally stable polymerases behaved in this assay, 16 ng coliphage λ genomic DNA were randomly amplified with 2 U Deep Vent, Vent (exo$^-$), *REDTaq*, *KlenTaq*, and *Pfu* DNA polymerases. Figure 2A shows that Deep Vent produced the most random amplification products (1.78 µg), followed by Vent (exo$^-$) (1.50 µg), which was greater than *Pfu* (0.974 µg), *KlenTaq* (0.94 µg), which was greater than *REDTaq* (0.86 µg).

To determine the limitations imposed by different reaction components, the concentrations of random 10-mers and dNTPs were varied in separate reactions. When 18 ng input coliphage λ

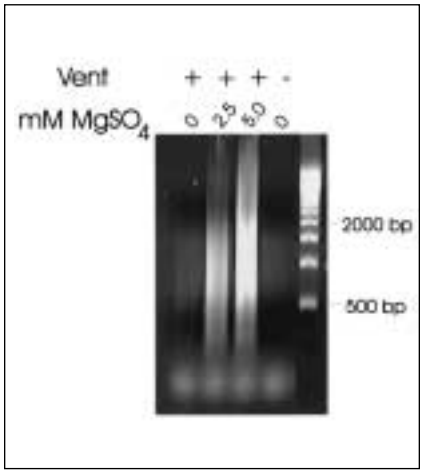

Figure 1. Random amplification of coliphage λ genomic DNA by thermal cycling. Coliphage λ genomic DNA (18 ng) was subjected to 25 thermal cycles (94°C for 1 min, 35°C for 1 min, and 72°C for 3 min) using the standard reaction conditions (i.e., 1 mM dNTPs, 34 µM random 10-mers, 2 U Deep Vent DNA polymerase, and 1× Vent buffer in a final volume of 50 µL) with varying concentrations of the $MgSO_4$.

DNA were used, the highest concentration of the random 10-mers tested (34 μM) gave the greatest amount of amplification products (Figure 2B). This reaction required that the highest amount of $MgSO_4$ (5 mM) also be used, presumably because the primers were binding Mg^{2+}. This result suggested that the primers were limiting under these reaction conditions. However, 17 μM of the random 10-mers were sufficient for maximum amounts of amplification product when 1.8 ng input DNA were used (Figure 2B, last two lanes). Together, these results suggest that if very small amounts of input DNA are being used (≤1.8 ng), then 34 μM of the random 10-mer primers with 5 mM $MgSO_4$ were sufficient for maximum product yield. As the amount of input DNA increases, however, it may be prudent to increase both the concentration of random 10-mers and $MgSO_4$. Similarly, when the concentration of dNTPs was varied from 0.025 to 10 mM, it was found that 0.5 mM dNTPs worked the best under our reaction conditions (very little difference was found between 1 and 0.5 mM dNTPs; data not shown). When more $MgSO_4$ was added to the higher dNTP reactions, the amount of products roughly equaled that found in the 0.5 mM dNTP reaction (data not shown), suggesting both that more dNTPs were not needed for this concentration of input DNA and that the excessive amounts of dNTP were absorbing Mg^{2+} in the reactions containing higher dNTP concentrations (i.e., >0.5 mM dNTPs, where fewer products were made). Together, these results show that our standard reaction conditions were

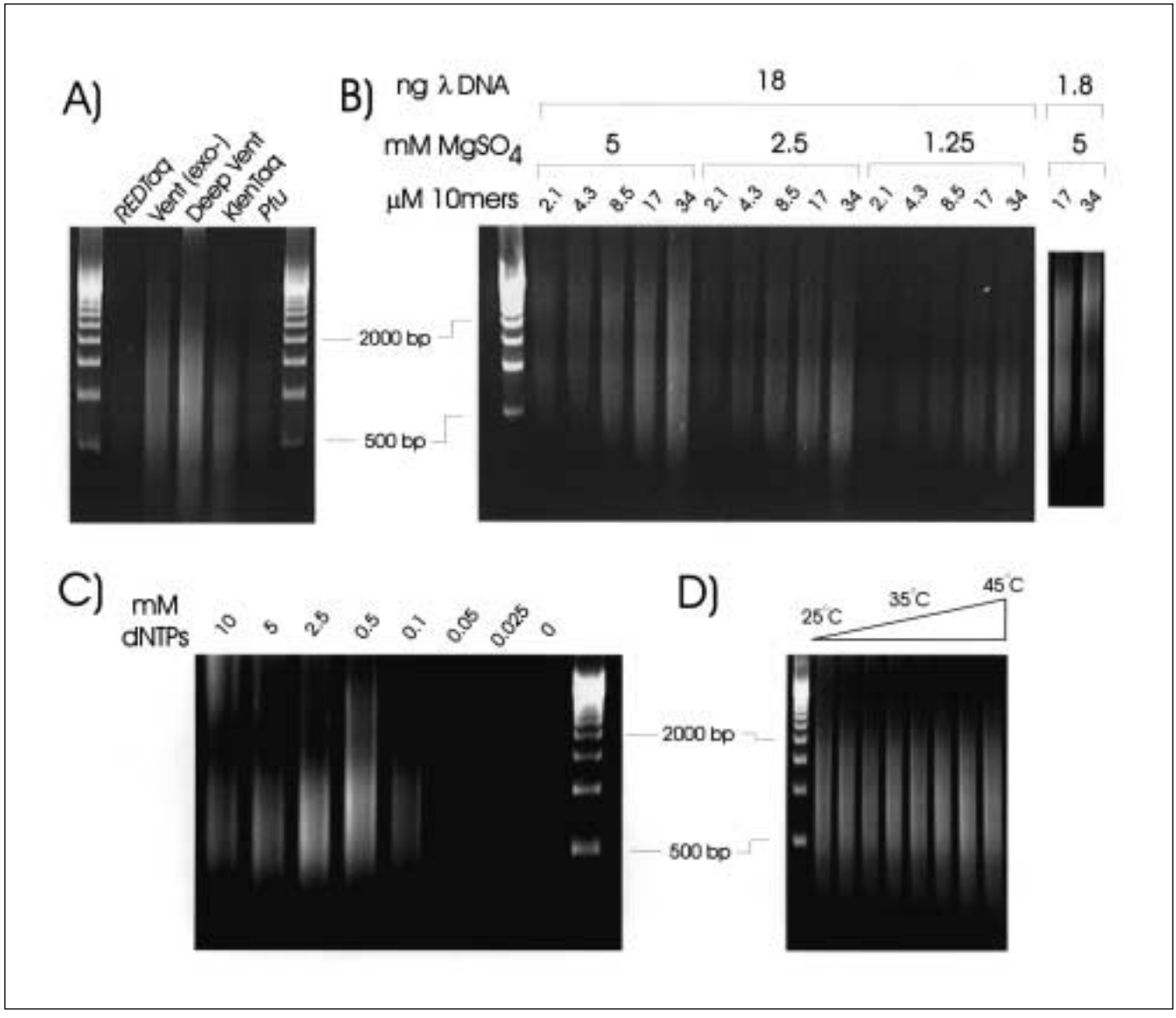

Figure 2. Optimization of RASL protocol. (A) Effects of various thermally stable polymerases on random amplification of coliphage λ genomic DNA. Coliphage λ genomic DNA (16 ng) was randomly amplified in reactions containing 2 U Deep Vent, Vent (exo-), *REDTaq*, *KlenTaq*, or *Pfu* DNA polymerases and supplemented with 5 mM of Mg^{2+}. (B) Effects of varying random 10-mer concentrations on the amount of amplification products. (C) Effects of varying dNTP concentration on the amount of amplification products. (D) Effect of varying annealing temperature on the amount of amplification products.

Table 1. Results from Coliphage λ Sequencing

1509	17019	31467
2011 (T to C; **142**; 1-bp gap; **254**)	19565 (T to A; **110**; T to G; **223**)	31674
2278	19012	31971 (G to A; **434**)
2987	19046 (G to T; **314**)	32098 (1-bp gap; **117**; 1-bp gap; **483**)
4470 (T to A; **435**)	19807 (G to C; **26**; C to A; **470**)	32193
6734 (1-bp gap; **398**)	20807	32228 (1-bp gap; **93**)
7586 (T to G; **74**; G to C; **339**)	24876	35980
7987 (A to T; **45**; A to T; **414**)	24909 (T to C; **372**;1-bp gap; **453**)	35483
8300 (1-bp gap; **176**)	25012 (T to A; **358**)	35782 (G to C; **432**; T to G; **457**)
8346	25908 (1-bp gap; **241**)	36583
8675 (A to G; **265**; 1-bp gap; **312**)	25994	36758
8882 (1-bp gap; **27**)	26878 (T to C; **64**)	37865 (T to C; **49**;1-bp gap; **125**;1-bp gap; **431**)
9489	27758 (G to A; **165**; 1-bp gap; **198**)	40134 (A to G; **267**; G to T; **285**)
9675 (1-bp gap; **334**; 1-bp gap; **376**)	27786 (T to A; **436**)	40276 (G to C; **143**; C to G; **452**)
12876 (T to A; **192**)	27893	42976 (G to A; **322**; A to G; **418**)
13873 (T to C; **81**; T to G; **153**)	29011 (T to C; **277**; G to T; **479**)	44390
13965	29150 (T to G; **178**; 1-bp gap; **235**)	44687
14576 (G to C; **258**; C to G; **312**; T to A; **397**)	29399 (A to T; **396**)	46105 (T to A; 1-bp gap)
15923 (A to T; **439**)	29981	46976 (G to C; T to G)
16598 (T to A; **292**; C to T; **455**)		

The starting position of each 500-bp fragment is given relative to the complete coliphage λ genome (J02459). Ambiguities are shown in parentheses and include gaps or base pair changes. No chimeric molecules were observed. In the case of base pair changes, the base pair in the JO2459 genome is reported first, and its substitution in the randomly primed library is reported second. The position of each mismatch or gap is shown in **bold**.

appropriate for small concentrations of input DNA and that it may be necessary to modify the conditions to maximize the yield of final products if larger amounts of input DNA are used.

The effect of the annealing temperature on the production of amplification products was also tested. As shown in Figure 2D, no discernable effect on quality or quantity of amplification products was detected when the annealing temperature was varied between 25°C and 45°C.

To determine if the makeup of the input DNA had significant effects on either the amount or quality of amplification products, reactions with dilutions of coliphage T7, λ, or T4 genomic DNAs were performed. Figure 3 shows that there were differences in the amount of amplification product from the three different genomes. Eighteen nanograms of genomic coliphage λ DNA produced greater than 1 µg amplification products, whereas 180 ng in-

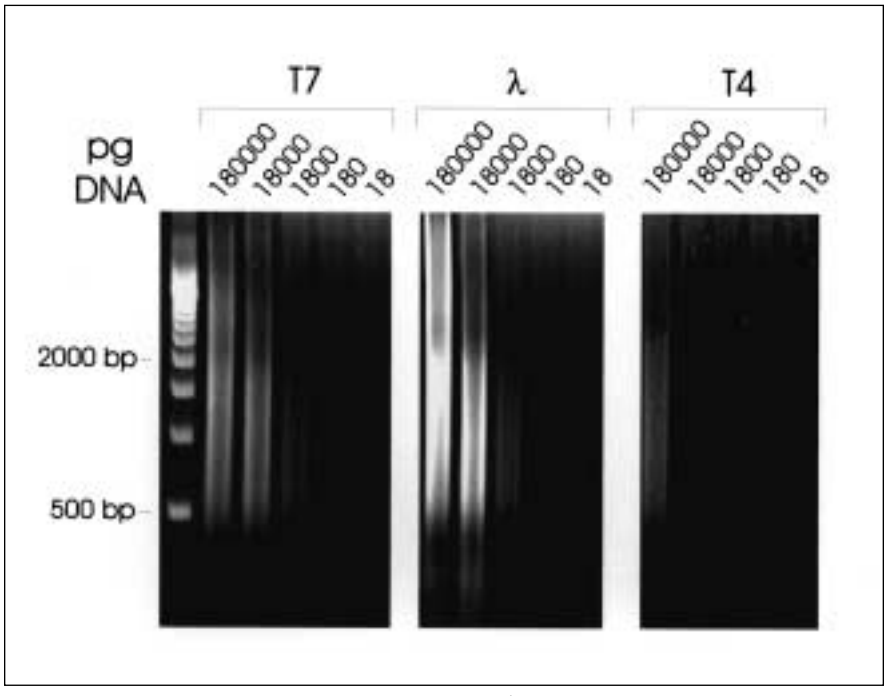

Figure 3. Effects of different input DNA (coliphage T7, λ, or T4) on the amount and size range of random amplification products.

put coliphage T7 and T4 genomic DNA were required to produce greater than 1 µg amplification products. The quality (i.e., the size range) of the random amplification products was equivalent among the three input DNAs.

Potential problems with using random amplification products for the construction of shotgun libraries include (*i*) biased amplification that favors some products over others, (*ii*) chimeric amplification products, and (*iii*) introduction of errors into the sequence by the thermally stable polymerase or imprecise annealing of random primers. The observation that the random amplification products were spread through a relatively large size range with no prominent bands argues against preferential amplification of selected sequences (e.g., Figure 1). However, to empirically rule out this possibility, a RASL of coliphage λ DNA was produced by randomly amplifying the genomic DNA, gel-isolating the 1000–2000-bp fragments, and cloning into pCR4Blunt-Topo. Fifty-eight colonies from the resulting RASL were then sequenced, and the first 500 bp of each sequence were compared against the GenBank® data-

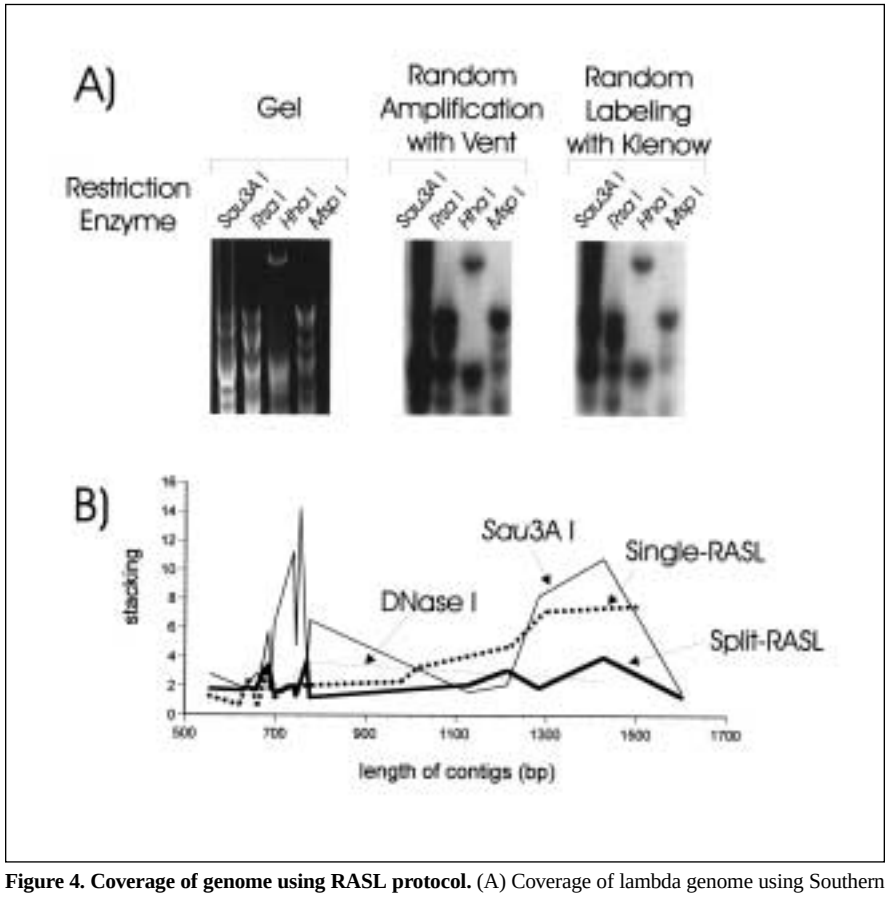

Figure 4. Coverage of genome using RASL protocol. (A) Coverage of lambda genome using Southern blotting. Gel = photograph of gel before transfer; Random amplification with Vent = probed with randomly amplified probe (i.e., made with Vent, random 10-mers, and thermal cycling); Random labeling with Klenow = probe made with Random Primed DNA Labeling Kit from Roche Molecular Biochemicals, which uses Klenow and random 6-mers. (B) Stacking characteristics of Vibriophage Φ16 Split-RASL versus Single-RASL versus the *Sau*3AI library versus Roseophage SIO1 DNase I library. Twenty contigs from each library were analyzed using the formula:

$$\text{Stacking} = \frac{\sum_{i=0}^{n} C_i}{|C|}$$

where C_i is the length (in base pairs) of the ith fragment of contig C and $|C|$ is the length of contig C.

base (Table 1). For all 58 sequences, almost perfect matches to complete coliphage λ genome (GenBank accession no. JO2459) were obtained. Only 59 ambiguities, either base substitutions or gaps, were found in 29 000 bp of sequence analyzed. This represents an error rate of 0.2%. By choosing the first 500 bp from each sequence, miscalling of bases during sequencing was effectively eliminated. Additionally, if miscalling of bases were a problem, it would be expected that the positions of the identified gaps and mismatches would be more common at the 3′ ends of the fragments. The positions of the identified gaps and mismatches were spread throughout the sequenced fragments (Table 1). Therefore, the changes were most likely due to (*i*) mistakes introduced during the RASL protocol or (*ii*) actual differences between the coliphage λ DNA we used in our studies and the λ DNA used in the original sequencing. We observed no chimeric molecules in the coliphage λ RASL.

The 58 sequences from the coliphage λ RASL appear to be spread throughout the genome (Table 1), strongly suggesting that the procedure amplifies all parts of the target DNA. However, to better determine the diversity of sequences represented in the library, Southern blots using coliphage λ genomic DNA as the target and radiolabeled random amplification products from coliphage λ genomic DNA as the probe were performed. It was assumed that if the random amplification products truly represent all of the target genome, then all of the bands on the blot should be bound by the probe. By comparing the gel and blots in Figure 4A, it was concluded that the random amplification probe contained regions complementary to all of the bands present in the gel. As a control, the membrane was stripped and probed with a coliphage λ probe made by random priming with Klenow and random hexamers. The random amplification probe had at least as good coverage as the Klenow-derived probe, again suggesting that the RASL method gave complete coverage of the genome. Similar results were obtained with Southern blots of coliphage T7 and *Haloferax volcanii* genomic DNA (data not shown).

To determine if the RASL technique could be used to obtain complete genomic libraries, we tested it on Vibriophage Φ16 (6). Previous attempts to make shotgun libraries of this marine phage had failed, despite numerous attempts with a variety of approaches (e.g., different DNA isolation techniques, shearing, and blunt-ending techniques). Two standard 50 µL random amplification mixtures containing 100 ng Vibriophage genomic DNA were assembled. One of the mixtures was evenly split into five 10-µL mini-reactions and aliquoted into separate tubes (Split-RASL). The second mixture was left as a single 50-µL reaction (Single-RASL). The six tubes were then cycled together. After thermal cycling, the tubes from the Split-RASL reaction were pooled. Separate amplification reactions were performed to avoid "jack-pots" (i.e., when some areas of the target DNA are overamplified in the earlier amplification steps), which would compromise the randomness of the resulting library. In conjunction with these random amplifications, 5 µg Vibriophage DNA were sheared using either a nebulizer or DNase I. The 500–2000-bp products from the two random amplification reactions, as well as the nebulizer and DNase I-sheared DNA were gel-isolated, blunt-ended with Klenow and T4 DNA polymerase, "tailed" with *REDTaq* DNA polymerase (to add a 3′ A), and cloned into pCR4-Topo. Both the random amplification protocols (i.e., Split- and Single-RASL) resulted in greater than 15 000 colonies each, while the physical shearing protocol only resulted in 68 colonies, and there were no colonies in the DNase I library. Another Vibriophage Φ16 shotgun library was made by partially digesting 5 µg DNA with *Sau*3AI, gel-isolating the 500–2000-bp fragments, and then cloning into pZErO-2 linearized with *Bam*HI. More than 200 colonies were obtained using this approach. Approximately 200 clones from the *Sau*3AI , the Split-RASL, and the Single-RASL were then sequenced (the DNase I and nebulizer-generated libraries were not sequenced because of the limited number of clones).

The average number of clones used to contribute any base (i.e., the stacking characteristics) within 20 contigs of the Vibriophage Φ16 Split-RASL, Single-RASL, *Sau*3AI, and a previously constructed DNase I library (made from Roseophage SIO1; see Reference 8) were determined using the equation shown in the legend for Figure 4B. As shown in Figure 4B, the stacking remained minimal as the contig size increased for Vibriophage Φ16 Split-RASL. In comparison, the Vibriophage Φ16 *Sau*3AI displayed a high degree of stacking, whereas very little stacking was associated with the Roseophage SIO1 DNase I library. The Single-RASL had a stacking pattern intermediate between the *Sau*3AI library and the Split-RASL. Taken together, these results show that the Split-RASL protocol provides coverage of the genome in a manner analogous to that obtained by DNase I and works on DNA that was essentially unclonable using other methods. Using the Split-RASL library, we have now completely sequenced the Vibriophage Φ16 genome (FR, VS, and AS; unpublished results). Our results

indicate that the stacking observed in the Single-RASL library was due to the fact that some areas of the target DNA were overamplified in the earlier steps and therefore contributed more to the final products. Because these early events are random, splitting the reaction into separate tubes helps provide more complete coverage.

As genomic sequencing of environmental isolates becomes more important, constraints such as the amount of DNA and its clonability will become important factors. The RASL protocol presented here avoids the problems associated with small amounts of DNA by amplifying the starting material (up to 93×). This may make it possible to sequence uncultured microbial genomes. For example, the RASL input DNA could be obtained from uncultured microbes sorted on a flow cytometer or from genomes isolated using pulse field gel electrophoresis (PFGE). Recently, we have made a cyanophage library using PFGE-isolated DNA. A second advantage of the RASL procedure is that, by first amplifying the DNA in vitro, this approach avoids difficulties associated with modified DNA that cannot be cloned into *E. coli*.

Our data show that RASLs for bacteriophage-sized genomes (i.e., approximately 50 kb) are relatively unbiased. However, we have no information about larger genomes. It seems likely that certain regions of the genome may be preferentially amplified, either due to sequence differences (e.g., high GC regions) or chance events in the early cycles of the random amplification. However, the Vibriophage Φ16 sequence contains both poly(A) and poly(GC) tracts. Preferential amplification of specific regions in the Vibriophage Φ16 genome was not observed, suggesting that sequence differences are not a major factor. We presume that the use of Vent DNA polymerase, which is highly processive, was responsible for getting past these problematic template tracts. While this finding is not conclusive, it does suggest that the RASL protocol will amplify difficult DNA stretches. The influence of chance events (i.e., jack-pots) in the early amplification steps is probably more of a problem. In our initial studies, we observed some bias in RASLs. This bias was eliminated in later experiments by performing separate amplification reactions and then combining the products (e.g., Single- versus Split-RASL) (Figure 4).

Besides sequencing uncultured or unclonable genomes, the RASL protocol is a rapid way of constructing a shotgun library from any DNA source. Because there is minimal DNA manipulation, inexperienced laboratories will be able to rapidly make high-quality shotgun libraries. Finally, because the RASL protocol takes very little input DNA, it will be useful when sequencing large vectors such as cosmids, fosmids, YACs, and BACs, which normally produce very little DNA during isolation (10).

ACKNOWLEDGMENTS

This work was sponsored by National Science Foundation grant no. DEB9972054. Anca M. Segall received partial support from NSF CAREER AWARD 9733332. Victor Seguritan received partial support from NIH-Minorities in Biomedical Research Support (MBRS) grant no. R25GM58906 to the San Diego State University College of Sciences.

REFERENCES

1. **Altschul, S.F., T.L. Madden, A.A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D.J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*:3389-3402.
2. **Andersson, B., M.A. Wentland, J.Y. Ricafrente, W. Liu, and R.A. Gibbs.** 1996. A "double adaptor" method for improved shotgun library construction. Anal. Biochem. *236*:107-113.
3. **Frangeul, L., K.E. Nelson, C. Buchrieser, A. Danchin, P. Glaser, and F. Kunst.** 1999. Cloning and assembly strategies in microbial genome projects. Microbiology *145*:2625-2634.
4. **Fraser, C.M. and R.D. Fleischmann.** 1997. Strategies for whole microbial genome sequencing and analysis. Electrophoresis *18*:1207-1216.
5. **Fuhrman, J.A. and L. Campbell.** 1998. Microbial microdiversity. Nature *393*:410-411.
6. **Kellogg, C.A., J.B. Rose, S.C. Jiang, J.M. Turmond, and J.H. Paul.** 1995. Genetic diversity of related vibriophages isolated from marine environments around Florida and Hawaii, USA. Mar. Ecol. Prog. Ser. *120*:89-98.
7. **Massana, R., E.F. DeLong, and C. Pedros-Alio.** 2000. A few cosmopolitan phylotypes dominate planktonic archaeal assemblages in widely different oceanic provinces. Appl. Environ. Microbiol. *66*:1777-1787.
8. **Rohwer, F., A. Segall, G. Steward, V. Seguritan, M. Breitbart, F. Wolven, and F. Azam.** 2000. The complete genomic sequence of the marine phage Roseophage SIO1 shares homology with non-marine phages. Limnol. Oceanogr. *45*:408-418.
9. **Stein, J.L., T.L. Marsh, K.Y. Wu, H. Shizuya, and E.F. DeLong.** 1996. Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. J. Bacteriol. *178*:591-599.
10. **Vaudin, M., A. Roopra, L. Hillier, R. Brinkman, J. Sulston, R.K. Wilson, and R.H. Waterston.** 1995. The construction and analysis of M13 libraries prepared from YAC DNA. Nucleic Acids Res. *23*:670-674.

Received 18 September 2000; accepted 28 December 2000.

Address correspondence to:

Dr. Forest Rohwer
Marine Biology Research Division
Hubbs Hall 4200
Scripps Institution of Oceanography
San Diego, CA 92037, USA
e-mail: forest@ucsd.edu