Amplification of high-quantity serial analysis of gene expression ditags and improvement of concatemer cloning efficiency

Zeji Du, Angela D. Scott, and Gregory D. May
The Samuel Roberts Noble Foundation, Ardmore, OK, USA

BioTechniques 35:66-72 (July 2003)

As a method for accurate quantitative and qualitative analysis of cell transcription, serial analysis of gene expression (SAGE) has been widely used in biological, medical, and pharmaceutical areas of research. The numerous steps required in the SAGE protocol, however, limit the efficiency of SAGE library construction. PCR amplification of ditags and concatemer separation are two such steps. Primer-2, which was used to amplify ditags in the original SAGE protocol, was prone to self-anneal and resulted in a decreased efficiency of the PCR amplification step. We have designed a new primer pair, which substantially enhances the efficiency of PCR amplification. Incorporating this modification reduces the number of reactions required for SAGE library construction. Sixty 50-µL reactions are now sufficient for SAGE library construction (up to an 80% reduction in comparison to the original protocol). Concatemer separation by column filtration and subsequent gel separation has been modified and has proven to be more efficient than simple heat treatment of the ligation mixture. Both of the above modifications may also be suitable for the newly developed LongSAGE protocol.

INTRODUCTION

Serial analysis of gene expression (SAGE) is an approach that allows for the rapid and detailed analysis of the transcriptome (1). SAGE samples short sequence tags (14 bases) that contain sufficient sequence information to assign identity to the specific mRNA transcript from which the tag was derived (2). Tags are ligated to form concatemers that are cloned and sequenced. The frequency at which each tag appears in the SAGE library is an accurate estimate of the relative abundance of its corresponding mRNA transcript. SAGE has recently been used in a wide range of research areas (3,4).

Because of the low efficiency of the original SAGE protocol [Veculescu, V.E., L. Zhang, B. Vogelstein, and K.W. Kinzler, 1997. Serial Analysis of Gene Expression: Detailed Protocol (Version 1.0c). Available from Johns Hopkins Oncology Center and Howard Hughes Medical Institute], a number of modifications were made (5–9), resulting in the revised SAGE 2000 version (http://www.sagenet.org). This revised version 1.0e protocol substantially increased SAGE efficiency and frequency of success in library construction. However, the revised SAGE protocol still required large-scale PCR amplification of the SAGE ditags. Typically, 200–300 50-µL reactions were required for the construction of one microSAGE library. To reduce the scale of required reactions, we have designed a new pair of primers for ditag amplification.

We have successfully constructed several SAGE libraries following the revised version of the SAGE 2000 protocol; nonetheless, we did encounter intermittent failure because of low-avedge-length concatemers due to contamination with smaller DNA fragments. Minimal success as it relates to small insert size has also been experienced by other researchers (personal communications during the SAGE 2001 Conference, Frontiers in Transcriptome Exploration, September 8–11, 2001; Coronado, CA, USA). To resolve this problem, we have incorporated column filtration purification of concatemers before gel separation. By using this modification in SAGE library construction, the average number of tags per clone was increased significantly.

MATERIALS AND METHODS

Plant Total RNA Extraction and SAGE Library Construction

Plant leaf explants derived from two-week-old Medicago truncatula (A17) plants were grown at 24°C for 3 weeks in embryogenic callus induction medium to induce the formation of embryogenic callus (10). Callus tissue was flash-frozen in liquid nitrogen and then ground to a fine powder using a prechilled mortar and pestle. Total RNA was extracted using TRIZol® reagent (Invitrogen, Carlsbad, CA, USA) according to the manufacturer’s protocol. The quality and concentration of RNA were determined by spectrophotometer at 260- and 280-nm wavelengths. SAGE libraries, starting with 10 µg total RNA, were constructed following the microSAGE protocol version 1.0e, with the exception of the modifications detailed below.

Modification of Primer Design

PCR primers were redesigned toward the 3’ end of the linker sequence. The sequence of the new primer pairs modified for ditag PCR amplification are as follows: P1′: 5′-CTAGGCTTAAATAGGGACATG-3′; P2′: 5′-TAAACATGTACGGGGACATG-3′. Primers were synthesized (Integrated DNA Technologies, Coralville, IA, USA) and gel-purified. Ditag PCR amplification was performed in 96-well PCR plates (TempPlate®, USA Scientific, Ocala, FL, USA) containing 50-µL reactions/well using the GeneAmp® PCR System 2700 (Applied Biosystems, Foster City, CA, USA) with the following conditions: 1 cycle of 94°C for 1 min, followed by 30 cycles of 94°C for 30 s, 54°C for 1 min, and 70°C for 1 min, 1 cycle of 70°C for 5 min, and hold at 4°C.

Purification and Separation of Ditag Ligation Product

The second modification was based on the use of the QIAquick™ PCR purification filter column (Qiagen, Valencia, CA, USA). After ligation of the ditags at 16°C for 4–5 h, 40 µL of the provided binding buffer were added to the 10 µL ligation reaction. The mixture was subsequently applied to a QIAquick
purification filter column, as described in the provided protocol. The loaded column was washed twice with 500 µL of the provided washing buffer. The concatemerized DNA was eluted by adding 40 µL ½ LoTE (1.5 mM Tris-HCl, pH 7.5, 0.1 mM EDTA) to the column and incubating at room temperature for 2 min, followed by centrifugation at 16,000×g for 1 min. The elution step may be repeated by using the first eluant if necessary. The volume of eluted DNA solution was concentrated to 10 µL through vacuum drying or by ethanol precipitation followed by resuspension of the DNA in 10 µL ½LoTE. Size separation of the concatemerized DNA was performed by PAGE on an 8% polyacrylamide gel. Concatemerized DNA in the 600–1500 bp range was extracted from the polyacrylamide gel using standard procedures.

### Cloning and Nucleotide Sequencing of Concatemerized DNA

The vector pUC19 was digested with the restriction enzyme SphI at 37°C for 3 h and subsequently dephosphorylated using calf intestinal phosphatase (New England Biolabs, Beverly, MA, USA) at 37°C for 1 h. Ligation of concatemers with the previously digested pUC19 vector was performed overnight at 16°C. Ligation products were electroporated into ElectroMAX™ DH10B Cells (Invitrogen). Blue/white color selection on solid LB agar medium containing 100 µg/mL ampicillin was performed. White colonies were cultured in TB (Terrific Broth) medium in 96-well culture blocks (Greiner Bio-One, Frickenhausen, Germany) at 37°C, 350 rpm, for 22 h. Plasmid DNA was prepared by the Biomek™ 2000 Laboratory Automation Workstation (Beckman Coulter, Fullerton, CA, USA) according to standard protocols. DNA sequencing was performed using the ABI Prism® 3700 DNA analyzer (Applied Biosystems). DNA was sequenced using the ABI Prism® Large Fragment Electrophoresis Kit (Applied Biosystems). For some DNA samples, DNA was extracted from the polyacrylamide gel using standard procedures.

### RESULTS

#### Enhancement of SAGE Ditag Amplification Using Modified PCR Primers

We have successfully constructed several SAGE libraries using the original primers. However, low yields of the 102-bp PCR products were occasionally experienced. Upon further examination of the original primer sequence, it was discovered that a large percent of the P2 oligonucleotide is self-complementary (Figure 1A). Figure 1B illustrates that the majority of original primer P2 formed dimers, thus reducing the PCR efficiency. To improve PCR efficiency, we designed a new primer pair located at the 3’ ends of each linker and designated as P1′ (5′-CTAGGCTTAAATAGGACATG-3′) and P2′ (5′-TAACGTAGCCGGGACATG-3′) correspondingly. These changes made the size of the new PCR product approximately 65 bp. PAGE analysis did not reveal dimer formation of the new P1′ and P2′ primers (Figure 1B).

To compare PCR yields between the original and modified primers, 25 50-µL reactions were performed for each pair. Following ethanol precipitation, PCR products were analyzed by PAGE on a 20% polyacrylamide gel. The yield of PCR products from primer pair P1′/P2′ was significantly higher than that derived from primers P1 and P2 (Figure 2, A and B). DNA quantitation was determined by the ethidium bromide-DNA dot assay as described in Appendix C of the microSAGE protocol, and the results are provided in Table 1. The yield of PCR product was enhanced 3.8-fold by the utilization of the modified primers.

### Table 1. PCR Efficiency of Original and Modified Primers

<table>
<thead>
<tr>
<th>Primer Sets</th>
<th>Ditags (ng/reaction)</th>
<th>Reactions Required (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original primers</td>
<td>52 ± 15</td>
<td>200–300</td>
</tr>
<tr>
<td>Modified primers</td>
<td>198 ± 26</td>
<td>60–80</td>
</tr>
</tbody>
</table>

*Results were obtained from 50 reactions (50 µL/reaction) and shown as X ± SD.

The total number of reactions (50 µL/reaction) for one SAGE library construction.
**SHORT TECHNICAL REPORTS**

Table 2. Distribution of Tag Number per Clone Obtained by the Protocol Version 1.0e and the Modified Method

<table>
<thead>
<tr>
<th>Tags/Clone (n)</th>
<th>&lt;30</th>
<th>30–40</th>
<th>&gt;40</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heating step</td>
<td>22</td>
<td>44</td>
<td>30</td>
</tr>
<tr>
<td>Column step</td>
<td>2</td>
<td>16</td>
<td>78</td>
</tr>
</tbody>
</table>

Only one direction primer (M13 Forward) was used for sequencing each clone. A single automated sequencing run can yield 600–1000 bp readable sequence.

---

primer pair, thus substantially reducing the scale of the reactions required for SAGE library construction. We have successfully constructed several SAGE libraries with 60 50-µL reactions using the P1'/P2' primer pair.

Since the sizes of the PCR products derived from the P1'/P2' primer pair were shorter than those utilizing the original PCR primers, the banding pattern of NlaIII restriction enzyme-digested PCR products is different than as described in the original protocol (Figure 3). Band 3 contains ditags from 24–27 bp in length. The lowest molecular weight band (band 4) contains digested linkers (approximately 18 bp).

Improvement of Cloning Efficiency of Concatemerized DNA by Adding Column Purification Step

As described in the SAGE protocol version 1.0e, concatemerized DNA was heated at 65°C for 5 min before separation by PAGE on an 8% polyacrylamide gel. The appropriate-size concatemer fraction was extracted from the gel and cloned into pUC19. Sequencing results from 96 clones showed that approximately 23% of the inserts had tag counts of less than 30 (Table 2). These results indicate that simply heating the ligation mixture was not sufficient to remove contaminating small DNA molecules. To resolve this problem, we incorporated a QIAquick PCR filter column purification step before gel separation. Use of the filter column resulted in a shift in concatemer population toward the desired 600–1500 bp size range (Figure 4). The effects of the column method versus the heating method in the standard protocol were compared (Table 2). Column purification results in an increase in size of cloned concatamers. No difference in cloned concatemer size was observed when a heating step was added after column purification (data not shown).

DISCUSSION

Since its development, SAGE has proven to be a powerful tool in transcriptome analysis for many aspects of fundamental biology (11–14). While the original protocol has been refined (5–9), we found that the efficiency of the current version was still limited by two important points practically: PCR amplification of ditags and concatemer cloning. SAGE library construction requires 10 µg PCR-amplified ditags, necessitating 200–300 50-µL reactions according to the version 1.0e protocol. This large-scale PCR and the following purification steps were time- and labor-intensive. Upon re-examination of the standard amplification primers, it was apparent that the formation of P2:P2 primer dimers was likely to occur (Figure 1A). To determine if low PCR efficiency was due in part to the formation of primer dimers, we designed a new primer pair toward the 3’ end of each linker. This resulted in greater gel separation between the digested linkers and the 26-bp ditags in the PAGE purification step. The theoretical size of the PCR product and the digested linkers should now be 65 bp (Figure 2B) and 18 bp (Figure 3), respectively. Thus, the linker fragments could be easily separated from ditags (24–27 bp) after NlaIII digestion. Alternatively, biotinylated primers may be used for removal of linker contamination by streptavidin-coated magnetic beads (6). Application of the modified primer design increased PCR product yield by 3.8-fold. As a result, 60–80 50-µL reactions should be sufficient for the construction of a single SAGE library (Table 1). This modification has been successfully applied in our laboratory for the construction of two of our most recent SAGE libraries.

---

**Figure 3. Ditag band pattern after NlaIII digestion.** A 20% polyacrylamide gel, stained with SYBR Green I, showed the bands pattern of 65-bp ditags after NlaIII digestion. M, 25-bp DNA ladder; bands 1 and 2, undigested and partially digested ditags, respectively; band 3, digested ditags from 24 to 27 bp; band 4, 18-bp linkers.

**Figure 4. Separation of concatamers by PAGE before cloning.** The 8% polyacrylamide gel was stained with SYBR Green I. M, kb DNA ladder (Invitrogen); lane 1, the concatamer sample according to SAGE protocol version 1.0e (with heating only); lane 2, concatamer sample with QIAquick PCR column filtration purification step. Column purification led to a shift in the banding pattern towards the appropriate-size DNA molecules. DNA bands (600–1500 bp) were excised from the gel and extracted for cloning.
Although the heating step is postulated to increase SAGE cloning efficiency (5), a small percentage of lower molecular weight DNA still existed in the purified concatemers (Figure 4), resulting in low cloning efficiency. To better remove lower molecular weight DNA from the concatemer mixture, we substituted a QIAquick PCR purification filter column for the heating step. Incorporation of the column in concatemer purification results in a significant loss of contaminating low molecular weight DNA and, as a result, improves cloning efficiency (Figure 4 and Table 2). For convenience, DNA inserts cloned in pUC19 were sequenced with M13 Forward primer only. A single automated sequencing run can yield 600–1000 nucleotides of high-quality DNA sequence—equivalent to 45–70 SAGE tags.

We have demonstrated that the combined use of modified primers for ditag PCR amplification and the incorporation of a filter column purification step for concatemer separation reduced the required number of reactions and enhanced the efficiency of SAGE library construction. These modifications substantially improve the performance of SAGE. The modifications described here may also be applicable to LongSAGE methodology (15).

ACKNOWLEDGMENTS

We thank Amy Martin and Sahra Self for their assistance in SAGE library sequencing, Xianzhi He for technical advice, and John Cooper and Lahoucine Achnine for critical examination of the manuscript. SAGE software and protocols were kindly provided by the Johns Hopkins School of Medicine. This work was sponsored by The Samuel Roberts Noble Foundation.

REFERENCES


Size-selection of cDNA libraries for the cloning of cDNAs after suppression subtractive hybridization

V. Lévesque, T. Fayad, K. Ndiaye, M. Nahé Diouf, and J.G. Lussier
Université de Montréal, Montreal, QC, Canada
BioTechniques 35:72-78 (July 2003)

Here we describe the establishment of size-selected cDNA libraries for the cloning of full-length cDNAs that were initially identified by suppression subtractive hybridization (SSH) technology as being differentially expressed. First, the SSH-cDNA fragments were used as 32P-probes to verify their level and differential pattern of expression by virtual Northern and to establish their corresponding full-length cDNA size. Second, cDNAs were separated by size on agarose gels and used to construct size-selected cDNA plasmid libraries, which were then screened by colony hybridization with the SSH-cDNA fragments. We conclude that the described approach complements SSH technology by allowing efficient cloning and characterization of the corresponding full-length cDNA from any desired cell type or species. This approach will give researchers the ability to specifically target and study differentially expressed genes in an efficient manner for functional genomic studies.