

2002

# In-gel purified ditags direct synthesis of highly efficient SAGE Libraries

Saroj P. Mathupala

*Barbara Ann Karmanos Cancer Institute, Wayne State University School of Medicine, ah1244@wayne.edu*

Andrew E. Sloan

*Barbara Ann Karmanos Cancer Institute, Wayne State University School of Medicine, asloan@neurosurgery.wayne.edu*

---

## Recommended Citation

Mathupala and Sloan *BMC Genomics* 2002, 3:20

doi:10.1186/1471-2164-3-20

Available at: <http://digitalcommons.wayne.edu/biomedcentral/228>

Methodology article

## "In-gel" purified ditags direct synthesis of highly efficient SAGE Libraries

Saroj P Mathupala<sup>1</sup> and Andrew E Sloan\*<sup>2</sup>

Address: <sup>1</sup>Department of Neurological Surgery, & Barbara Ann Karmanos Cancer Institute, Wayne State University School of Medicine, 4160 John R. Road, Suite 930, Detroit, Michigan 48201, USA and <sup>2</sup>Department of Neurological Surgery, & Barbara Ann Karmanos Cancer Institute, Wayne State University School of Medicine, 4160 John R. Road, Suite 930 Detroit, Michigan 48201, USA

E-mail: Saroj P Mathupala - ah1244@wayne.edu; Andrew E Sloan\* - asloan@neurosurgery.wayne.edu

\*Corresponding author

Published: 1 August 2002

Received: 29 May 2002

BMC Genomics 2002, 3:20

Accepted: 1 August 2002

This article is available from: <http://www.biomedcentral.com/1471-2164/3/20>

© 2002 Mathupala and Sloan; licensee BioMed Central Ltd. This article is published in Open Access: verbatim copying and redistribution of this article are permitted in all media for any non-commercial purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** SAGE (serial analysis of gene expression) is a recently developed technique for systematic analysis of eukaryotic transcriptomes. The most critical step in the SAGE method is large scale amplification of ditags which are then concatenated for the construction of representative SAGE libraries. Here, we report a protocol for purifying these ditags via an *in situ* PAGE purification method. This generates ditags free of linker contaminations, making library construction simpler and more efficient.

**Results:** Ditags used to generate SAGE libraries were demarcated *in situ* on preparative polyacrylamide gels using XC and BPB dyes, which precisely straddle the ditag band when a 16% PAGE gel (19:1 acrylamide:bis, 5% cross linker) is used to resolve the DNA bands. Here, the ditag DNA was directly excised from gel without visualization via EtBr or fluorescent dye staining, resulting in highly purified ditag DNA free of contaminating linkers. These ditags could be rapidly self ligated even at 4°C to generate concatemers in a controlled manner, which in turn enabled us to generate highly efficient SAGE libraries. This reduced the labor and time necessary, as well as the cost.

**Conclusions:** This approach greatly simplified the ditag purification procedure for constructing SAGE libraries. Since the traditional post-run staining with EtBr or fluorescent dyes routinely results in cross contamination of a DNA band of interest by other DNA in the gel, the dry gel DNA excision method described here may also be amenable to other molecular biology techniques in which DNA purity is critically important.

### Background

SAGE (serial analysis of gene expression) technique [1], which has replaced differential display [2] as the primary technique for comparative analysis of transcriptome profiles between two or more tissues, relies on the quantitation of 3' termini of transcripts. Here, short (10–14 BP)

dsDNA sequences known as tags adjacent to the 3' most NlaIII site of cDNA generated from a transcriptome are isolated, and then dimerized to form ditags. The ditags are then PCR amplified, pooled, and then concatenated together for identification using a minimum amount of sequencing runs, where a 600–800 bp sized concatemer will

contain approximately 50 tags that represent the same number of transcripts.

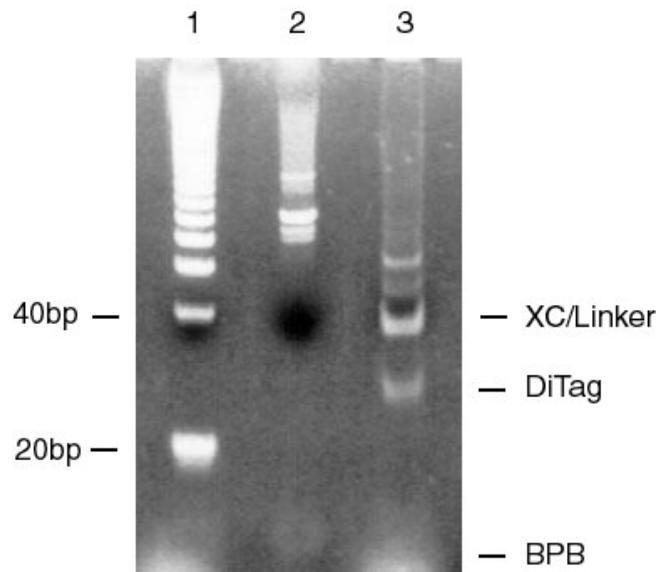
For practical manipulations, each concatemer is cloned into a suitable vector, and thousands of such clones (SAGE library) are sequenced to get a profile of the expression pattern of the transcriptome of a given tissue of interest.

Although a SAGE library construction project may take between 4 to 5 weeks, the most crucial part is the final stage, in which the amplified ditags are concatemerized to obtain self-ligated products in the size range of 600–800 bp. Prior to the concatemerization step, linker DNA, which are used to PCR amplify the ditags, need to be effectively removed as they tend to "poison" the concatemerization reaction by capping the ditags. The standard procedure for removal of linker DNA involves the prior labeling of linkers with biotin, which in turn enables them to be removed later using streptavidin affinity capture. For this, the commonly followed SAGE protocol [3] uses streptavidin coupled magnetic beads to capture and retain the linker against a magnetic field while the ditags are separated out. This method is not only costly, due to the large amounts of magnetic beads needed, but also time consuming, as multiple wash and recovery cycles are required [3].

Here, we report a simple procedure involving classical molecular biology using PAGE to separate the ditags from contaminating DNA, with co-migratory visual tracking dyes used to localize the ditags *in situ*. The visual dyes BPB and XC described here have been routinely used in laboratories to visualize DNA and RNA migration rates during both polyacrylamide and agarose gel electrophoresis [4]. However, they have not been used as described herein, for *in situ* visualization, localization and excision of DNA from polyacrylamide or agarose gels, as a method for purification of specific DNA bands of interest. In the method described here, the PAGE gel/cross-linker concentrations have been optimized/adjusted such that the migrating ditag band is demarcated by the dyes BPB and XC, both of which straddle the ditag DNA band. This allows for the direct excision of the ditag band from the gel following electrophoresis. Since the gel is not stained to visualize DNA prior to isolation of the ditag DNA band during this procedure, cross-contamination with linker DNA (which migrates behind the ditag) or other misprimed PCR products is prevented, resulting in recovery of extremely pure ditag DNA that can be efficiently concatemerized for SAGE library construction.

## Results and Discussion

A typical electrophoretic pattern obtained under the conditions listed in Materials and Methods section are shown in Figure 1. The electrophoresis is terminated when the

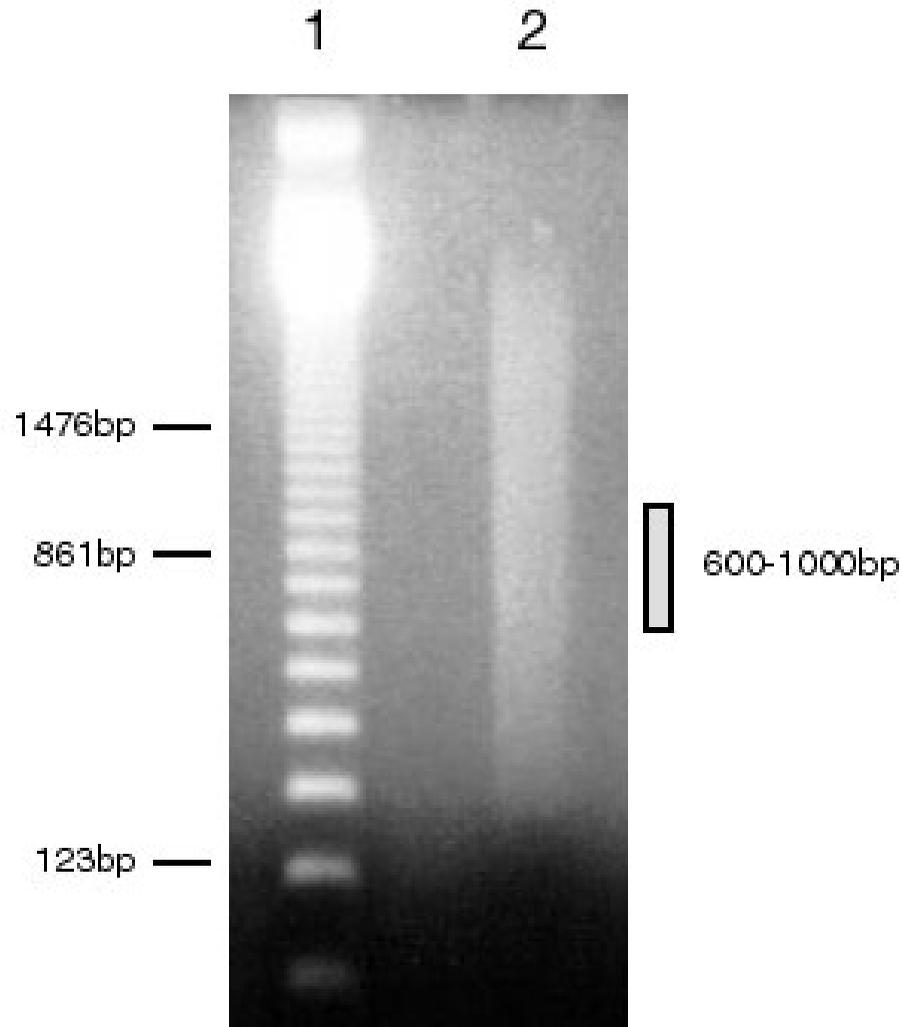


**Figure 1**  
Native PAGE analysis of NlaIII digested 102 bp PCR product. Lane 1, 20 bp markers; Lane 2, undigested product; Lane 3, products after Nla III digestion. Positions of 26 bp ditags, 40 bp residual linker, XC and BPB dye fronts are indicated.

BPB dye front reaches the front edge of gel. At this stage, the 26 bp ditag DNA is well separated from both the BPB and XC dye bands. As it can be observed, the XC dye band migrates precisely at 40 bp, which coincides with the molecular weight of the linker DNA that needs to be completely separated from the 26 bp ditag DNA to prevent 'poisoning' of the concatemers during SAGE library construction.

The use of XC and BPB dyes allow the precise visualization of the outlier border for the 26 bp ditag DNA. Using the two dye tracks as guides, the gel fragment between them can be excised from the post-run gel prior to any staining for DNA visualization. If necessary, the excised gel fragment can then be briefly stained with EtBr to precisely demarcate the DNA band, in order to trim away residual gel material to enhance the concentration of the recovered DNA. Since the ditag DNA has been separated from contaminating 'linker' in the rest of the gel, any post-run staining and manipulation will not introduce extraneous DNA bands back into the 26 bp ditag of interest.

It is quite possible to expand the method to demarcate and isolate DNA (or, for that matter protein) bands of interest using different ranges of gel concentrations (4%–20%) where a range of tracking dyes can be used to demarcate a given DNA band of interest. For example, in addition to XC and BPB, visual dyes such as Orange G and



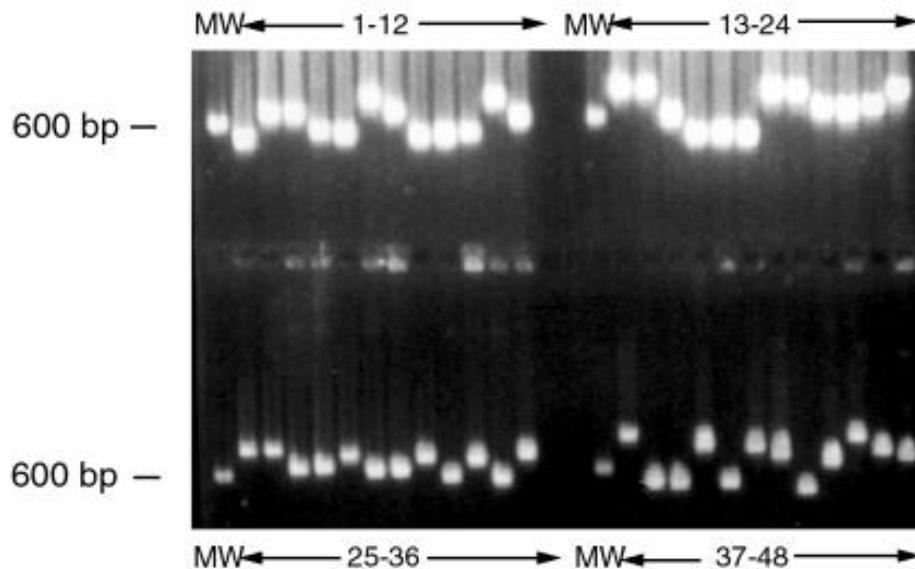
**Figure 2**  
Agarose gel analysis of concatemers. Lane 1, 123 bp DNA ladder; Lane 2, Concatemers (1  $\mu$ l)

Bromocresol Green can be used to distinguish and separate DNA of different molecular weight ranges.

Proof of principle of the above purification method can be seen upon ligation of isolated ditags to form concatamers, as shown in Figure 2. Here, the ditags were ligated at 4°C over a period of 90 min. The low temperature allowed us to precisely control the size range of the concatemers that were formed, and the reaction could be stopped when the desired size range was achieved (centered around 800 bp; 600–1000 bp). Since the resultant concatemers were in the desired size range for SAGE library construction, subsequent ligations were highly efficient, maximizing the number of productive clones that could be generated per ng of ditags. This is in marked contrast to the standard SAGE concatmer formation reaction process, where >90% of the concatemers resulted in the size of

>5000 bp which is wasted as it is unclonable or results in clones that cannot be subjected to high throughput sequencing to generate SAGE library data [1].

We have used the above method to construct three SAGE libraries from human brain specimens in which we were able to generate approximately 250 clones per 0.5  $\mu$ l of the standard SAGE concatemer-plasmid vector ligation reaction (10  $\mu$ l), where the SAGE library was constructed using a starting amount of 500 ng mRNA. We routinely obtain 600–800 bp concatamers in 95% of the clones that are tested by PCR amplification as shown in Figure 3. Sequencing of the PCR products yielded approximately 25 ditags per clone on average. In contrast, use of the standard SAGE procedure [3] with 2.5  $\mu$ g of brain tumor derived mRNA yielded 100 clones per 0.5  $\mu$ l of the ligation reaction, with 17 ditags per clone on average, resulting in a



**Figure 3**

Size distribution of inserts in pUC 18 plasmid clones. MW, 600 bp molecular weight marker. Other lanes (1–48), individual clones after 25 cycle PCR amplification of inserts with M13 forward and reverse primers. Most clones are  $\geq 600$  bp

SAGE library of approximately 68,000 tags (Data not shown).

Therefore, the modifications allow us to obtain an approximately 250,000 tag SAGE library from 50 mg clinical tissue (50  $\mu$ g total RNA; 500 ng mRNA), using 25–26 cycle PCR at the large scale ditag amplification stage [3]. Therefore, an 18-fold greater efficiency in SAGE library yield was achieved using the techniques outlined above.

### Conclusions

In summary, we have used two simple and straightforward biochemical methods, i.e., *in situ* gel purification of ditags, and controlled ditag concatemerization by ligation at 4°C, to significantly improve the efficiency of the standard SAGE method. The former enabled us to get highly purified ditags for downstream processing, while the latter enabled us to synthesize concatemers essentially in the size range needed to construct SAGE libraries. This latter step significantly minimized the formation of very high molecular weight concatemers that are unusable when constructing a SAGE library. In addition, gel purification of ditags resulted in significant time and cost recoveries, as streptavidin-magnetic beads are not needed in linker-ditag separation.

### Methods

Total RNA was isolated from 50 mg clinical specimens using RNA-Stat (Tel-Test, Inc.). Messenger RNA was purified using Oligo-dT spin columns (Oligotex, Qiagen). Superscript Choice System (Invitrogen Cat. No. 18090) was used for cDNA synthesis in the presence of biotinylated oligo dT (Roche Biochemicals). All linker oligonucleotides were synthesized according to the standard SAGE protocol [3]. The SAGE library construction steps up to the Nla III digestion of linker-ditag molecules were carried out according to the established SAGE procedure with minor modifications, including the use of non-biotinylated primers for scale-up PCR of ditags, as the method outlined here obviates the need for biotinylation.

To isolate pure ditags for concatemerization, we followed the procedure outlined below; The Nla III digested linker-ditag molecules were phenol-chloroform extracted and ethanol precipitated according to the standard SAGE procedure and resuspended in 30  $\mu$ l TE. 6  $\mu$ l  $\times$  6 loading dye (40% sucrose, 0.25% BPB and 0.25% XC) was added to above and loaded into 3 wells of a 0.75 mm thick 16% gel (29:1, acrylamide:bis; also can be prepared by mixing equal volumes of 19:1 and 37:1 acrylamide:bis) using a Bio-Rad minigel apparatus. Electrophoresis (60 V) was carried out at 4°C using pre-cooled (4°C)  $\times$  1TBE running buffer. When the BPB dye front reached the front edge of gel, the gel/glass plate sandwich was removed from the

apparatus, and the smaller glass plate separated out. While maintaining the gel on the larger glass plate, the region between the XC and BPB dye bands (Fig. 2, lane 3) was excised into cold TE, and briefly stained (30 sec) with EtBr (10 µg/ml). Residual gel material was excised from around the 26 bp band under low intensity UV, and the resulting gel fragment processed by the "crush-and-soak" method as described in the standard SAGE procedure. DNA was eluted using a single microcentrifuge tube, but at 4°C using 300 µl TE, and the filtrate (280 µl) ethanol precipitated (28 µl MgCl<sub>2</sub>, 28 µl 3 M NaOAc, pH 5.2, 2 µl 5 mg/ml linear acrylamide, 3 µl 20 mg/ml glycogen, and 900 µl cold ethanol) by centrifugation (16,000 g, 40 min) after incubation at -80°C for 1 hour. The DNA pellet was washed twice with 75% cold ethanol (2 × 200 µl), and re-suspended in 7 µl TE<sub>0.1</sub>.

To concatemerize of the ditags, 2 µl × 5 ligase buffer and 1 µl T4 DNA ligase (5 U/µl) were added to above and incubated at 4°C for 1.5 hrs. 1 µl aliquots were analyzed by agarose minigel electrophoresis (1%) every 30 min, until the concatemers reached a median size of 800 bp. The entire ligation mix was then separated on a 1% agarose minigel prestained with EtBr (4 V/cm; 50 V)(Fig. 2, lane 2). Gel region containing concatemers in the size range of 600 bp to 1100 bp were excised (~ 200 mg gel slice) and the DNA eluted in to 50 µl 10 mM Tris, pH 8.5, by spin-column chromatography (QiaQuick Gel Extraction kit, Qiagen). This was concentrated by ethanol precipitation (25 µl NH<sub>4</sub>OAc, 2 µl 20 mg/ml glycogen, 225 µl cold ethanol; -80°C, 1 hr) at 16,000 g for 40 min (4°C). The pellet was washed with 75% cold ethanol (200 µl), and re-suspended in 5 µl TE<sub>0.1</sub>. The isolated concatemers were ligated into 0.5 µg of Sph I digested, dephosphorylated pUC 18 vector in a 10 µl ligation reaction according to standard procedures to generate the SAGE library.

A detailed protocol can be obtained from the corresponding author (A.E.S.) via email [asloan@neurosurgery.wayne.edu].

### Authors' contributions

S.P.M. carried out the SAGE methodological studies. A.E.S. conceived of the study, provided the materials, and participated in its design and coordination. Both authors participated in the preparation and approval of the final manuscript.

### List of abbreviations

BPB, bromophenol blue; XC, xylene cyanol; TE<sub>0.1</sub>, 10 mM Tris-HCl, 0.1 mM EDTA; TE, 10 mM Tris, 1 mM EDTA, pH 8.0; EtBr, ethidium bromide; PAGE, polyacrylamide gel electrophoresis; TBE, Tris-borate-EDTA buffer.

### Acknowledgements

This work was supported by the George Kamberos Award of the American Brain Tumor Association to A.E.S. The authors acknowledge the able assistance of Prahlad Parajuli, Ph.D., Pingyang Yu, M.D. and Ruping Wu, as well as helpful discussions with John Kamholtz, M.D., Ph.D.

### References

1. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression**. *Science* 1995, **270**:484-487
2. Liang P, Pardee AB: **Differential Display of Eukaryotic messenger RNA by means of the polymerase chain reaction**. *Science* 1992, **257**:967-971
3. Velculescu VE, Zhang L, Zhou W, Traverso G, St Croix B, Vogelstein B, Kinzler KW: **Serial Analysis of Gene Expression Detailed Protocol Ver. 1.0d.**, 1999
4. Sambrook J, Fritsch EF, Maniatis T: *in Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, New York, 1989

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMedcentral will be the most significant development for disseminating the results of biomedical research in our lifetime."

Paul Nurse, Director-General, Imperial Cancer Research Fund

Publish with **BMC** and your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours - you keep the copyright

Submit your manuscript here:

<http://www.biomedcentral.com/manuscript/>

 **BioMedcentral.com**

[editorial@biomedcentral.com](mailto:editorial@biomedcentral.com)