


See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/10877599>

Shotgun Library Construction in a Day

Article in BioTechniques · March 2003
DOI: 10.2144/03342bm02 · Source: PubMed

CITATIONS
0


3 authors:



Irene Rombel

14 PUBLICATIONS 1,002 CITATIONS

SEE PROFILE




Stephen albert Johnston

Arizona State University

284 PUBLICATIONS 16,339 CITATIONS

SEE PROFILE

READS
76



Adel Talaat

University of Wisconsin–Madison

116 PUBLICATIONS 3,690 CITATIONS

SEE PROFILE

- ing and immunofluorescence staining of archival human liver transplant biopsies. *J. Histochem. Cytochem.* 49:1321-1322.
2. **Johnson, K.L., J.L. Nelson, D.E. Furst, P.A. Mc Sweeney, D.J. Roberts, D.K. Zhen, and D.W. Bianchi.** 2001. Fetal cell microchimerism in tissue from multiple sites in women with systemic sclerosis. *Arthritis Rheum.* 44:1848-1854.
3. **Johnson, K.L., D.K. Zhen, and D.W. Bianchi.** 2000. The use of fluorescence *in situ* hybridization (FISH) on paraffin-embedded tissue sections for the study of microchimerism. *BioTechniques* 29:1220-1224.
4. **Korbling, M., R.L. Katz, A. Khanna, A.C. Ruifrok, G. Rondon, M. Albitar, R.E. Champlin, and Z. Estrov.** 2002. Hepatocytes and epithelial cells of donor origin in recipients of peripheral-blood stem cells. *N. Engl. J. Med.* 346:738-746.
5. **Sigurdsson, S., S.K. Bodvarsdottir, K. Ananthawat-Jonsson, M. Steinarsdottir, J.G. Jonasson, H.M. Ogmundsdottir, and J.E. Eyfjord.** 2000. p53 abnormality and chromosomal instability in the same breast tumor cells. *Cancer Genet. Cytogenet.* 121:150-155.

Dr. Khosrotehrani was supported by the "Bourse Lavoisier du ministère des affaires étrangères", France. We thank Pr Selim Aractingi and Caroline Ledanff for helpful discussions on immunofluorescence techniques. Address correspondence to Dr. Kirby L. Johnson, Division of Genetics, Department of Pediatrics, Tufts-New England Medical Center, Box 394, 750 Washington Street, Boston, MA 02111, USA. e-mail: kjohnson@lifespan.org

Received 3 October 2002; accepted 11 December 2002.

**Kiarash Khosrotehrani,
Helene Stroh, Diana
W. Bianchi, and Kirby
L. Johnson**
*Tufts-New England Medical
Center
Boston, MA, USA*

**For reprints of this or
any other article, contact
Reprints@BioTechniques.com**

Shotgun Library Construction in a Day

BioTechniques 34:244-250 (February 2003)

The generation of random genomic libraries is an important component of several key technologies, including shotgun sequencing of genomic DNA (2) and genomic SELEX for isolation of protein binding ligands (4). We are particularly interested in the generation of random libraries for expression library immunization (ELI), a technique that screens genomic libraries for DNA fragments that protect animals against infection from the pathogen in question, thus providing leads for vaccine candidates (1). Traditionally, random genomic libraries have been constructed using DNA that has been fragmented either by physical shearing (such as nebulization or sonication) or by partial digestion with a restriction enzyme. However, there are a number of practical disadvantages associated with these methods: (i) they are time consuming, and (ii) they require large amounts of DNA (e.g., both ELI and SELEX require 25 µg genomic DNA as the starting material). This high demand is driven by DNA loss during size fractionation and the relatively inefficient processes of blunt-end ligation or of using adaptors for cloning (6). A protocol that uses amounts of genomic DNA in the nanogram range has been described (3), but this method still has the disadvantage of being fairly labor intensive and lengthy.

We have developed a rapid and facile protocol for constructing genomic libraries from nanogram amounts of DNA. The first part of the technique for producing insert DNA uses a similar principle to one previously described (4). The DNA is initially amplified using oligonucleotide primers that have two distinct parts: a sequence of random nucleotides that is used to prime the genomic DNA and a fixed sequence of nucleotides that is common to every oligonucleotide in the mixture [5'-CC-UCGUCGUCGUCUG(N)₉-3', where underlining denotes the constant region]. The average product length is influenced by the length of the random region of the template, as well as by the

annealing temperature. We empirically determined that a length of nine random nucleotides and an annealing temperature of 25°C resulted in the production of DNA fragments within the 300–800 bp size range, which was optimal for constructing libraries for ELI. To amplify the DNA, a 20-µL reaction was set up with 10–100 ng genomic DNA that had been purified by standard methods, 2 µL Klenow 10× buffer, 0.3 mM dNTPs (Promega, Madison, WI, USA), and 50 ng random primers (synthesized on an ABI 394 DNA/RNA Synthesizer; Applied Biosystems, Foster City, CA, USA). To generate the first strand of DNA, the mixture was first heated at 93°C for 3 min, chilled on ice for 2 min, and treated with 5 U Klenow enzyme (Promega). After a 5-min pre-incubation stage on ice, the tube was transferred to a thermal cycler, where it was subjected to incubations of 25 min at 25°C and 5 min at 50°C. The reaction was then heated at 93°C for 2 min and chilled on ice for 2 min. To synthesize the second strand of DNA, an additional 5 U Klenow were added to the reaction, which was then subjected to the same series of incubations as described above. To remove the random primers and unincorporated nucleotides, the reaction was passed over a gel filtration column (Edge Biosystems, Gaithersburg, MD, USA) and centrifuged for 2 min at approximately 360× g to produce an eluate that contained the purified dsDNA product. Interestingly, we did not observe a visible difference in the first Klenow step whether (i) the DNA was nebulized or untreated, or (ii) we used 10 or 100 ng starting material (not shown). We did not test below the 10-ng range of DNA.

To produce sufficient amounts of DNA for cloning purposes, this product provided the template in a subsequent PCR step that used a primer corresponding to the constant region of the random primer mixture used for the initial amplification step (5'-CC-UCGUCGUCGUCUG-3'), since this complemented one of the two template strands. In addition, this primer contained five UTP residues required for subsequent linkage of the PCR products (see below). PCR was carried out in a 100-µL volume containing 2 µL purified dsDNA product generated by

Benchmarks

the Klenow enzyme, 10 μ L *Taq* 10 \times buffer (Promega), 2 mM $MgCl_2$, 0.5 M betaine, 0.2 mM dNTPs, and 1 μ M PCR primer. The reaction was overlaid with mineral oil and heated for 3 min at 94°C in a thermal cycler, at which time 5 U *Taq* DNA polymerase (Promega) were added. The reaction was then subjected to 10 cycles at 94°C for 1 min, 50°C for 1 min, and 72°C for 1 min, and 20 cycles at 94°C for 1 min, 45°C for 1 min, and 72°C for 1 min. The PCR products were purified through QIAquick™ PCR purification columns (Qiagen, Valencia, CA, USA), using 50 μ L distilled water for elution, and elec-

trophoresed on a 1.5% gel. To ensure that the product generated was specific, control reactions that contained no template DNA or no primers were included. These controls appeared blank on the gel, while the complete reaction gave a short, intense smear centered at approximately 500 bp (not shown).

The described two-step process for making the DNA insert, which is usually the most laborious, time-consuming, and technically demanding part of library preparation is very rapid (<6 h) and involves relatively little hands-on work. To make the library in a rapid and highly efficient manner, we by-

passed ligase-based cloning and took advantage of the deoxyuracil residues present in the terminal 15-bp regions of each PCR product by linking them with a plasmid vector that contained compatible ends. The presence of deoxyuracil residues allow the generation of 15-base single-stranded regions by treatment with uracil N-glycosylase (UDG) (5). These cohesive ends are then capable of annealing to the compatible ends of the vector. To prepare the plasmid vector, we linearized our vector of choice (pCMVi-UB; Reference 6) with *Bgl*II, dephosphorylated, gel-purified, and ligated with an oligonucleotide that contained a *Bgl*II overhang in addition to the 15-mer constant region present at the termini of all of the PCR products (5'-P-GATCC-CTCGCTCGTCTG-3', where P is a phosphate group and the constant region is underlined). The ligation was carried out overnight at 16°C in the presence of 2.5% PEG 8000, and the ligated vector was gel-purified to remove PEG, excess oligonucleotide, and unligated vector molecules. The genomic library was prepared in 10 μ L reaction containing 50 ng vector DNA and different concentrations of the PCR product insert, ranging from 1 to 6 μ L of the 100- μ L product, in addition to 1 μ L Promega Restriction Buffer C and 1 U UDG enzyme (New England Biolabs, Beverly, MA, USA). The reaction was incubated for 30 min at 37°C, after which time samples of pre- and post-UDG-treated DNA were electrophoresed on a 1% gel to determine the efficiency of the linkage step. Finally, a 1- μ L aliquot was transformed into 40 μ L electrocompetent *E. coli* cells, at which point the library construction was completed. This linkage part of the procedure can be achieved in about 2 h, provided that the vector has been prepared ahead of time. Based on the number of colonies observed, we estimate that 1 μ g untreated DNA would yield approximately 5×10^6 cfu, which corresponds to 100–400 genome equivalents for a typical bacterium. Figure 1 shows the entire shotgun library construction process.

To determine the size distributions of the cloned DNA fragments and to ascertain whether there was any overt bias for certain sequences, we carried

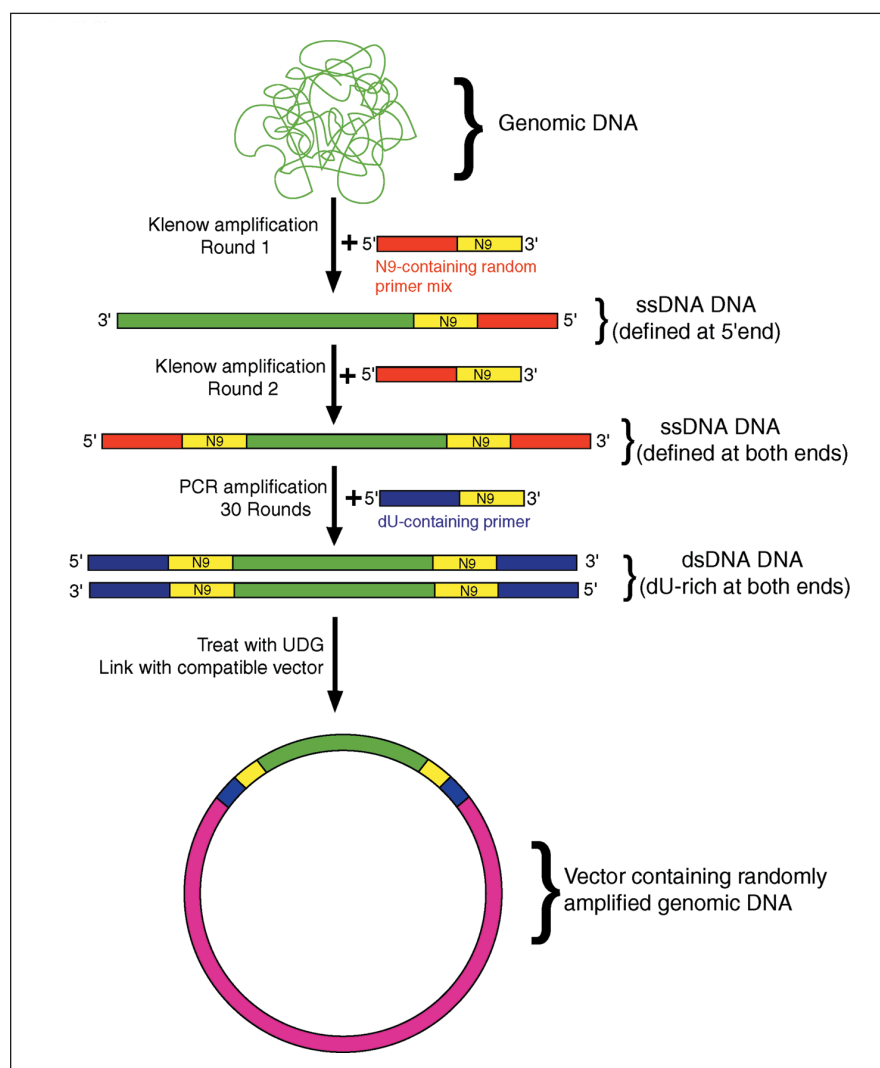
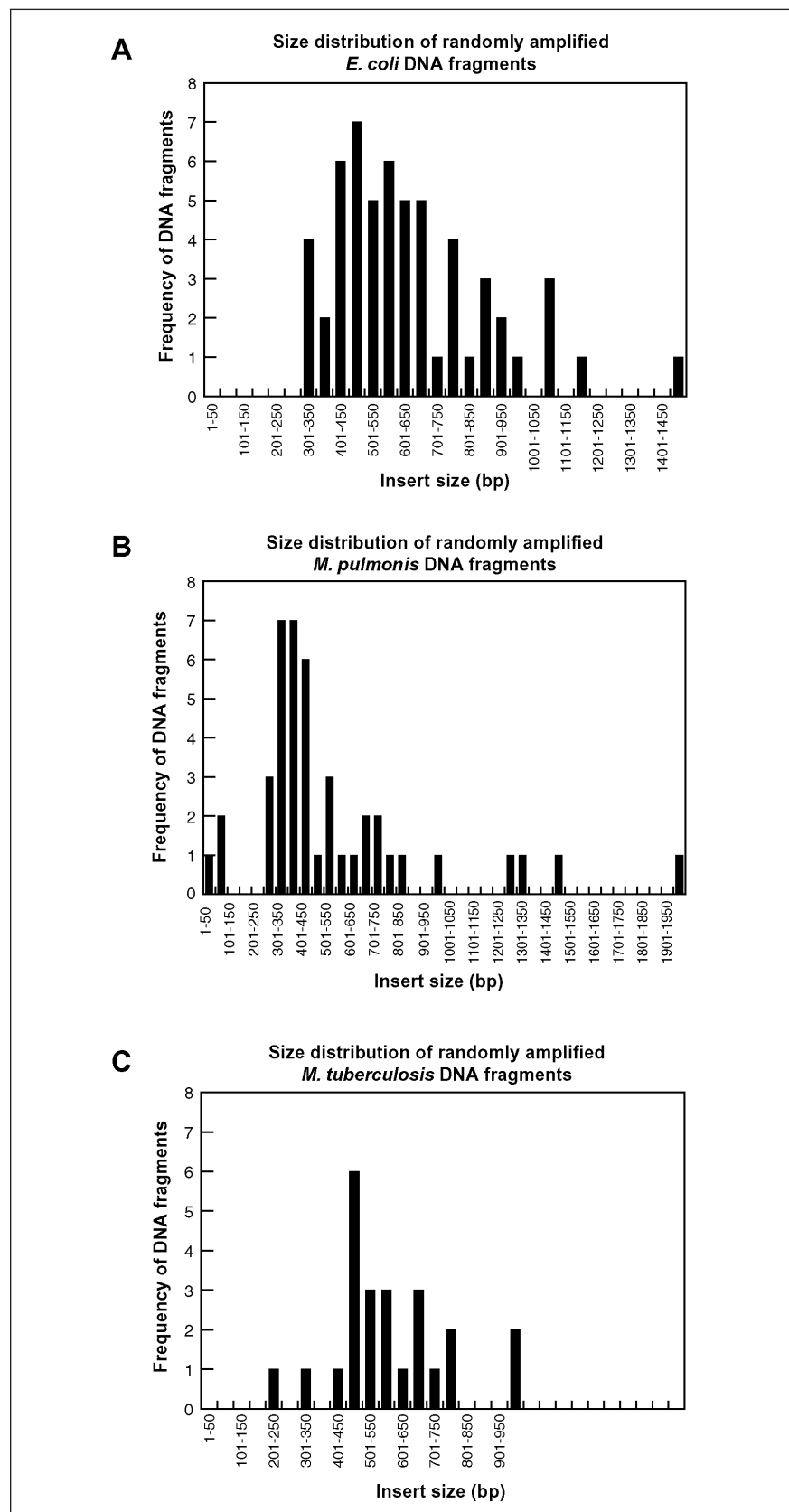


Figure 1. Procedure for shotgun library construction. Genomic DNA (green) is amplified with Klenow DNA polymerase using a primer that contains nine random nucleotides (yellow) linked to a fixed 16-mer sequence (red). Two rounds of amplification are required to produce ssDNA molecules that are defined by the fixed 16-mer sequence at each end. These molecules are then PCR-amplified using a primer that corresponds to the fixed sequence and that contains several dU residues (blue). The resultant PCR product is then linked to a vector with compatible ends in a reaction that involves UDG.

Benchmarks

out colony PCR sequencing. Furthermore, to determine whether genomes with different G+C compositions would affect the outcome, we prepared and analyzed genomic libraries from three bacteria: *E. coli*, *Mycoplasma pulmonis*, and *Mycobacterium tuberculosis*, which have respective G+C contents of 51%, 28%, and 65%. Analysis of 57, 43, and 24 clones from *E. coli*, *M. pulmonis*, and *M. tuberculosis*, respectively, revealed average insert sizes of 640, 540, and 580 bp, as was desired (Figure 2). Despite a 37% difference in G+C content between the two most extreme examples tested (*M. pulmonis* and *M. tuberculosis*), the observed difference in average PCR product size between these two organisms was only 100 bp, less than 20% in variation. BLAST analysis of the cloned DNA revealed that all of the sequences were unique, indicating that there was no overt bias during amplification.

To further assess the genome coverage of a typical library, we hybridized the *M. tuberculosis* library to DNA microarrays designed for the *M. tuberculosis* genome (7,8). The fluorescent dye-labeled nucleotide (Cy3-dCTP; Amersham Biosciences, Piscataway, NJ, USA) was incorporated during the PCR amplification stage (using 2.5 μ M Cy3-labeled dCTP+7.5 μ M dCTP and 10 μ M dGTP, dTTP, and dATP) and used to probe a spotted microarrays of PCR products representing 95% of the *M. tuberculosis* genome. Mycobacterial genomic DNA (labeled with Cy5-dye using a standard nick translation protocol) was co-hybridized with the Cy3-labeled DNA library to assess the hybridization efficiency for each spot. The Cy3-labeled library gave hybridization signals above the background levels in greater than or equal to 81% of the genes, indicating that the insert DNA prepared by the method described above gives fairly good coverage of the genome in question. When the signals from the PCR-labeled library were compared to those of the labeled genomic DNA, 88% of the genes were found to have a ratio ranging from 1 to 2, as shown in the histogram in Figure 3. A relatively small number of genes (12%) had ratios greater than 2, indicating that these genes had multiple representations in the library, as might



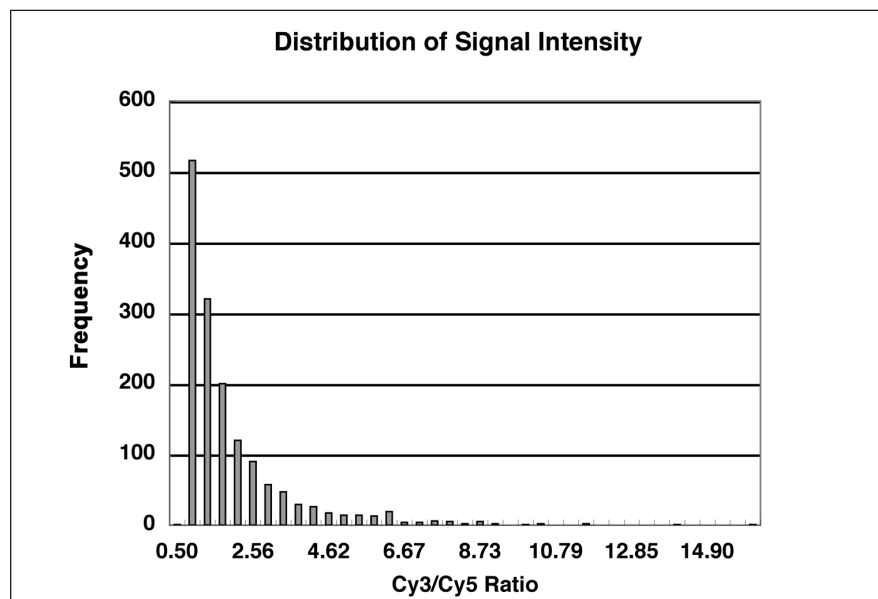


Figure 3. Distribution of signal intensities from the *M. tuberculosis* genomic library. The histogram represents the frequency of signals generated from the hybridized PCR library (Cy3-labeled) relative to the signals generated from the hybridized genomic DNA (Cy5-labeled).

be expected from an amplification-based strategy. To our knowledge, this is the first time that such analysis has been conducted to assess the representation of a genomic library.

In summary, we have developed a quick, facile protocol for constructing libraries from relatively small amounts of DNA. Using our method, we have shown that a representative, random library can be constructed in a single day. It has previously been shown that it is possible to vaccinate an animal using an entire genomic expression library and elicit a protective response (1). Therefore, it may be possible to create a genetic vaccine for even an unknown pathogen in a single day using this method. Given that in certain situations, such as biothreats or outbreaks of zoonoses, there may not be time to identify and isolate protective antigens from a complete genome, immunization with a shotgun random library may be a viable course of action.

REFERENCES

1. Barry, M.A., W. Lai, and S.A. Johnston. 1995. Protection against mycoplasma infection using expression library immunization. *Nature* 377:632-635.
2. Fraser, C.M. and R.D. Fleishmann. 1997. Strategies for whole microbial genome se-

quencing and analysis. *Electrophoresis* 18:1207-1216.

3. Rohwer, F., V. Seguritan, D.H. Choi, A.M. Segall, and F. Azam. 2001. Production of shotgun libraries using random amplification. *BioTechniques* 31:108-115.
4. Singer, B.S., T. Shtatland, D. Brown, and L. Gold. 1997. Libraries for genomic SELEX. *Nucleic Acids Res.* 25:781-786.
5. Smith, C., P.J.R. Day, and M.R. Walker. 1993. Generation of cohesive ends on PCR products by UDG-mediated excision of dU, and application for cloning into restriction digest-linearized vectors. *PCR Methods Appl.* 2:328-332.
6. Sykes, K.F. and S.A. Johnston. 1999. Genetic live vaccines mimic the antigenicity but not pathogenicity of live viruses. *DNA Cell Biol.* 18:521-531.
7. Talaat, A.M., S. Howard, W. Hale, R. Lyons, H. Garner, and S.A. Johnston. 2002. Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis* during growth from logarithmic to stationary phases. *Nucleic Acids Res.* 30:e104.
8. Talaat, A.M., P. Hunter, and S.A. Johnston. 2000. Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. *Nat. Biotechnol.* 18:679-682.

We are grateful to Ross Chambers for critical reading of the manuscript. This work was supported by DARPA grant no. MDA972-97-1-0013 to S.A.J. Address correspondence to Dr. Irene Rombel, Center for Biomedical Inventions, UT Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75390-9185, USA. e-mail: irene.rombel@utsouthwestern.edu

Received 10 July 2002; accepted 6 November 2002.

Irene T. Rombel, Adel M. Talaat, and Stephen Albert Johnston
*University of Texas
Southwestern Medical Center
Dallas, TX, USA*

Normalization of a cDNA Library Cloned in λ ZAP by a Long PCR and cDNA Reassociation Procedure

BioTechniques 34:250-254 (February 2003)

Some mRNAs are considerably more abundant than others in any mRNA population (4). In sequencing expressed sequence tag (EST) libraries, the identification of relatively rare mRNAs is hindered by redundant identification of abundant mRNAs (1). To remove the abundant sequences or "normalize" libraries, the kinetics of DNA reassociation have been utilized (14). After strand separation, abundant sequences reassociate more rapidly than rarer sequences. Hence, at a suitable point of initial concentration multiplied by time (C_0t), the single-stranded fraction will be greatly enriched for less abundant sequences and can be separated from the double-stranded fraction by hydroxylapatite (HAP) chromatography and then cloned (14).

We encountered this situation during studies on scabies, a contagious disease that afflicts people living in overcrowded housing in the developing and the developed world (11) and livestock (6). Molecular studies on the causative organism *Sarcoptes scabiei* have been extremely limited because of the difficulty of obtaining mites. Recently, cDNA libraries from *S. scabiei* have been constructed and EST programs commenced (8,10,12). During sequencing