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

Benchmarks

of libraries of scabies cDNAs (8) cloned in the bacteriophage vector λ ZAP (13), the rate of acquisition of new sequences slowed markedly, and after 9216 runs, normalization was necessary. mRNA-based approaches were not options in the present case, as there was no suitable source of additional mRNA. While effective methods for normalization have been described (5), we were unable to find a detailed protocol exactly applicable to a λ ZAP library. We have developed a protocol based on long and accurate PCR (3) to amplify the cDNA sequences and to then normalize and re-clone them in this vector. As the λ ZAP system is suitable for EST programs, the procedure may be generally useful.

During preparation of the initial scabies cDNA libraries (8), three different size classes of cDNA were fractionated by chromatography on Sepharose® CL-2B and cloned separately so that selection against longer sequences could be partially compensated by sequencing more clones from the library with the largest inserts. The titers of the λ ZAP libraries Yv4, Yv5, and Yv6 were 2.7×10^9 , 1.5×10^{10} , and 4.0×10^9 pfu/mL (8).

Primers were 35 bases long, exactly overlapping the *Eco*RI and *Xho*I sites of the λ ZAP vector.

L. (*Eco*RI) 5'-CCTGCAGGTCGA-CACTAGTGGATCCAAAGAATTC-3'

R. (*Xho*I) 5'-GATGGGCCCCGCGG-CGCTCTAGAAAGTACTCTCGAG-3'

8–12 μ L of the λ ZAP libraries were used as template in 100- μ L long-PCR systems. PCR buffer and Tac/Pfu 15/1, as in Reference 3, was used, except that KlenTaq was replaced by AmpliTaq Gold® (Applied Biosystems, Foster City, CA, USA). The reactions were heated to 95°C immediately after adding the enzyme mixture, as *Pfu* DNA polymerase can degrade oligonucleotides at 0°C or room temperature. The PCR mixture was heated to 95°C for 10 min to activate the AmpliTaq Gold and then cycled at 95°C for 30 s, 60°C for 30 s, and 68°C for 5 min extension for 15 cycles. Seven 100- μ L systems were used to generate approximately 10 μ g cDNA of each library. Six of these were pooled, and the DNA was precipitated with 2.5 volumes of ethanol after adding 0.1 volumes of 4 M sodium acetate. Primers were not re-

moved, as they were needed as blockers during the subsequent reassociation. To enable monitoring of the elution off HAP, dCTP was partially replaced by α [³²P]dCTP in one of the seven PCRs by using a dNTP mixture of 2 mM dGTP, dTTP, dATP, and 0.2 mM dCTP. Labeled PCR products were pooled, unincorporated nucleotides were removed using a MicroSpin S-300 HR column (Amersham Biosciences, Piscataway, NJ, USA), and the DNA was precipitated as described above.

Each pool of cDNA (10–15 μ g) was dissolved in 18 μ L deionized formamide and the labeled cDNA in 12 μ L formamide. All the Yv4 and Yv6 cDNA, 12 μ L of the more concentrated Yv5 cDNA, and all the labeled cDNA were combined with 24 μ L oligo dT (5 μ g/ μ L) and 6 μ L each PCR primer (2 μ g/ μ L) and melted at 80°C for 3 min in a PCR machine with a heated lid. Twelve microliters of 10 $\times$ TE and 12 μ L 5 M NaCl were then added. The reaction mixtures were then incubated at 42°C for 16 h (C_0 t approximately 100) in three aliquots of 40 μ L using 0.2-mL PCR tubes and a PCR machine with heated lid.

Each of the three samples was then set on ice for 5 min and immediately added to 200 μ L 10 mM sodium phosphate at 60°C and to approximately 0.2 mL packed HAP (prewashed in 10 mM sodium phosphate, pH 6.8, at 60°C) in a 1.5-mL Eppendorf® centrifuge tube. The HAP-cDNA mixture was then subsequently washed twice with 10 mM sodium phosphate, twice with 20 mM sodium phosphate, three times with 140 mM sodium phosphate, twice with 150 mM sodium phosphate, twice with 160 mM

sodium phosphate, and once with 0.4 M sodium phosphate. Wash conditions were 750 μ L volume, 5 min with gentle agitation at 60°C (no vortex, no pipetting), and 10 s spin at 3000 rpm. All eluates were monitored by measuring the radioactivity. Below 0.5% of the total counts were eluted from the HAP with the first four 10–20 mM washes, 25%–27% (corresponding to the ssDNA) were eluted with the 140–160 mM washes, 63%–65% (corresponding to the dsDNA) were eluted by the 400 mM wash, and 10% remained bound to the column. The fraction that initially bound to HAP but was eluted with 140–160 mM sodium phosphate was collected, pooled, and purified using the QIAquick™ PCR Purification Kit (Qiagen, Valencia, CA, USA), following the manufacturers protocol, with the exception of using one-fifth of the recommended volume of QIAquick binding buffer.

The DNA was precipitated and washed with 70% ethanol and then copied into dsDNA in an *E. coli* DNA polymerase I system with 5-methyl-dCTP utilizing the *Xho*I primer described above, under the conditions described in the λ ZAP Express cDNA cloning protocol (Stratagene, La Jolla, CA, USA). The hemimethylated products of this procedure were processed identically to the double-stranded cDNA generated in the λ ZAP Express cDNA cloning protocol. Following Sephacryl fractionation, three libraries, Yv7 (longest fraction; 0.9×10^5 pfu), Yv8 (3×10^5 pfu), and Yv9 (1.5×10^5 pfu), were generated. After one round of amplification and mass excision of phagemids from Yv7–9, as per the

Figure 1. Frequency of occurrence of clones in the 20 most abundant contigs from libraries Yv4–6, after normalization. The 20 contigs containing the most clones encountered in 9216 runs on libraries Yv4–6 were searched for using BLASTn in 14 880 runs on libraries Yv7–9.

Benchmarks

λ ZAP manufacturer's instructions, 120 clones were examined, and 116 were found to contain inserts ranging from 0.5 to 2 kb.

Clones were sequenced from the T3 primer using BigDye™ terminator and an ABI PRISM® 3700 capillary DNA analyzer (Applied Biosystems). The program phred (7) was used for automatic base calling and trimming of vector and low-quality sequences (phred score <20). Contigs were assembled through *phrap* (9) using the default values, except for minscore (60) and min-match (28) to increase the stringency of the contig assembly.

Long PCR amplification was found to be optimal and efficient at an input of 5×10^7 to 1×10^8 pfu λ phage/100 μ L system. As the cDNA was to be size-fractionated again after normalization, and to aid in achieving a high enough C_0t in a timeframe in which the DNA would undergo minimal depurination and thermal hydrolysis, the amplified fractions were pooled. After incubation overnight to a C_0t of approximately 100 in three independent experiments, over 99.5% of the total 32 P-labeled cDNA bound to the HAP in 10 mM sodium phosphate, and 25%–27% of this behaved as ssDNA that could be washed off the HAP by 140–160 mM sodium phosphate. After HAP fractionation and purification as described, the single-stranded fraction could readily be re-cloned in the λ ZAP system.

The sequences of the most abundant 20 contigs of the pre-normalized sequences were checked for their presence in the 14 880 normalized clones sequenced to date after entering them into a local database, using BLAST (2). Before normalization, the most abundant 20 contigs accounted for 16.4% of 9216 clones. After normalization, the same 20 contigs accounted for 0.87% of 14 880 clones (Figure 1).

Bonaldo et al. (5) noted that cross-hybridizing diverged sequences seem to escape normalization in all of the procedures that they have investigated. This was considered advantageous because it minimizes the likelihood that a rare member of a gene family might be excluded from the final normalized library as a result of cross hybridization with a more prevalent but diverged sequence. This appears also to be true for the pro-

cedure described here, as 33 members of the group 3 allergen homologue multigene family were present in the first 14 880 of the normalized clones sequenced, and 21 of these were sequences that had not been encountered before (data not shown). Of the clones examined that had been represented in the pre-normalized libraries, all were identical to or showed no more SNPs than the sequences from the pre-normalized libraries (the libraries were from several hundred mites; data not shown). The use of a thermostable DNA polymerase with an editing function in this procedure should minimize errors due to base misincorporation (3) and minimize errors due to recombination of incomplete PCR products via annealing to related but non-identical templates. We have not observed any examples of recombination among sequences analyzed in detail to date. As the highly efficient commercially available λ ZAP Express system is suitable for EST projects, the normalization procedure presented here may be generally useful.

REFERENCES

1. Adams, M.B., M.B. Soares, A.R. Kerlavage, C. Fields, and J.C. Venter. 1993. Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. *Nat. Genet.* 4:373-380.
2. Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
3. Barnes, W.M. 1994. PCR amplification of up to 35-kb DNA with high fidelity and high yield from lambda bacteriophage templates. *Proc. Natl. Acad. Sci. USA* 91:2216-2220.
4. Bishop, J.O., J.G. Morton, M. Rosbash, and M. Richardson. 1974. Three abundance classes in HeLa cell messenger RNA. *Nature* 250:199-204.
5. Bonaldo, M., G. Lennon, and M.F. Soares. 1996. Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res.* 6:791-806.
6. Davis, D.P. and R.D. Moon. 1990. Dynamics of swine mange: a critical review of the literature. *J. Med. Entomol.* 27:727-737.
7. Ewing, B., L. Hillier, M. Wendl, and P. Green. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res.* 8:175-185.
8. Fischer, K., D.C. Holt, P. Harumal, B.J. Currie, S.F. Walton, and D.J. Kemp. Generation and characterization of cDNA clones from *Sarcoptes scabiei* var. *hominis* for an expressed sequence tag library; Identification of homologues of house dust mite allergens. *Am. J. Trop. Med. Hyg.* (In Press.)
9. Gordon, D., C. Abajian, and P. Green. 1998. Consed: a graphical tool for sequence finishing. *Genome Res.* 8:195-202.
10. Harumal, P., M. Morgan, S.F. Walton, D.C. Holt, J. Rode, L.G. Arlian, B.J. Currie, and D.J. Kemp. Identification of a homologue of a house dust mite allergen in a cDNA library from *Sarcoptes scabiei* var. *hominis* and evaluation of its vaccine potential in a rabbit/*S. scabiei* var. *canis* model. *Am. J. Trop. Med. Hyg.* (In Press.)
11. Kemp, D.J., S.F. Walton, P. Harumal, and B.J. Currie. 2002. The scourge of scabies. *Biologist* 15:1-6.
12. Mattsson, J.G., E.L. Ljunggren, and K. Bergstrom. 2001. Paramyosin from the parasitic mite *Sarcoptes scabiei*: cDNA cloning and heterologous expression. *Parasitol.* 122:555-562.
13. Short, J.M., J.M. Fernandez, J.A. Sorge, and W.D. Huse. 1988. Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. *Nucleic Acids Res.* 16:7583-7600.
14. Soares, M.B., M.F. Bonaldo, P. Jelene, L. Su, L. Lawton, and A. Efstratiadis. 1994. Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. USA* 91:9228-9232.

This work was supported by a Medical Genomics grant from the Australian NHMRC, by NHMRC grant no. 137205, and by an NHMRC Senior Principal Research Fellowship to D.J.K. The sequencing program was carried out at the Australian Genome Sequencing Facility established in 1997 by a Federal Government Major National Research Facility grant. Address correspondence to Dr. David J. Kemp, Queensland Institute of Medical Research, Post Office Royal Brisbane Hospital, Herston, Qld, 4029, Australia. e-mail: davek@qimr.edu.au

Received 13 September 2002; accepted 13 November 2002.

K. Fischer, D.C. Holt, P. Wilson¹, J. Davis¹, V. Hewitt¹, M. Johnson¹, A. McGrath¹, B.J. Currie², S.F. Walton², and D.J. Kemp
*Queensland Institute of Medical Research
Queensland
¹University of Queensland
Queensland
²Menzies School of Health Research
Northern Territory, Australia*