PCR-based immortalization and screening of hierarchical pools of cDNAs

Maurizio D'Esposito^{1,2}, Richard Mazzearella¹, Gina Pengue¹, Carmela Jones¹, Michele D'Urso² and David Schlessinger^{1,*}

¹Department of Molecular Microbiology and Center for Genetics in Medicine, Washington University Medical School, St Louis, MO 63110, USA and ²International Institute of Genetics and Biophysics, CNR, Naples 80125, Italy

Received May 28, 1994; Revised and Accepted August 23, 1994

ABSTRACT

Starting from sequences of at least 60 bp, PCR-based screening has been developed to recover cDNAs from libraries without the necessity for hybridization or extensive DNA extraction steps. The method maintains the indefinite availability of even scarce cDNA libraries and provides an estimate of the relative abundance of the mRNA species. Isolation of a cDNA clone can be done in less than a week. cDNAs were isolated that were cognate for fragments of expressed sequences and for an exon predicted from genomic sequence.

INTRODUCTION

Molecular biologists have continued to put considerable effort into the production of high quality, full-length cDNA libraries, and the recovery of full-length cDNAs remains a major step in efforts to understand the function of genes. Searches for particular species in cDNA collections now include a number of methods based on the use of genomic DNA, for example in recombination-based (1) and direct selection (2,3) methods. In an increasing number of cases, bits of putative or demonstrated expressed sequence from a gene are available (4–10); but short fragments are difficult to use in hybridization experiments against cDNA libraries and often give significant background.

An alternative approach with high specificity is potentially provided by PCR-based screening. Using PCR and a variety of pooling schemes, genomic clones of interest have been identified among stored arrays of yeast artificial chromosome (YAC; 11,12) or λ libraries (13). Here we have adapted a hierarchical, tree-based scheme for the PCR-based screening of cDNA libraries with oligomer pairs inferred from sequences as short as 60 bp.

MATERIALS AND METHODS

Libraries

Two cDNA libraries have been used: a teratocarcinoma NT2/D1 cell library (14) and a human 8 week embryo library (15), both made in the vector λ gt10. These libraries provide a variety of

expressed sequences, since they derive from relatively undifferentiated cell types, and have yielded a number of cDNA clones (16–20) with an average insert size greater than 2 kb.

Pooling scheme

N-dimensional screening with pools of arrayed clones could provide the advantage of built-in scoring redundancy (11); but arrays of cDNAs analogous to those used to make YAC pools would number 700,000 or more to include rare species. Assembling such arrays would be prohibitively difficult as a standard first step. Instead, pools for sib selection have been constructed from dilutions of a lysate. To make the 20 top-Level (Level 1) pools (21; Figure 1, described in the text), a plate stock of phages was eluted with 1 ml of SM overnight for each Level 1 subpool. Then 500 microliters of phage lysate from each of 12 subpools were combined to give a Level 1 pool. DNA was purified for PCR assays from the 20 top-level (Level 1) pools using cetyl-trimethylammonium bromide (22). This purification improves the sensitivity and reproducibility of screening of Level 1 pools, but is not necessary with less complex pools. 10 ng of the purified phage DNA is the substrate for PCR screening at Level 1. For a positive in a Level 1 pool, two microliters of a 10-fold dilution of phage lysate was screened from each Level 1 subpool. Similar dilutions of pools, subpools, and individual phage stocks are employed in the screening assays at Levels 2 and 3. The titers of the phage pools have thus far been stable for 15 months at 4°C, and replicate pools are stored at –20°C in 7% dimethylsulfoxide.

PCR conditions

The testing of the primers pairs was done with TNK100 Buffer (23) (10 mM Tris, pH 8.6; 100 mM KCl; 1.5 mM MgCl₂; 5 mM NH₄Cl), 100 μ M dNTPs, 0.02 units/ml *Taq* polymerase, 0.5 nmol/ml of each primer and 10 ng/ml DNA in a 15 μ l reaction. PCR reactions were done with an initial denaturing cycle of 2.5 min at 94°C. The subsequent 30 cycles were done with a denaturing time of 30 s at 94°C, an annealing time of 45 s at 55°C and extension time of 45 s at 72°C.

*To whom correspondence should be addressed

Figure 1. Pooling scheme for PCR screening of a cDNA library. The top level (Level 1) consists of 20 pools, with each pool being made from subpools stored in 2 six well plates containing a total of about 36,000 phage. Once the positive Level 1 subpool is identified, about 7500 phage are replated and stored in subpools in 12 pairs of six well plates for Level 2 screening. After the identification of a Level 2 subpool and replating, 130 phage clones are picked and arranged into 13 pools for Level 3 screening, which identifies the positive phage.

Genomic Southern blots

1 μ g of YAC DNA and 10 μ g of human genomic DNA was digested with *EcoRI* and the products were separated by electrophoresis in a 0.8% agarose gel and transferred onto an ONCOR membrane according to manufacturer's guidelines. DNA on the filter was hybridized to a gel-purified cDNA radiolabeled by random priming (24).

Northern blot

Tissue distribution of the cDNAs were assessed by hybridization of gel-purified cDNAs by blot hybridization to human (poly A⁺) RNAs from a series of adult tissues (Clontech). The final wash was done with 0.2 $\times$ SSC /0.2 % SDS at 65°C for the actin probe and at 50°C for the EST and conserved STS probes.

RESULTS

Design of screening protocol

In previous work, the regional localization of specific ESTs on the X chromosome was determined (25,26). We developed a PCR-based screening strategy for the isolation of cognate cDNAs.

The technique employs sib selection based on a hierarchical, top-down pooling scheme. The scheme is outlined in Figure 1 and the assays described in Methods. In brief, it identifies single phage from an initiative total of 720,000 phage equivalents, containing more than 3.5 equivalents for even a very rare cDNA

Table 1. Fraction of complex cDNA pools positive for X-specific ESTs by PCR assays

EST	Positive pools (of 20 tested)	
	Embryo	Teratocarcinoma
00133 (DXS1006E)	8	4
00169 (DXS1008E)	0	8
00205 (DXS1012E)	16	16
00276 (DXS1014E)	0	0
00354 (DXS1005E)	7	4
00442 (DXS1013E)	0	0
00574 (DXS1119E)	1	0
00737 (DXS1115E)	0	0
00827 (DXS1009E)	7	6
00858 (DXS1180E)	0	0
00887 (DXS1116E)	10	18
00896 (DXS1117E)	2	20
01014 (DXS1010E)	14	6
01029 (DXS1011E)	0	10
01560 (DXS1004E)	0	8
01879 (DXS1118E)	0	8
02087 (DXS1179E)	0	10

Each pool contains about 36,000 PFU.

(that is, 1 copy/200,000 phage). Thus, if representation of mRNA sequences were random, there is a 97% chance that rare mRNAs would be included.

Pools at three levels are constituted from dilutions of the original stock. Dilutions are plated on agar by standard methods

Figure 2. Northern analysis of EST00205 and EST01014. A multiple tissue mRNA blot was hybridized with probes derived from cDNAs for either EST00205 (panel A), EST01014 (panel B) or actin (panel C).

to obtain 240 phage lysates, each with a complexity of 3,000 phage. These are then recombined into 20 pools (Level 1 pools), each containing 36,000 phage equivalents with 10^{10} – 10^{11} phage particles/ml. Each top level pool (Figure 1) is reconstituted from 12 subpools of 3,000 phage each (Level 1 subpools) conveniently stored in 6-well microtiter dishes.

Level 2 pools are similarly constructed by plating 2.5 equivalents of a Level 1 subpool (about 7500 phage). In the infrequent case when a phage in a Level 1 subpool is missing in the derived Level 2 subpools (statistical expectation of 8%), it can be recovered by replating the Level 1 subpool.

Each of 12 Level 2 pools is again reconstituted from 12 Level 2 subpools, with each subpool containing 50 phage equivalents. Once a well positive in a Level 2 subpool is identified, phage are replated to test single phage (Level 3). 130 single phage, or again about 2.5 equivalents of DNA, are arrayed in 10 rows of 13. Row and column pools are then screened to identify positive clones.

Isolation of a cDNA clones encoding EST01014 and EST00205

Each of 17 PCR primer pairs for expressed sequence tags (ESTs) was scored on the human teratocarcinoma and embryo libraries (Table 1). The relative abundance of each EST varies in each of the cDNA libraries. Two of the more abundant cDNAs (1/30,000–1/72,000) were EST01014 and EST00205. Both are located in Xq28 in a region which contains many disease genes (27).

Figure 3. Recovery of a cDNA for an STS derived from a YAC clone. Panel A. PCR amplification of an STS from human X and hamster DNA. An STS derived from the end of YAC yWXD1689, was tested with 10 ng total human DNA (lane 1), a human X-only hamster hybrid (lane 2), a hamster hybrid containing only Xq24-q28 (lane 3), a human 7-only containing hamster hybrid (lane 4), hamster DNA (lane 5), yeast DNA (lane 6) or no DNA (lane 7). Panel B. Southern analysis of the cDNA recovered with STS sWXD1080. DNA from YACs, yWXD1689 (lane 1), yWXD801 (lane 2), yWXD660 (lane 3) and total human DNA (lane 4) was digested with *Eco*RI and hybridized to the cDNA insert of the STS positive phage.

A cDNA for EST00205 was isolated, starting from a STS of 118 bp, from the embryo library. This 600 bp clone was then used to screen a teratocarcinoma NT2/D1 library by hybridization. Subsequently, a 4 kb cDNA clone for this gene was identified. When a 2 kb subcloned fragment was used as a probe in a multiple tissue Northern analysis, several forms of the mRNA were seen with a predominant 2 kb species in most tissues and a predominant variant >9.5 kb in brain (Fig. 2A). After standardization for the amount of messenger RNA using an actin probe (Fig. 2C) the relative levels of expression were determined. EST00205 appears to be most abundant in kidney, skeletal muscle, and heart tissues, less abundant in pancreas, liver and brain and expressed little or not at all in liver or placenta.

Northern analysis of the 2.1 kb cDNA recovered for the second EST, EST01014, showed it to be almost full-length (Figure 2B). This gene appears to be most highly expressed in pancreas, kidney, and brain tissue, more moderately expressed in skeletal muscle, placenta, and heart, and expressed at lower levels in lung and kidney tissue, with a message size of about 2.5 kb in all tissues. A 274 bp portion of the cDNA has also been independently isolated as XAP-3 using direct selection (28).

The localization of the EST00205 cDNA was confirmed by hybridization to YACs in the region centromeric to color vision, whereas the EST01014 cDNA was confirmed to be contained in the region telomeric to color vision (26; G.Palmieri and M.D'Urso, unpublished data). Partial sequence data showed a poly(A) tail at the presumptive 3' end of both genes but no significant homology to other sequences in Genbank.

Isolation of a cDNA clone for an evolutionarily conserved sequence

In a further application, a cDNA for an evolutionarily 'conserved sequence' (29) has been recovered. The primer pair, developed from the end-sequence of a yeast artificial chromosome insert of human DNA, gave a product of the same size from hamster

DNA as well as human DNA (Fig. 3A). A few pools were positive in the teratocarcinoma cDNA library, indicating that this is indeed part of a gene, rather rarely expressed in this source (probably $\sim 1/200,000$; data not shown). Subsequently, a cDNA clone of 1.85 kb was isolated. To demonstrate that it mapped back to the original YAC from which the STS derived, a Southern experiment was performed (Fig. 3B). When the entire cDNA was used as a probe, a band of 3.3 kb was seen in an *EcoRI* digestion of two YACs known to overlap and to be present in Xq25-q26 (R.Nagaraja, work in progress), yWXD1689 (lane 1) and yWXD801 (lane 2), as well as in total human DNA (lane 4). A third indifferent YAC, yWXD660 (lane 3), showed no hybridization with the probe.

DISCUSSION

The PCR-based screening scheme leads to the rapid isolation of a cDNA. The only requirement is an EST, STS, or enough sequence to develop a PCR primer pair from the transcribed region sought.

Single phage can be isolated through successive pools within a week. Also, once a probe is isolated for a particular STS, all positive subpools can be replated at high multiplicity (5 equivalents, or 15,000 phage/plate) to enhance the probability of finding all other cDNA clones in the library. The method lends itself to automated screening of wells, and even cDNAs from complex tissues could be analyzed with a moderate increase in the number of top-level pools.

Like hybridization-based methods, this approach would still tend to find a number of members of a gene family for many STSs, but the specificity of the assay is increased both by the requirement for an essentially complete match to the PCR oligonucleotide sequences and by the frequent availability of genomic site-specific ESTs or sequences from the 3'-untranslated regions of mRNAs.

In mapping efforts, thousands of PCR primer pairs for unique sequences are being developed. Some of these will represent transcribed sequences, and will be augmented by the sequencing of many regions to suggest putative exons. A battery of cDNA libraries would provide a reasonable test for transcription of the sequence. In initial attempts, PCR primer pairs developed from putative exons have indeed permitted the isolation of corresponding cDNAs (30). Protocols like that presented here could thus facilitate isolation of cDNAs and development of a transcription map of the genome.

ACKNOWLEDGEMENTS

We would like to thank Dr Giuseppe Pilia for helpful suggestions concerning the cDNA pooling strategy. This work was supported by NIH Grant #HG00247, the NATO Collaborative Research Grant #CRG 920110 and funds from C.N.R. Progetto Finalizzato Genoma Umano and Telethon grant 417.

REFERENCES

- Hanzlik, A.J., Osemlak-Hanzlik, M.M., Hauser, M.A. and Kurnit, D.M. (1993) *Nat. Genet.* 3, 44–48.
- Lovett, M., Kere, J. and Hinton, L.M. (1991) *Proc. Natl Acad. Sci. U.S.A.* 88, 9628–9632.
- Parimoo, S., Patanjali, S.R., Shukla, H., Chaplin, D.D. and Weissman, S.M. (1991) *Proc. Natl Acad. Sci. U.S.A.* 88, 9623.

- Duyk, G.M., Kim, S.W., Myers, R.M. and Cox, D.R. (1990) *Proc. Natl Acad. Sci. U.S.A.* 87, 8995–8999.
- Uberbacher, E. C. and Mural, R. J. (1991) *Proc. Natl Acad. Sci. U.S.A.* 88, 11261–11265.
- Guigo, R., Knudsen, S., Drake, N. and Smith, T. (1992) *J. Mol. Biol.* 245, 45–56.
- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merrill, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. and Venter, J.C. (1991) *Science* 252, 1651–1656.
- Adams, M.D., Dubnick, M., Kerlavage, A.R., Moreno, R., Kelley, J.M., Utterback, T.R., Nagle, J.W., Fields, C. and Venter, J.C. (1992) *Nature* 355, 632–634.
- Adams, M.D., Kerlavage, A.R., Fields, C. and Venter, J.C. (1993) *Nature Genetics* 4, 256–267.
- Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C. and Venter, J.C. (1993) *Nature Genetics* 4, 373–380.
- Barillot, E., Lacroix, B. and Cohen, D. (1991) *Nucl. Acid Res.* 19, 6241–6247.
- Green, E.D. and Olson, M.V. (1991) *Proc. Natl Acad. Sci. U.S.A.* 87, 1213–1216.
- Israel, D.I. (1993) *Nucl. Acid Res.* 21, 2627–2631.
- Skowronski, J., Fanning, T.G., and Singer, M.F. (1988) *Mol. Cell Biol.* 8, 1385–1397.
- Simeone, A., Gulisano, M., Acampora, D., Stornaiuolo, A., Rambaldi, M., and Boncinelli, E. *EMBO J.* 11, 2541–2550 (1992).
- Pengue, G., Calabro, V., Cannada-Bartoli, P., De Luca, P., Esposito, T., Taillon-Miller, P., LaForgia, S., Druck, T., Huebner, K., D'Urso, M., and Lania, L. (1993) *Hum. Mol. Genet.* 2, 791–796.
- Ciccociola, A., Dono, R., Obici, S., Simeone, A., Zollo, M. and Persico, M.G. (1989) *EMBO J.* 8, 1987–1991.
- Simeone, A., Acampora, D., Pannese, M., D'Esposito, M., Stornaiuolo, A., Gulisano, M., Kastury, K., Druck, T., Huebner, K. and Boncinelli, E. (1993) *Proc. Natl Acad. Sci. U.S.A.* 91, 2250–2254.
- Acampora, D., D'Esposito, M., Faiella, A., Pannese, M., Migliaccio, E., Morelli, F., Stornaiuolo, A., Nigro, V., Simeone, A. and Boncinelli E. (1989) *Nucl. Acids Res.* 17, 10385–10402.
- Faiella, A., D'Esposito, M., Rambaldi, M., Acampora, D., Balsofiore, S., Stornaiuolo, Mallamaci, A., Migliaccio, E., Gulisano, M., Simeone, A., and Boncinelli, E. (1991) *Nucl. Acids Res.* 19, 6541–6545.
- Wong, G.G., Witek, J.S., Temple, P.A., Wilkens, K.M., Leary, A.C., Luxenberg, D.P., Jones, S.S., Brown, E.L., Kay, R.M., Orr, E.C., Shoemaker, C., Golde, D.W., Kaufman, R.J., Hewick, R.M., Wang, E.A. and Clark, S.C. (1985) *Science* 228, 810–815.
- Del Sal, G., Manfioletti, G. & Schneider, C. (1989) *Biotechniques* 7, 514–519.
- Blanchard M.M., Taillon-Miller P., Nowotny P. and Nowotny V. (1993) *PCR methods Appl.* 2, 234–240.
- Feinberg, A. P. and Vogelstein, B. (1984). *Anal. Biochem.* 137, 266–267.
- Parrish, J.E. and Nelson, D.L. (1993) *Human Mol. Genet.* 11, 1901–1911.
- Mazzarella, R. and Srivastava, A. (1994) *Human Mol. Genet.* 3, 1095–1101.
- Schlessinger, D., Mandel, J.-L., Monaco, A.P., Nelson, D.L. and Willard, H.F. (eds) (1993) *Cyto. and Cell Genet.* 64, 148–194.
- Sedlacek, Z., Korn, B., Konecki, D.S., Siebenhaar, R., Coy, J.F., Kioschis, P., and Poustka, A.M. *Hum. Mol. Genet.* 2, 1865–1869 (1993).
- Mazzarella R., Montanaro V., Kere J., Reinbold R., Ciccociola A., D'Urso M. and Schlessinger D. (1992) *Proc. Natl Acad. Sci. U.S.A.* 89, 3681–3685.
- Zollo, M., Mazzarella, R., Bione, S., Toniolo, D., Schlessinger, D., D'Urso, M. and Chen E.Y. submitted.