



## Mapping expressed sequence tag sites on yeast artificial chromosome clones of *Arabidopsis thaliana* DNA.

F D Agyare, D A Lashkari, A Lagos, et al.

*Genome Res.* 1997 7: 1-9

Access the most recent version at doi:[10.1101/gr.7.1.1](https://doi.org/10.1101/gr.7.1.1)

---

**References** This article cites 19 articles, 5 of which can be accessed free at:  
<http://genome.cshlp.org/content/7/1/1.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Copyright © Cold Spring Harbor Laboratory Press

## RESEARCH

# Mapping Expressed Sequence Tag Sites on Yeast Artificial Chromosome Clones of *Arabidopsis thaliana* DNA

Francis D. Agyare,<sup>1,3,5</sup> Deval A. Lashkari,<sup>2,4,5</sup> Apostolos Lagos,<sup>1</sup>  
Allen F. Namath,<sup>2</sup> Gus Lagos,<sup>1</sup> Ronald W. Davis,<sup>2</sup> and  
Bertrand Lemieux<sup>1,2,6</sup>

<sup>1</sup>Department of Biology, York University, Ontario M3J 1P3 Canada; <sup>2</sup>Stanford DNA Sequence and Technology Center, Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305

We describe a method for efficient parallel mapping of expressed sequence tag (EST) sites onto yeast artificial chromosome (YAC) clones. The strategy involves an initial YAC clone pooling scheme that minimizes the number of required PCR amplifications. This is followed by parallel analysis of PCR amplicons of EST sequences. Using this method, we have screened 600 EST sites in combinatorial pools of 3449 YAC clones that contain *Arabidopsis thaliana* DNA inserts. The presence of these genes on YACs was detected by amplifying EST sequences with PCR and analyzing the reaction products by agarose gel electrophoresis. Of the 600 ESTs, 271 were found to map to individual YACs. Software tools are presented that allow for the automated analysis of this electrophoresis data. Suggestions for the scale-up of this method to map large genomes are discussed.

*Arabidopsis thaliana* is one of the most amenable genetic model organisms available for the study of plant biology (Somerville and Meyerowitz 1994). The genome of *Arabidopsis* is estimated to contain 100–120 million base pairs (Mbp) separated into five chromosomes. Large insert genomic libraries have been developed using yeast artificial chromosome (YAC) vectors (Ecker 1990; Ward and Jen 1990; Grill and Somerville 1991; Creusot et al. 1995) and P1 phage (Liu et al. 1995). Several large-scale efforts are under way to generate a physical map of the *Arabidopsis* genome. The efforts have culminated in the recent reports of complete contig maps of chromosomes 4 and 2 of the *Arabidopsis* genome by Schmidt et al. (1995) and Zachgo et al. (1996), respectively. Systematic cDNA sequencing programs have been initiated to generate expressed sequence tags (ESTs) (Hofte et al. 1993; Newman et al. 1994). This EST database was assembled to facilitate the identification of gene coding sequences in the ge-

nomeric DNA sequence contigs that will be assembled by the systematic genome sequencing efforts under way in several laboratories worldwide. However, EST sequences can also serve as landmarks in the physical mapping of the genome.

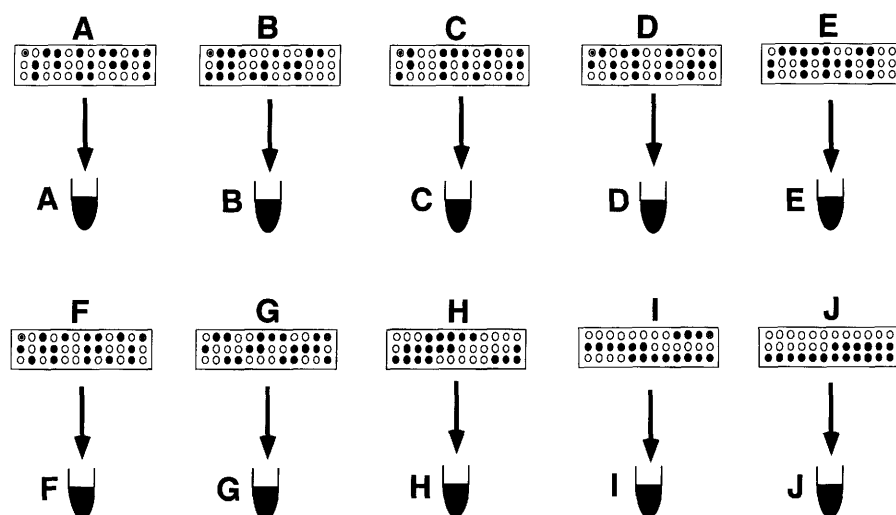
We have initiated a project aimed at the systematic mapping of ESTs to YAC clones of *Arabidopsis* genomic DNA. As the integration of the genetic and physical map of the genome of *Arabidopsis* becomes more accurate, the mapping of ESTs will take on greater importance as it will aid in the selection of candidate genes in map-based cloning efforts. The mapping of very large numbers of ESTs, however, is not a trivial technical task, and traditionally most high throughput EST content mapping strategies have made use of a combination of colony hybridization and PCR screening (Green and Olson 1990). However, without extensive use of laboratory automation, the process of scaling-up YAC colony hybridization experiments can often lead to a loss of quality in data acquisition. Moreover, simple pooling strategies (e.g., pooling a 96-well plate by column and by row) are not robust enough for large-scale PCR-based screens because of the numerous false positives and false negatives encountered with this method. To overcome these difficulties, various

**Present addresses:** <sup>3</sup>Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114; <sup>4</sup>Synteni, Inc., Palo Alto, California 94306.

<sup>5</sup>These authors contributed equally to this work.

<sup>6</sup>Corresponding author.

E-MAIL [Blemieux@cmgm.stanford.edu](mailto:Blemieux@cmgm.stanford.edu); FAX (415) 725-6044.



**Figure 1** *j*-detector pooling key. This key can be used to pool DNA samples isolated from groups of clones, DNA samples isolated from individual clones, or cells of individual clones. DNA samples or clones are arrayed in a microtiter plate in a 3-row  $\times$  12-column matrix. The black circles in the grids indicate the clones (or DNA samples) that are sampled for each *j*-detector. There are 10 *j*-detector pools in all, and each clone (or DNA sample) shares no more than three pools with other clones. For example, clone 1 (shown as a black dot in a white circle) is in pools A–D and F, whereas clone 12 (shown as a solid circle) is in clones A, C, F, G, and I. These clones share pools A, C, and F such that observing an EST PCR product in the latter as well as in any of pools B, D, G, or I would be sufficient to assign the product to either clone 1 or 12.

combinatorial pooling strategies have been devised (Barillot et al. 1991; Balding and Torney 1994).

To simplify EST mapping, we have developed a content mapping strategy relying entirely on the PCR detection of sequences in combinatorial pools. This strategy is based on the use of combinatorial pools, the structures of which are based on t-design theory (Beth et al. 1986). For example, the *j*-detector pooling scheme of Balding (1994) has the dimensions  $n = 36$ ,  $v = 10$ ,  $t = 3$ , and  $k = 5$ , where  $v$  is the number of pools,  $t$  is the number of pools shared by any two clones, and  $k$  is the number of pools with a given clone. The sampling key used to assemble these 10 pools, presented in Figure 1, was originally proposed by Balding and Torney (1994) to map sequence-tagged sites (STSs) on YAC clones of human DNA. As outlined by Balding (1994), the advantage of this pooling scheme is that clones share no more than three pools in common (Fig. 1) such that detection of a sequence in four pools is theoretically sufficient to identify which clone among the 36 contains the sequence. Unlike the  $N$ -dimensional pooling strategy proposed by Barillot et al. (1991), a single pooling design is used for the entire screening process. This greatly simplifies the development of

laboratory workstation software scripts for the automated assembly of these pools.

To minimize the number of PCR reactions, we have used “super *j*-detector” pools assembled with pooled DNA samples from large numbers of YACs to assign ESTs to primary pools of 36 YAC clones. In a secondary step, *j*-detector pools of these primary pools are screened for EST content. This strategy combines two essential qualities for PCR-based screens: (1) It requires very few PCR reactions to map most ESTs to at least one YAC; and (2) it has a built-in tolerance for one false-negative result. We have also developed a simple data handling algorithm to convert experimental data obtained with this pooling scheme into sequence content mapping information. Although we did not systematically screen the entire set of available YAC

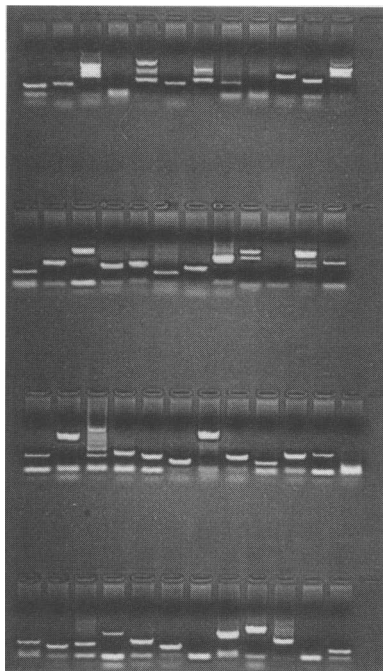
banks with each EST probe, our results show that this PCR-based sequence content mapping is feasible.

## RESULTS

The first step in any PCR-based mapping effort is to design and produce oligonucleotides with which to prime DNA synthesis. The technical and economic feasibility of this approach is, to a large extent, possible because of recent advances in automated multiplex oligonucleotide synthesis (AMOS) technology (Lashkari et al. 1995) and the development of a DNA sequence information database derived by the cDNA sequencing efforts at the Michigan State University/Department of Energy (MSU/DOE) Plant Research Laboratory (Newman et al. 1994) and the French *Arabidopsis* research network (Hofte et al. 1993). These data were used to design a total of 4000 oligonucleotides that could amplify 2000 *Arabidopsis* gene coding sequences. Primers were synthesized on a 20-nmole scale using a high throughput DNA synthesizer capable of producing 96 primers simultaneously in a convenient 96-well format (Lashkari et al. 1995). The primers, which have an average

length of 21 bases, are positioned within DNA sequences such that PCR products produced with cDNA templates would range between 100 and 450 bp. As introns in *Arabidopsis* genes are of modest size, 60% of the 1920 primer pairs tested on plant DNA gave PCR products. Figure 2 shows an example of the electrophoretic patterns obtained with PCR products generated with plant template DNA. Of the EST primers that gave PCR products, 40% were found to generate multiple bands. We opted to screen YAC banks with ESTs that gave single PCR products with plant DNA template.

The second step of this mapping effort was to isolate DNA from a set of 3449 YAC clones that were sampled from four different ordered YAC libraries. Of the first 1145 clones we screened, 934 were from the CEPH/INRA/CNRS collaborative group (CIC) (Creusot et al. 1995), 114 were from the EG bank (Grill and Somerville 1991), 97 were from the EW bank (Ward and Jen 1990), and the remaining 2304 clones were from the yUP library (Ecker 1990). The clones from the EG and EW libraries that we selected did not contain repetitive sequences or organelle DNA (Schmidt et al. 1994). Some of the EG

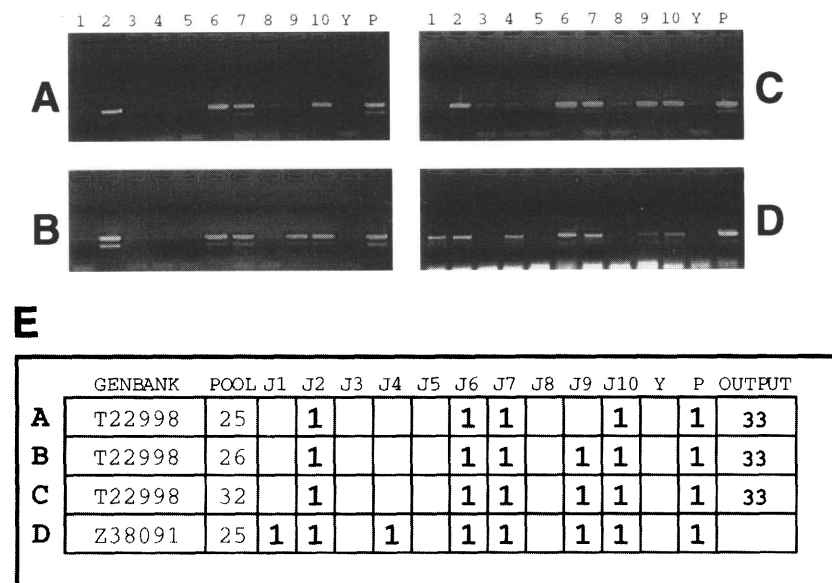


**Figure 2** Example of positive-control PCR reaction products obtained with oligonucleotides generated by AMOS. Plant DNA (100 ng) is used as template to amplify EST coding sequences by PCR. The products were separated with a 1% agarose gel stained with ethidium bromide before photography.

and EW clones have been assigned to mapped contigs based on hybridizations of RFLP probes to YACs (Hwang et al. 1991). This combined YAC library was sectorized into 96 “primary” pools of 36 clones each. DNA was isolated from these primary pools as well as from 960 *j*-detector pools assembled by sampling the 96 primary pools according to the key presented in Figure 1.

To reduce the number of PCR reactions needed to map an EST to a YAC, 36 DNA samples, each of which is extracted from a primary pool of 36 clones, were used to generate a set of 10 super *j*-detector pools. As one positive and one negative PCR controls are tested with each primer pair, the super *j*-detector step of the screening process allows us to screen 1296 YAC clones with 12 PCR reactions. The detection of four, five, or six PCR products in the super *j*-detector pools allows us to assign an EST to an individual primary pool (i.e., 36 possible clones). An EST that maps to a given primary pool is re-screened with a set of 10 *j*-detector pools assembled from the clones comprising the primary pool. The detection of four, five, or six PCR products in the *j*-detector pools allows the mapping of the EST to an individual clone in the pool. Each *j*-detector pool occupies one of the first 10 wells of a microtiter plate row. Wells 11 and 12 contain yeast and *Arabidopsis* DNA templates as negative and positive controls, respectively. Therefore, the two-step PCR-based screen outlined above can identify an individual YAC that contains a given EST from a set of 1296 YACs in 24 PCR reactions.

Figure 3 presents an example of the raw data obtained from a *j*-detector PCR screen. The PCR products from a complete set of *j*-detector amplifications, along with the accompanying positive and negative controls, occupy a single row of wells in the agarose gels used to analyze the reaction products. The synthesis of a PCR product and its size are compared to the products generated with the plant template DNA-positive control and the yeast DNA-negative control; this enables us to distinguish between a plant EST PCR product and a nonspecific PCR product. Although we performed our PCR amplifications with a single EST primer pair at a time, it should be possible to include multiple primer pairs in each reaction if the respective PCR amplicons have significantly different sizes. Some of the EST primers give PCR products in more than five *j*-detector pools (Fig. 3D). This result indicates that there is a possibility that more than one YAC contains this particular EST. Overall, we found that 6% of ESTs gave such a result with super *j*-detector pools.



**Figure 3** Gel electrophoresis separation of EST PCR amplification products obtained with *j*-detector pools. (A–D) Products obtained with *j*-detector template DNA are separated in lanes 1–10; the negative and positive controls, consisting of yeast DNA (Y) and plant DNA (P), are in lanes 11 and 12. In this particular example, EST T22998 was detected in three YACs (i.e., clone 33 from pool 25, clone 33 from pool 26, and clone 33 from pool 32) shown in A–C. The primers for EST Z38091 gave seven products in pool 25 and could not be mapped to a single clone (D). (E) An example of the spreadsheet compilation of electrophoresis data presented in (A–D). The first column lists EST clone GenBank numbers; column 2, the YAC primary pool being screened with the *j*-detector scheme; columns 3–12, the results of PCR amplification with *j*-detector pool DNA template; columns 13 and 14, the negative and positive PCR controls consisting of yeast (Y) and plant DNA (P) template; and column 15, the output cell for the clone-identifying algorithm. The numeral 1 in columns 3–12 signifies that a PCR product with the same size as the positive control has been detected in that particular column. The numeral 1 in the control columns indicates that clearly detectable PCR products were generated.

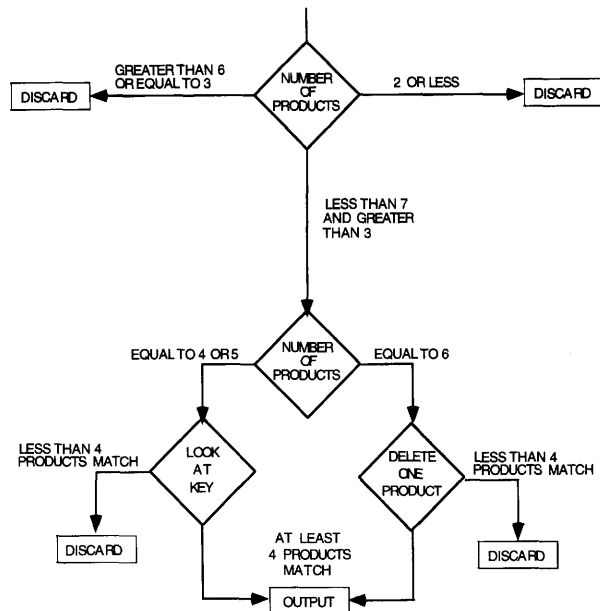
Results of the PCR amplifications are recorded on a Lotus 1-2-3 spreadsheet where each *j*-detector occupies a column and each EST a row. Figure 3E gives an example of the spreadsheet in which electrophoresis data are compiled and analyzed. It has one column that contains the clone number of the EST primers followed by 10 columns for entering the existence of plant DNA PCR reaction products. In addition, there are negative and positive control columns; these are followed by a series of columns indicating the positive pool or clone numbers identified using an algorithm that automatically identifies the YAC primary pool (or YAC clone) containing the EST. Columns that contain 1s indicate a positive PCR reaction product (i.e., a product with

the size of the plant-positive control). Blank entries are treated as 0 values by the algorithm and indicate either no reaction products or products of the incorrect size when compared to the positive control.

We have found that 31% of the ESTs can be mapped by screening the first 1296 YAC clones. Those that could not be mapped to a YAC in the first 36 primary pools are used to screen the next set of 36 primary pools. This second round of screening allowed us to map an additional 13% of the ESTs. Screening of the final set of 24 primary pools allowed us to map another 5% of the ESTs. In this feasibility study, we did not screen the entire set of available YACs with all of the ESTs.

Figure 4 presents a logical diagram of the decision-making process undertaken by an algorithm we have designed to automatically identify primary pools from tabulated electrophoresis data. The logic behind the software script takes into consideration the built-in redundancy of the pooling scheme and the occurrence of inevitable false negatives and false positives associated with PCR screening protocols. Four or five positive PCR products will give a unique result, whereas six plant PCR products can be the result of a false positive in cases where an EST maps to a single clone in the pool or a false negative in cases where two or more YAC clones contain the EST. When six products are detected, a macro can be used to automatically remove data from one *j*-detector (i.e., a column) and recalculate a limited number of possible outputs. These additional outputs are exported to cells to the right of the first output cell. Although not all of these possible outputs are confirmed as positives in the second stage of the screen, we have found that this solution is effective at resolving cases where more than one clone contains an EST. If there are more than six products, the program does not generate an output.

An example of the software script for the seventeenth primary pool or clone is expanded in Figure



**Figure 4** Logic diagram of the *j*-detector key algorithm.

5. The entire algorithm is a repetition of this code in which each *j*-detector cell is treated as a binary number (i.e., either 1 for a plant PCR product or 0 for no PCR product or a nonspecific product). Initially, the code determines whether there are at least four plant PCR products. If there are less than four products, then the program does not generate an output. If there are at least four, then the algorithm advances to the second step and determines whether there are six or fewer plant PCR products. In the case of four or five products, to confirm that these products correspond to a particular clone, the script verifies that none of the products appear in any of the five columns that should be empty in the case of a given clone and thus exclude the EST from mapping to that particular clone. If any one product fails to meet this criteria, the algorithm will conclude that the product cannot correspond to its clone number output and move on to the next clone number. Therefore, by using both the presence and absence of products to match the *j*-detector key, it is possible to identify a clone number with only four plant PCR products instead of five. In cases where there are six products, if the distribution of five products matches the key for a particular clone, the output is appended with a + to warn the operator that an additional YAC primary pool may contain the EST.

The results from super *j*-detector and *j*-detector pool screening are entered in separate files using the same spreadsheet matrix. The results of these

screens are merged by a PASCAL program that uses the spreadsheet files as input and generates a text file table where each EST and the YAC to which it mapped are listed on a common line. A sampling of our EST mapping data is presented in Table 1, along with YAC DNA sizing information and the map positions of YACs that have been assigned to mapped contigs with RFLP probes, are indicated.

To demonstrate the reliability of these data, we have probed Southern blots of individual clones with PCR products generated with plant DNA template. The results of five hybridizations to YAC DNA compared to a negative control of yeast DNA are shown in Figure 6. The EST probes hybridize to YAC clones but not to the yeast DNA-negative control and thus serve to validate our PCR-based sequence content mapping protocol.

## DISCUSSION

We have demonstrated that a screen for sequence content in a genomic library can be based entirely on PCR with minimal use of instrumentation. On the basis of YAC clone insert sizing information (Ecker 1990; Creusot et al. 1995), we calculate that the average insert size of the screened YACs is 300 kbp and that, assuming a genome size of 120 Mbp, each super *j*-detector contains 3.2 genome equivalents. Considering that 3.2 genome equivalents

```

1) @IF(@SUM(J1..J10)>3,
2)   @IF(@SUM(J1..J10)<7,
3)     @IF(@SUM(J1..J10)<6,
4)       @IF(J1<>1,
5)         @IF(J2<>1,
6)           @IF(J6<>1,
7)             @IF(J7<>1,
8)               @IF(J10<>1,"17" ," ")," ")," ")," ")
9)             @IF(J3=1,
10)            @IF(J4=1,
11)            @IF(J7=1,
12)            @IF(J8=1,
13)            @IF(J9=1,"17+" ," ")," ")," ")," ")
14)          ),
15)        ""),
16)      ""))
  
```

{\*\*END\*\*}

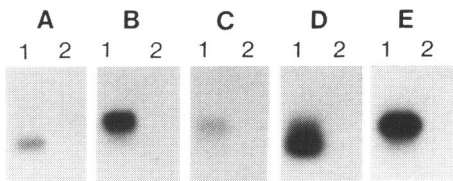
**Figure 5** Breakdown of the software script for the portion of the *j*-detector key algorithm that identifies clone 17.

<b>Table 1. EST Content Mapping Data</b>			
<b>GenBank accession no.</b>	<b>YAC anchored</b>	<b>YAC size (kb)</b>	<b>Chromosome no. and contig</b>
T04705	CIC3A6	510	3-FUS6
T04492	CIC3A3	1500	5-g17337
T04389	CIC4F1	480	1-g6838
T04335	CIC4D5	460	2-nga32
T04651	CIC5H3	460	5-nga151
T04657	yUP11C10	240	5-tt4-g5962
T04285	yUP13A1	250	1-m310
T04150	yUP13C10	300	1-m532
T04681	yUP15D11	260	2-GA1-g6844
T04280	yUP8H3	275	1-g4552-g16595
T04473	yUP3B5	350	3-g4125-AP3
T04814	CIC8B7	550	1-m219
T04814	yUP11B1	280	1-g8986
T04814	yUP13A6	270	1-m219
T43565	yUP4G9	110	1-g5972-nga63
T42542	yUP10H2	250	1-g5972-nga63
T41886	yUP12D7	290	1-ACC2-nga59
T43324	yUP14H2	120	4-pCITf3-pCITd23
T13795	yUP10B12	230	1-m402-m254
T20562	CIC4D4	470	5-nga139
T21228	CIC3B1	670	5-m555
T21115	CIC6H1	530	1-nga111
T22998	CIC10C9	480	1-g15894
T22943	CIC4D5	460	2-nga32
T22529	CIC3B1	670	5-m555
T21894	yUP2B4	300	chimeric
T22365	yUP3B5	350	3-g4125-AP3
T21295	yUP12D11	240	1-m335
T22185	yUP12D7	290	1-ACC2
T22792	yUP15B4	220	1-m241-m333
T21062	yUP11E5	190	1-m241-m333
T21562	yUP11E5	190	1-m241-m333
T21225	CIC5D1	765	1-g4026
T22676	CIC10B3	320	2-gpa1
T20705	yUP22H6	330	5-g6856
T21512	yUP24F12	180	2-m104
T20560	CIC10C9	480	1-g5972-nga63
Z34259	CIC10B3	320	2-GPA1
Z38091	CIC10C9	480	1-g15894
Z37622	CIC10E10	450	1-g4121
Z30817	yUP11E2	310	3-MR1-g6220
Z33694	yUP15F4	120	3-gapc-m302
Z33730	yUP11E4	320	4-CYCA1
Z37611	CIC3D3	555	5-Lfy-g17337
Z38044	yUP19B5	140	4-pCITd71
Z46821	yUP23A5	280	4-m210
Z37281	yUP24F2	290	3-m4523
Z34616	yUP22A9	250	3-m4014
Z35751	yUP2B4	300	chimeric
T04345	CIC2E12	640	2-GPA1
T04143	yUP2G1	250	5-nga225
T04229	yUP5H7	240	5-pCITd94-mi97

GenBank accession numbers for the ESTs were obtained from the GenBank database. The yUP YAC clone sizes were taken from the *Arabidopsis thaliana* Genome Center (ATGC) database, the CIC YAC clone sizes were obtained from D. Bouchez (INRA, Versailles, France), and the EG and EW YAC clone sizes were obtained from Hwang et al. (1991) and Putterill et al. (1993). All of the EST and YAC clones are available from the *Arabidopsis* Biological Resource Center (ABRC). The contig data were obtained from the ATGC database.

should provide a 96% probability of mapping any given EST (Clark and Carbon 1976), the low frequency of mapped ESTs that we have observed (31%) may be due to the fact that we did not equalize the number of yeast cells prior to the extraction but, rather, relied on a less accurate, visual estimation of the density of cell lawn in the microtube assemblies. Therefore, ESTs may not have been mapped to some of the slower growing clones because the amount of template DNA was too low. This problem may be particularly acute in the super *j*-detector pools. The solution to this problem would be to equalize the amount of DNA from each clone in all of the pools.

It is difficult to estimate the depth of the YAC library with results derived from this feasibility study because we did not systematically screen the entire set of 3449 available YACs with each EST probe. It should be noted, however, that the method described in Results could be used to screen larger numbers of YACs to identify multiple YACs for each EST. Although it is difficult to estimate the exact number of positive clones that contain a given EST when more than six of the *j*-detector pools yield PCR products, this situation did not occur frequently (e.g., with only 6% of the 600 ESTs in the case of the super *j*-detector pools). We conclude that a YAC primary pool size of 36 is adequate for an EST content mapping effort using these YAC banks. Balding and Torney (1994) are working on the development of an *l*-detector scheme with the dimensions  $n = 132$ ,  $v = 12$ ,  $t = 4$ ,



**Figure 6** Southern blot hybridization results of YAC clone DNA blot probed with EST PCR products from positive control reaction. DNAs from YAC clones EW5E6, CIC10C9, CIC10F7, CIC10F2, and yUP11F10 (lane 1, A–E, respectively) and the yeast negative controls (lane 2) were digested with *EcoRI* and probed with ESTs Z37283, T42819, T42804, T43565, and Z35749, respectively. Hybridization signal was detected in the YAC clone lanes but not in the yeast negative control lanes.

and  $k = 6$ . The algorithm we described in Results could easily be adapted to this larger *l*-detector scheme with few modifications. This modification could allow for a single-step PCR-based screen of a “minimum overlap” YAC bank of the *Arabidopsis* genome, which would be a rapid and efficient method for mapping cloned sequences onto YACs.

We suggest that a similar approach can be used for the characterization of total DNA libraries of organisms with larger genomes by (1) modifying the detector design to have larger dimensions, (2) using genomic banks with larger insert sizes, and (3) isolating DNA from individual clones and quantifying this DNA before pooling. Ewens et al. (1991) have presented some interesting calculations that, when projected graphically (Matallana et al. 1992), indicate the relationship between YAC insert size and the number of probes used to map a given genome through a random probe approach. For example, in the case of corn (a  $3 \times 10^9$ -bp haploid genome), using 2000 probes to screen a YAC library with an 800-kbp average insert size would generate a contig map composed of 2800 islands. Applying a two-step STS screening strategy to this particular example would require partitioning YACs into sets of 12 super *l*-detector pools and their corresponding sets of *l*-detector pools. Each of the super *l*-detector pools would contain 2.3 genome equivalents. Considering that the super *j*-detector pools that have been used successfully to screen *Arabidopsis* YACs for sequence content have 3.2 genome equivalents, the aforementioned super *l*-detector pools should be able to map ~25% of the STS to one YAC with as little as 28 reactions.

## METHODS

### YAC Clone Manipulation

Yeast clones were grown individually in Beckman minitube assemblies according to Agyare et al. (1997). We have developed a software script for the Biomek-1000 which assembles the *j*-detector pools proposed by Balding (1994). This script is stored at [http://www.yorku.ca/ftp/york\\_other/cgat/](http://www.yorku.ca/ftp/york_other/cgat/). The configuration of our instrument is 250- $\mu$ l tips,  $13 \times 100$ -mm tubes in tray 1, 96-well flat-bottom plate in tray 2, and 1-ml minitube box (Beckman) in tray 3 and a single-tip P200 tool. All liquid handling is done with the single-channel P200 tool using the generic pipetting function. Tray 3 is designated as the source and tray 1 the destination and 150  $\mu$ l of cells from the source tubes is delivered into each of the five tubes in the destination tray according to the *j*-detector pattern of Balding (1994) such that each glass tube will contain 2.7 ml of resuspended cells. A volume of 50  $\mu$ l of the cell suspension is mixed three times between each transfer to ensure that a uniform amount of cells are transferred. This mixing was done at the bottom level of the 1-ml minitubes to obtain best results. DNA was extracted from pools of individually grown clones to limit bias because of differential growth of clones. The minitube assemblies used to grow cells were visually examined and recorded with a CCD camera to ensure significant growth of all constituent clones before pooling for DNA extraction. The DNA extraction protocol is a modification of the method of Hoffman and Winston (1987) as outlined in Agyare et al. (1997). Plant DNA was extracted according to Rogers and Bendish (1988).

### Oligonucleotide Design

Oligonucleotide sequences were selected from EST sequence files using the PRIMER program developed at the Whitehead Institute (Cambridge, MA). Primer design was controlled by a set of predetermined criteria stored in a file that is accessed by PRIMER. Our primers were selected for an optimal length of 20 bases  $\pm$  2 based on their melting temperatures being between 60°C and 65°C. The length of the PCR products expected based on the sequence files can also be selected by PRIMER such that we usually set several ranges to obtain alternative primer pairs for each EST; for example, a range of 100–150 bases, one of 151–250, one of 251–300, one of 301–350, and one of 351–400. The syntheses were performed as described in Lashkari et al. (1995). The sequence of the primers are posted at York University ([http://www.yorku.ca/ftp/york\\_other/cgat/](http://www.yorku.ca/ftp/york_other/cgat/)).

### PCR Amplifications

All of the components needed for PCR amplification were added to a master mix in the following proportions for 96-well microamp tray assembly: 206  $\mu$ l of  $10\times$  PCR reaction buffer, 206  $\mu$ l of 2 mM dNTPs, 186  $\mu$ l of 15 mM  $MgCl_2$ , 720  $\mu$ l of sterile  $ddH_2O$  and, 20  $\mu$ l of *Taq* DNA polymerase (Perkin Elmer); the enzyme is added just prior to dispensing 168  $\mu$ l of this master mix into eight tubes. As each row of the assembly is used to screen one EST on the complete *j*-detector set, 20  $\mu$ l of the appropriate forward and reverse primer stock solutions (10 pmole/ $\mu$ l) was added to each tube. A volume of 14  $\mu$ l of this mixture was dispensed into each well of the microamp

## AGYARE ET AL.

assembly with a BioHit 8-channel pipette. A volume of 5  $\mu$ l of template DNA solutions (20 ng/ $\mu$ l) is added to the microamp assembly with a 12-channel pipette. The assembly is centrifuged for 30 sec to ensure that all reagents are mixed. Amplifications were performed in a Perkin Elmer system 9600 thermal cycler with an initial denaturation at 95°C for 1 min followed by 40 cycles of 94°C for 30 sec, 62°C for 30 sec, 72°C for 30 sec, and a final extension at 72°C for 5 min. Products were separated by electrophoresis at 120 V for 1 hr through 2% (wt/vol) agarose gels prestained with ethidium bromide (Sambrook et al. 1989). The banding patterns of these gels were recorded with an Alpha Innotech gel documentation system.

## Data Availability

The processed mapping data are listed on the *Arabidopsis thaliana* Genome Center (ATGC) Database (<http://cbil.humgen.upenn.edu>), as well as the *Arabidopsis thaliana* Database (AtDB) (<http://genome-www.stanford.edu/arabidopsis/EST2YAC.html>), which currently contain 271 EST map entries.

## ACKNOWLEDGMENTS

This work was supported by grant GO-12403 from the Canadian Genome Analysis and Technology (CGAT) program to B.L. and by grant NIH1P01 HG00205 from the National Institutes of Health (NIH) to R.W.D. We thank Dr. D. Flanders (Stanford University) for posting our data on the AtDB, Dr. J.R. Ecker (University of Pennsylvania) for posting our EST mapping data on the AtGC server as well as in the yUP YAC bank, Dr. D. Bouchez (INRA, Versailles) for providing the CIC YAC bank before publication, Drs. T. Newman (Michigan State University), and E. Retzel (University of Minnesota) for assistance in correlating EST clone numbers and sequence files. We also thank Dr. J. Quackenbush and Dan Shoemaker (Stanford University) for critically reading the manuscript.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

## REFERENCES

Agyare, F.D., G. Lagos, D. Lashkari, R.W. Davis, and B. Lemieux. 1997. Mapping of cloned sequences on yeast artificial chromosomes. In *Arabidopsis protocols* (ed. J.M. Martinez-Zapater and J. Salinas), Humana Press (In press).

Balding, D.J. 1994. Design and analysis of chromosome physical mapping experiments. *Phil. Trans. R. Soc. London B.* **344**: 329–335.

Balding D.J. and D.C. Torney. 1994. YAC library pooling scheme for PCR-based screening. In *The human genome project: Deciphering the blueprint of heredity* (ed. N.G. Cooper), p. 135. University Science Books, Hong Kong.

Barillot, E., B. Lacroix, and D. Cohen. 1991. Theoretical analysis of library screening using an N-dimensional pooling strategy. *Nucleic Acids Res.* **19**: 6241–6247.

Beth, T., D. Jungnickel, and H. Lenz. 1986. *Design theory*. Cambridge University Press, Cambridge, England.

Clark, L. and J. Carbon. 1976. A clone bank containing synthetic ColEI hybrid plasmids representative of the entire *E. coli* genome. *Cell* **9**: 91–99.

Creusot, F., E. Fouilloux, M. Dron, J. Lafleuriel, G. Picard, A. Billault, D. Le Paslier, D. Cohen, M.-E. Chaboute, A. Durr, J. Fleck, C. Gigot, C. Camilleri, C. Bellini, M. Caboche, and D. Bouchez. 1995. The CIC library: A large insert YAC library for genome mapping in *Arabidopsis thaliana*. *Plant J.* **8**: 763–770.

Ecker, J.R. 1990. Pulse field gel electrophoresis and yeast artificial chromosome analysis of the *Arabidopsis* genome. *Methods* **1**: 186–194.

Ewens, W.J., C.J. Bell, P.J. Donnelly, P. Dunn, E. Matallana, and J.R. Ecker. 1991. Genome mapping with anchored clones: Theoretical aspects. *Genomics* **11**: 799–805.

Green, E.D. and M.V. Olson. 1990. Systematic screening of yeast artificial chromosome libraries by use of the polymerase chain reaction. *Proc. Natl. Acad. Sci.* **87**: 1213–1217.

Grill, E. and C.R. Somerville. 1991. Construction and characterization of a yeast artificial chromosome library of *Arabidopsis* which is suitable for chromosome walking. *Mol. & Gen. Genet.* **226**: 484–490.

Hoffman, C.S. and F. Winston. 1987. A Ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *E. coli*. *Gene* **57**: 267–272.

Hofte, H., T. Desprez, J. Amselem, H. Chiapello, M. Caboche, A. Moisan, M.-F. Jourjon, J.-L. Charpentreau, P. Berthomieu, D. Guerrier, G. Giraudat, F. Quigley, F. Thomas, D.-Y. Yu, R. Mache, M. Raynal, R. Cooke, F. Grellet, M. Delseny, Y. Parmentier, G. de Marcillac, C. Gigot, J. Fleck, G. Phillips, M. Axelos, C. Bardet, D. Tremousaygue, and B. Lescure. 1993. An inventory of 1,152 expressed sequence tags obtained by partial sequencing of cDNAs from *Arabidopsis thaliana*. *Plant J.* **4**: 1051–1061.

Hwang, I., T. Kohchi, B. Hauge, H.M. Goodman, R. Schmidt, G. Cnops, C. Dean, S. Gibson, K. Iba, B. Lemieux, V. Arondel, L. Danhoff, C.R. Somerville. 1991. Identification and map position of YAC clones comprising one third of the *Arabidopsis* genome. *Plant J.* **1**: 367–374.

Lashkari, D., S.P. Hunicke-Smith, R.M. Norgen, R.W. Davis, and T. Brennan. 1995. An Automated multiplex oligonucleotide synthesizer: Development of high-throughput, low-cost DNA synthesis. *Proc. Natl. Acad. Sci.* **92**: 7912–7915.

Liu, Y.-G., N. Mitsukawa, A. Vazquez-Tello, and R.F. Whittier. 1995. Generation of a high-quality P1 library of *Arabidopsis* suitable for chromosome walking. *Plant J.* **7**: 351–358.

Matallana, E., C.J. Bell, P. Dunn, M. Lu, and J.R. Ecker.

1992. Genetic and Physical linkage of the *Arabidopsis* genome. Methods for anchoring YACs. In *Methods in Arabidopsis research* (ed. C. Koncz, N.-H. Chua, and J. Schell), pp. 144–169. World Scientific, Singapore.

Newman, T., F.J. de Bruijn, P. Green, K. Keegstra, H. Kende, L. McIntosh, J. Ohlrogge, N. Raikel, S. Somerville, M. Thomashow, E. Retzel, and C.R. Somerville. 1994. Genes galore: A summary of methods for accessing results from large-scale partial sequencing of anonymous *Arabidopsis* cDNA clones. *Plant Physiol.* **106**: 1241–1255.

Putterill J., F. Robson, K. Lee, and G. Coupland. 1993. Chromosome walking with YAC clones in *Arabidopsis*: Isolation of 1700 kbp of contiguous DNA on chromosome 5, including a 300 kbp region containing the flowering-time gene CO. *Mol. & Gen. Genet.* **239**: 145–157.

Rogers, S.O. and A.J. Bendish. 1988. Extraction of DNA from plant tissues. In *Plant Molecular Biology manual A6* (ed. S.B. Gelvin and R.A. Schilperoot), pp. 1–10. Kluwer Academic Press, Dordrecht, Belgium.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Schmidt, R., J. Putterill, J. West, G. Cnops, F. Robson, G. Coupland, and C. Dean. 1994. Analysis of clones carrying repeated DNA sequences in two YAC libraries of *Arabidopsis thaliana* DNA. *Plant J.* **5**: 735–744.

Schmidt, R., J. West, K. Love, Z. Lenchan, D. Lister, H. Thompson, D. Bouchez, and C. Dean. 1995. Physical map and organization of *Arabidopsis thaliana* chromosome 4. *Science* **270**: 480–483.

Somerville, C.R. and E.M. Meyerowitz. 1994. Introduction. In *Arabidopsis* (ed. C.R. Somerville and E.M. Meyerowitz), pp. 1–6. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Ward, E.R. and G.C. Jen. 1990. Isolation of single-copy-sequence clones from a yeast artificial chromosome library of randomly shared *Arabidopsis thaliana* DNA. *Plant Mol. Biol.* **14**: 561–568.

Zachgo, E.A., M.L. Wang, J. Dewdney, D. Bouchez, C. Camilleri, S. Belmonte, L. Huang, M. Dolan, and H.M. Goodman. 1996. A physical map of chromosome 2 of *Arabidopsis thaliana*. *Genome Res.* **6**: 19–25.

Received May 10, 1996; accepted in revised form December 3, 1996.