



## Screening of Gene-Associated Polymorphisms by Use of In-Gel Competitive Reassociation and EST (cDNA) Array Hybridization

Koshichi Gotoh and Michio Oishi

*Genome Res.* 2003 13: 492-495

Access the most recent version at doi:[10.1101/gr.434103](https://doi.org/10.1101/gr.434103)

---

**References** This article cites 7 articles, 4 of which can be accessed free at:  
<http://genome.cshlp.org/content/13/3/492.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](#).

---

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." In the center, there is a white box with the text "LEARN MORE". On the right, there is a woman wearing a red and white superhero cape and mask, and the logo for "CELLECTA" which consists of a cluster of green dots.

CRISPR and RNAi Genetic Screening.  
Your new superpower.

LEARN MORE

CELLECTA

---

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

---

Cold Spring Harbor Laboratory Press

## Methods

# Screening of Gene-Associated Polymorphisms by Use of In-Gel Competitive Reassociation and EST (cDNA) Array Hybridization

Koshichi Gotoh<sup>1</sup> and Michio Oishi

Human Gene 2, Kazusa DNA Research Institute, 2-6-7, Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Japan

In-gel competitive reassociation (IGCR) is a method of differential subtraction to enrich polymorphic DNA restriction fragments between two DNA samples without probes or specific sequence information. Here, we show that by combining IGCR and expressed sequence tags (EST) array hybridization, polymorphic DNA fragments associated with genes in complex higher organisms (*Arabidopsis thaliana*) can be effectively screened, demonstrating that this procedure offers a simple and efficient way to obtain gene-associated polymorphic DNA markers.

[The following individuals kindly provided reagents, samples, or unpublished information as indicated in the paper: N. Sakurai, D. Shibata, and C. Kuwata.]

Polymorphic DNA markers are very useful tools for tracing genetic traits that affect phenotypes as well as for studying biological diversity among species. Construction of polymorphic DNA marker libraries or databases, therefore, is considered to be the next important step after the genome sequences of major model organisms are determined. At present, however, random shotgun sequencing is the strategy of choice for identifying polymorphic markers. We previously reported that highly extensive polymorphic DNA libraries could be constructed without probes or specific sequence information by employing a differential subtraction procedure that is based on in-gel competitive reassociation (IGCR) (Inoue et al. 1996). Considering the importance of collecting gene-associated polymorphic DNA markers in various aspects of genome-wide studies, we show here, by a combination of IGCR and expressed sequence tags (EST) (cDNA) array hybridization, an effective strategy to screen gene-associated polymorphic DNA markers in complex genomes using two strains of the higher plant, *Arabidopsis thaliana*, as a model.

## RESULTS

### In-Gel Competitive Reassociation (IGCR)

To screen DNA fragments with restriction fragment length polymorphism (RFLP) between two complex genomes, we first performed IGCR for differential subtraction of polymorphic DNA fragments between the genomic DNA of two strains of the higher plant *A. thaliana*; the Landsberg *erecta* strain (as target), and Columbia strain (as reference) (Fig. 1). As the control, we employed the same IGCR procedure using Landsberg *erecta* DNA as both target and reference DNA and used the library as the standard for the enrichment of RFLP fragments. Although the entire genome sequence of the Columbia strain has been determined (The *Arabidopsis* Genome Initiative 2000), that of the Landsberg *erecta* strain has not yet been completely determined. To monitor the efficiency of enrichment, a single copy adenovirus-2 DNA was added to the

target Landsberg *erecta* genomic DNA. Electrophoretic patterns of *Taq* I-digested Columbia and Landsberg *erecta* genomic DNA as well as control and IGCR subtraction samples are shown in Figure 2A. The adenovirus-2 fragments added to target DNA were actually enriched 3- to 20-fold in IGCR sample compared with the control sample (Fig. 2B), indicating that the differential subtraction was successfully carried out under the condition described above.

### EST Macroarray Hybridization

To screen specific polymorphic DNA fragments associated with genes among the IGCR sample, we performed EST (cDNA) array hybridization (Asamizu et al. 2000) employing two sets of DNA samples, the IGCR and control sample (see above), as probes for hybridization. After hybridization, ratios of the signal intensities of each spot obtained with the IGCR DNA probe to the control DNA probe (IGCR/control) were determined. It is expected that if a polymorphic DNA fragment associated with a specific EST is enriched by IGCR, the IGCR DNA probe would give higher signal intensity compared with the control DNA probe for the specific EST spot. A portion of the arrays, which contained 480 ESTs, was used for the experiment (each in duplicate, see Fig. 3A) and calculated the average ratios (IGCR/control) of the intensities for each of the duplicated EST spots. All the EST spots were renumbered from 1–480 and are shown diagrammatically in Figure 3B in order of signal intensity from the highest ratio to the lowest.

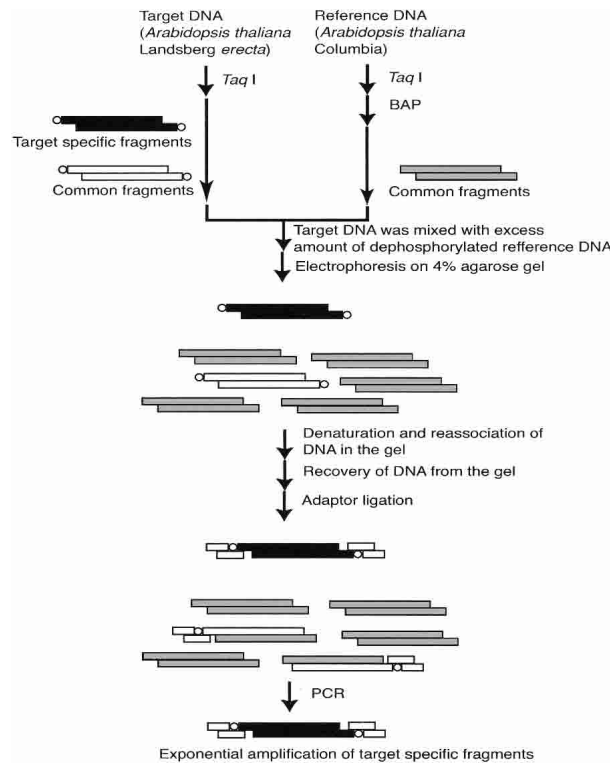
### Analysis of Differentially Hybridized ESTs

To examine whether differentially hybridized ESTs with higher IGCR/control ratios were actually the result of the enrichment of the corresponding gene-associated polymorphic DNA fragments, we performed Southern hybridization using some of the EST inserts as probes against *Taq* I treated genomic DNA of both Landsberg *erecta* and Columbia strains, as well as against the IGCR and control DNA samples. EST clones that belong to three categories in respect to the IGCR/control ratios were used as probes; five clones that gave the highest signal ratios (clones 1, 3, 4, 5, and 6; average ratio 3.92), five clones randomly selected from a group that gave relatively high ratios (11, 12, 13, 14, 27, and 32; average ratio 2.35) and

<sup>1</sup>Corresponding author.

E-MAIL [koshichi@kazusa.or.jp](mailto:koshichi@kazusa.or.jp); FAX 81-438-52-3946.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.434103>.



**Figure 1** Schematic diagram of in-gel competitive reassociation (IGCR) procedure. Closed bars, target-specific DNA fragments; open bars, common DNA fragments derived from target DNA; shaded bars, common DNA fragments derived from reference DNA; short open bars, adaptors for PCR amplification. BAP, bacterial alkaline phosphatase. Circles at the end of the bars represent phosphate moieties. Target-specific restriction fragment length polymorphism products, which are enriched after IGCR, are shown at the bottom of the chart (closed bars with adapters).

10 clones randomly selected from a group that gave virtually no increase in relative intensity (52, 85, 116, 241, 263, 316, 337, 366, 445, and 468; average ratio 1.12) (Fig. 3). As shown in Figure 4, in Southern hybridization, all the EST clones with the highest ratios as well as those with relatively high ratios (except for clone 4) exhibited clearly visible RFLP bands (lane *a* for Columbia genomic DNA and lane *b* for Landsberg *erecta* genomic DNA). The RFLP bands present in the Landsberg *erecta* DNA (lane *b*) were actually enriched in the IGCR samples (lane *c* contains the control sample and lane *d* contains the IGCR sample). On the other hand, when EST clones that showed the lowest ratios (clones 52–468) were used as probes, most of the clones (8 out of 10 clones except for clones 337 and 366) did not exhibit RFLP patterns (compare lanes *a* and *b*) and the bands were not enriched by IGCR (compare lanes *c* and *d*). These results show that IGCR is very effective in screening polymorphic DNA fragments, and that EST macroarray hybridization allowed us to identify genes associated with individual polymorphic DNA fragments. Sequence analysis of these gene-associated polymorphic clones showed that six clones possess single-point mutations at the *Taq* I recognition site (Nos. 1, 3, 6, 13, 27, and 337); four have insertion-deletions (Nos. 11, 14, 32, and 366) and two have both a point mutation at the *Taq* I site and insertion-deletions (Nos. 5 and 12).

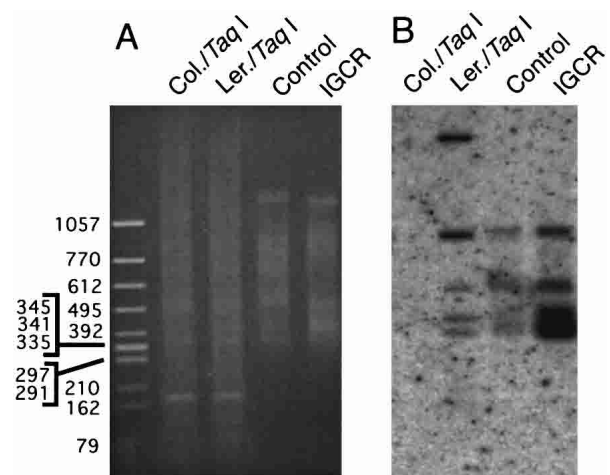
## DISCUSSION

Although all the polymorphic fragments examined in this study showed clearly recognizable RFLP bands by Southern hybridization, we found that RFLP fragments different in only several base pairs (406 bp versus 398 bp, data not shown) could be enriched, indicating that RFLP fragments derived from small insertions and deletions can also be effectively screened. Gene-associated polymorphic DNA marker libraries may become even more extensive by constructing similar DNA libraries derived from genomic digests produced by other restriction enzymes.

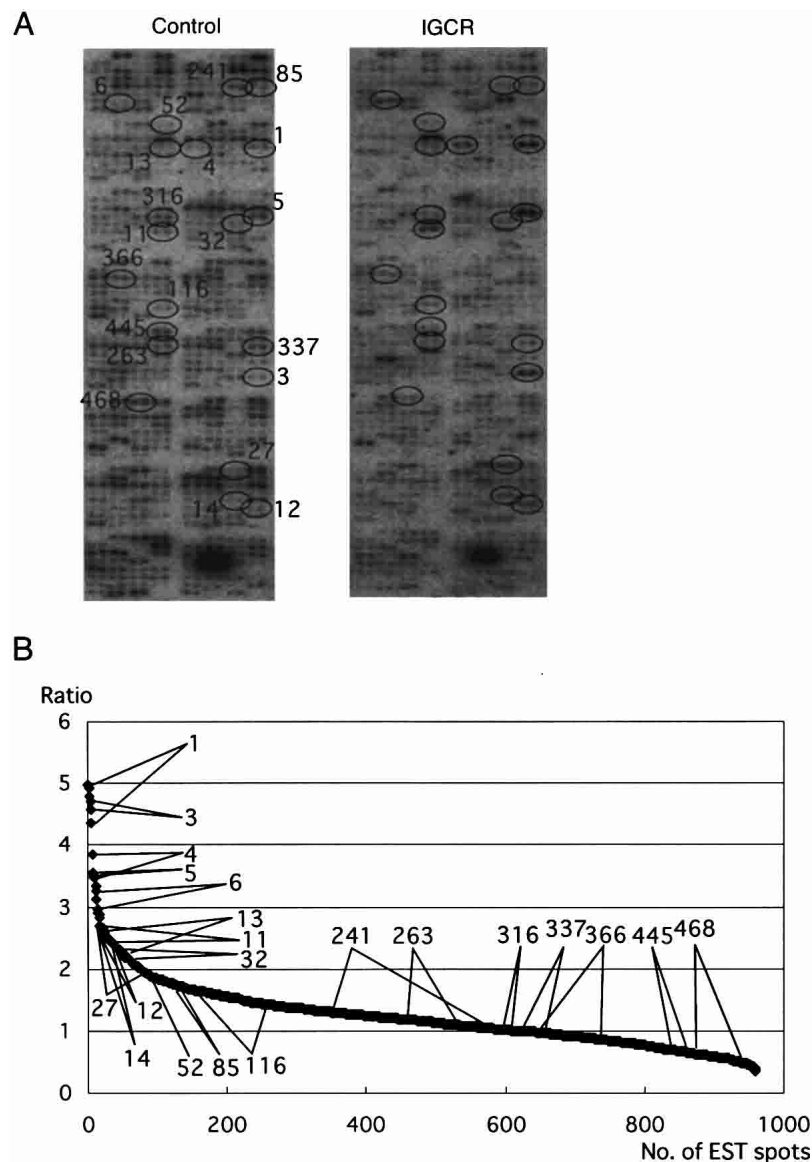
In theory, similar differential subtraction procedures, like RDA (Lisitsyn and Wigler 1993; Rosenberg et al. 1994), can be applied for the same purpose as presented here. IGCR, however, offers more versatile applications in constructing RFLP DNA marker libraries, particularly from organisms with highly complex genomes, as reassociation in the gel, rather than in solution, not only avoids illegitimate reassociation between DNA fragments carrying similar, but not identical, base sequences, but also substantially increases reassociation efficiency, which is critical in dealing with complex genomes.

Because the size of the EST used in this study is ~1.2 kb, and the average size of restriction fragments produced by 4-base cutters is 256 bp, the frequency of the *Taq* I site in the EST is estimated to be 4.7 *Taq* I sites per EST (1200/256). If all the EST clones showing relatively high signal intensity ratios (>2) are associated with RFLP fragments, ~10% of the ESTs should contain RFLP fragments (Fig. 3B), indicating one polymorphism per 47 *Taq* I sites. This corresponds to one polymorphism per 188 bp ( $4 \times 47$ ). This estimate is roughly the same as the observed frequency of polymorphisms in the two strains (Chang et al. 1988; Bergelson et al. 1998).

In summary, we showed that polymorphic DNA fragments associated with genes can be effectively screened by combining IGCR and EST array hybridization. The procedure is relatively simple and effective, as evidenced by the observation that almost all ESTs that gave stronger signals by the



**Figure 2** Enrichment of adenovirus-2 DNA by IGCR. (A) Electrophoretic patterns of DNA stained with ethidium bromide. DNA size markers ( $\Phi$ X174-*Hinc* II digest, TaKaRa, Japan) are also shown. (B) Southern hybridization by a 4.7-kb *Bam* HI adenovirus-2 DNA fragment as the probe. Col./*Taq* I, *Taq* I-digested Columbia genomic DNA (500 ng); Ler./*Taq* I, *Taq* I-digested Landsberg *erecta* genomic DNA (500 ng) with a single copy adenovirus-2 DNA; Control, control subtraction sample (300 ng); IGCR, IGCR DNA sample (300 ng).



**Figure 3** Screening of gene-associated restriction fragment length polymorphism (RFLP) fragments by EST macroarray. (A) Radioactive images of an EST macroarray after hybridization. (Left panel) Probed with the control DNA sample. (Right panel) Probed with the IGCR DNA sample. (B) Ratios (IGCR/control) of signal intensities at each spot were calculated (averaged from duplicate samples), re-numbered from 1 (the highest) to 480 (the lowest) and plotted in that order.

IGCR DNA probe than by the control probe were associated with polymorphic DNA fragments. This would be particularly useful in obtaining gene-associated polymorphic (RFLP) DNA marker libraries between two related organisms in which either one or both of the genomic sequences are not well understood.

## METHODS

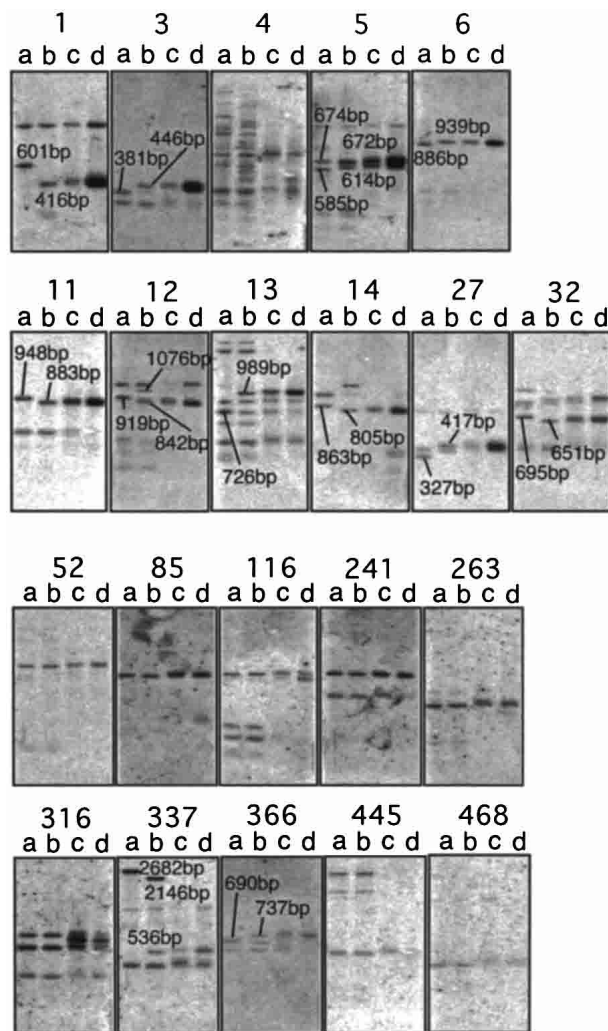
### Genomic DNA and the Adaptors

Genomic DNA of two *A. thaliana* strains, Columbia and Landsberg *erecta*, were kindly supplied by Dr. N. Sakurai and Dr. D. Shibata (Kazusa DNA Research Institute). Two *Taq* I adaptors (A and R) were used for PCR amplification of whole genomic

DNA. The adaptors were prepared by annealing synthetic oligonucleotides 5'-AAATGGATCCTTGCGGCCGCAT3' and 5'-CGATGCGGCCGC-3' for adaptor A and 5'-AGCACTCTCCAGCCTCTCACCGAT3' and 5'-CGATCGGTGA-3' for adaptor R. The 5'-AAATGGATCCTTGCGGCCGCAT3' and 5'-AGCACTCTCCAGCCTCTCACCGAT3' were used as PCR primers for adaptor A and R, respectively.

### IGCR Procedure

The IGCR procedure was modified from the original report (Inoue et al. 1996) as described below and is shown diagrammatically in Figure 1. Genomic DNA of the two *Arabidopsis* strains (Landsberg *erecta* as target DNA and Columbia as reference DNA) was digested with *Taq* I. *Taq* I-digested adenovirus-2 DNA (60 pg, equivalent to a single copy, Life Technologies, Inc.) was added to the *Taq* I-digested Landsberg *erecta* DNA (200 ng) as a marker to monitor the enrichment of polymorphic DNA fragments. The Landsberg *erecta* DNA (with adenovirus-2 DNA) and Columbia DNA digests were ligated to adaptor A (10 pmole to each 10 ng of DNA) and were PCR amplified using primer A (60 pmole) in a 50- $\mu$ L reaction under the following cycling conditions: 72°C for 5 min, then 10 amplification cycles (94°C for 20 sec, 68°C for 10 min), followed by a final extension at 68°C for 10 min. The amplified DNA was digested with *Taq* I again and purified by Qiaquick PCR purification kit (Qiagen). The PCR-amplified Columbia genomic DNA (20  $\mu$ g) was dephosphorylated by bacterial alkaline phosphatase (5.6 units) in a 400- $\mu$ L reaction, that contained 150 mM Tris-HCl (pH 8.0) and 10 mM MgCl<sub>2</sub> at 56°C for 1 h. The resulting solution was then phenol-extracted and the DNA precipitated with ethanol. To ensure complete dephosphorylation, the dephosphorylation step was repeated twice more. The Landsberg *erecta* DNA (100 ng) was mixed with 10  $\mu$ g of the dephosphorylated Columbia DNA, and electrophoresed on an agarose gel (4% agarose, Nakalai Tesque, Japan) at 120 V for 5 h in TAE buffer. A portion of the gel containing ~0.2- to 1.0-kb DNA fragments was excised and soaked twice in DNA denaturation buffer (0.5 M NaOH, 0.9 M NaCl) at 45°C for 45 min. The gel was then incubated four times at 45°C for 20 min each in a fresh reassociation buffer containing 50% formamide, 50 mM Tris-HCl (pH 8.0), 0.9 M NaCl, 10 mM EDTA, 10% polyethylene glycol 8000, then further incubated in the same buffer at 45°C for 35 h and washed five times with TAE buffer at room temperature for 20 min. DNA was recovered from the gel in a dialysis bag by electroelution and purified using the Qiaquick PCR purification kit (Qiagen). The DNA (total, ~450 ng) was ligated to adaptor R (100 pmole) and a quarter of the sample was amplified by PCR with 60 pmole of primer R in a 50- $\mu$ L reaction, using the following cycling conditions: 72°C for 5 min, then 20 amplification cycles (94°C for 20 sec, 68°C for 10 min), followed by final extension at 68°C for 10 min. The PCR product (1  $\mu$ L) was subjected to a second round of PCR cycles for nine cycles. The amplified DNA was digested with *Taq* I and purified by the Qiaquick PCR purification kit (Qiagen). In parallel (con-



**Figure 4** Southern hybridization analysis. Membranes that are essentially the same as shown in Figure 2 were subjected to Southern hybridization using EST clones as probes. Lane *a*, *Taq* I-digested Columbia genomic DNA (500 ng), lane *b*, *Taq* I-digested Landsberg *erecta* genomic DNA (500 ng) with a single copy adenovirus-2 DNA, lane *c*, control IGCR DNA sample (300 ng), and lane *d*, IGCR sample (300 ng). The numbers shown on each panel are those of the clones shown in panel *B* in Figure 3. When restriction fragment length polymorphism (RFLP) DNA bands were clearly shown, the genomic regions corresponding to the RFLP bands for both Columbia and Landsberg *erecta* strains were PCR-amplified and the products were sequenced. The size (bp) of each pair of the RFLP fragments is indicated.

trol subtraction), the same IGCR procedure was performed using 100 ng of Landsberg *erecta* DNA (with adenovirus DNA)

and 10  $\mu$ g of dephosphorylated Landsberg *erecta* DNA (with adenovirus DNA).

### Hybridization with *Arabidopsis* EST Macroarray

Inserts of 11,520 EST (cDNA) clones from *A. thaliana* Columbia (Asamizu et al. 2000) were spotted, in duplicate, onto nylon membranes (8  $\times$  12 cm). As probes, DNA samples (100 ng) were labeled with [ $\alpha$ - $^{32}$ P] dCTP (Amersham Biosciences; 3000 Ci/mmol) and purified through a G-50 spin column (Amersham Biosciences). The DNA was then mixed with sonicated pBluescript SK-vector (5  $\mu$ g; Stratagene), poly(dA)-poly(dT) (5  $\mu$ g) in 200  $\mu$ L of 4  $\times$  SSC, 0.1% SDS. The samples, after heat treatment (95°C for 5 min) and incubation (65°C for 30 min), were mixed with 14 mL of hybridization buffer (0.5 M Na<sub>2</sub>HPO<sub>4</sub>, pH 7.2, 1 mM EDTA, 7% SDS). The hybridization buffer was divided into two lots (7 mL each), with each lot being used for the hybridization (65°C for 16 h) of two membranes of EST macro array (8  $\times$  12 cm). After hybridization, the filters were washed twice with 40 mM Na<sub>2</sub>HPO<sub>4</sub> (pH 7.2), 1 mM EDTA, 1% SDS at 65°C for 15 min, and once with 0.1  $\times$  SSC, 0.2% SDS at 65°C for 30 min. The filters were exposed to an imaging plate (BAS2000; Fuji Film) for the detection of hybridized signals.

### ACKNOWLEDGMENTS

We thank Dr. N. Sakurai and Dr. D. Shibata for providing the *A. thaliana* genomic DNA and Mr. C. Kuwata for providing the EST macroarray. This work was supported by the Kazusa DNA Research Institute Foundation.

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

- The *Arabidopsis* Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.
- Asamizu, E., Nakamura, Y., Sato, S., and Tabata, S. 2000. A large scale analysis of cDNA in *Arabidopsis thaliana*: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries. *DNA Res.* **7**: 175–180.
- Bergelson, J., Stahl, E., Dudek, S., and Kreitman, M. 1998. Genetic variation within and among populations of *Arabidopsis thaliana*. *Genetics* **148**: 1311–1323.
- Chang, C., Bowman, J.L., DeJohn, A.W., Lander, E.S., and Meyerowitz, E.M. 1988. Restriction fragment length polymorphism linkage map for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci.* **85**: 6856–6860.
- Inoue, S., Kiyama, R., and Oishi, M. 1996. Construction of highly extensive polymorphic DNA libraries by in-gel competitive reassociation procedure. *Genomics* **31**: 271–276.
- Lisitsyn, N. and Wigler, M. 1993. Cloning the differences between two complex genomes. *Science* **259**: 946–951.
- Rosenberg, M., Przybylska, M., and Straus, D. 1994. "RFLP subtraction": A method for making libraries of polymorphic markers. *Proc. Natl. Acad. Sci.* **91**: 6113–6117.

Received May 6, 2002; accepted in revised form December 6, 2002.