Nucleic Acids Research, 1998, Vol. 26, No. 21 **4901–4909**

Cloning and stable maintenance of DNA fragments over 300 kb in *Escherichia coli* with conventional plasmid-based vectors

Quanzhou Tao and Hong-Bin Zhang*

Department of Soil and Crop Sciences and Crop Biotechnology Center, Texas A&M University, College Station, TX 77843-2123, USA

Received July 8, 1998; Revised and Accepted September 11, 1998

ABSTRACT

Bacterial artificial chromosome (BAC) and P1-derived artificial chromosome (PAC) systems were previously developed for cloning of very large eukaryotic DNA fragments in bacteria. We report the feasibility of cloning very large fragments of eukaryotic DNA in bacteria using conventional plasmid-based vectors. One conventional plasmid vector (pGEM11), one conventional binary plasmid vector (pSLJ1711) and one conventional binary cosmid vector (pCLD04541) were investigated using the widely used BAC (pBelo-BAC11 and pECBAC1) and BIBAC (BIBAC2) vectors as controls. The plasmid vector pGEM11 yielded clones ranging in insert sizes from 40 to 100 kb, whereas the two binary vectors pCLD04541 and pSLJ1711 yielded clones ranging in insert sizes from 40 to 310 kb. Analysis of the pCLD04541 and pSLJ1711 clones indicated that they had insert sizes and stabilities similar to the BACs and BIBACs. Our findings indicate that conventional plasmid-based vectors are capable of cloning and stably maintaining DNA fragments as large as BACs and PACs in bacteria. These results suggest that many existing plasmid-based vectors, including plant and animal transformation and expression binary vectors, could be directly used for cloning of very large eukaryotic DNA fragments. The pCLD04541 and pSLJ1711 clones were shown to be present at at least 4–5 copies/cell. The high stability of these clones indicates that stability of clones does not seem contingent on single-copy status. The insert sizes and the copy numbers of the pCLD04541 and pSLJ1711 clones indicate that *Escherichia coli* can stably maintain at least 1200 kb of foreign DNA per cell. These results provide a new conceptual and theoretical basis for development of improved and new vectors for large DNA fragment cloning and transformation. According to this discovery, we have established a system for large DNA fragment cloning in bacteria using the two binary vectors, with which several very large-insert DNA libraries have been developed.

INTRODUCTION

The ability to clone large (>100 kb) DNA fragments is crucial to physical mapping, positional cloning and molecular analysis of complex eukaryotic genomes. The first large DNA fragment cloning system was reported in 1987 by Burke *et al*. (1). This system was based on yeast artificial chromosomes (YACs) and allowed cloning and maintenance of DNA fragments up to 1000 kb in yeast (1), a quantum leap relative to cosmids (40–50 kb). Because of its large DNA fragment cloning capacity, the YAC system was quickly adopted for research on the genomes of humans and other species (for example 2–5). However, several difficulties of YACs have limited the utility of YAC libraries, including their high level of chimerism (6), occasional insert instability (6) and difficulty of purifying cloned insert DNA, which tends to be contaminated with yeast host chromosomal DNA (7).

Alternative systems using bacteria as the hosts were soon developed. In 1992, Shizuya *et al*. (8) reported large DNA fragment cloning in *Escherichia coli* using a bacterial artificial chromosome (BAC) system based on the *E.coli* fertility (F-factor) plasmid. Two years later, Ioannou *et al*. (9) reported similar facts using a P1-derived artificial chromosome (PAC) system, which combined the features of the bacteriophage P1 (10) and the F-factor-based BAC (8) cloning systems. Both BAC and PAC are capable of cloning and stably maintaining DNA fragments >300 kb in *E.coli*. While the insert sizes of BACs and PACs are somewhat smaller than YACs, they have several major advantages over YACs, including their low levels of chimerism, facility and speed of insert DNA purification and high stability in the host cells. Therefore, BACs and PACs have quickly assumed a central position in genome research. BAC and PAC libraries have been developed for dozens of plants, animals and humans (for example 11–14; http://hbz.tamu.edu).

While clearly powerful, BAC and PAC vectors have a number of limitations. The most significant among these is their limited variety and diversity. Although significant efforts have been made to develop BAC vectors to meet the different needs of genomics research, only two vectors, pBeloBAC11 (13) and its derivative pECBAC1 (14), have been so far developed for large-insert BAC library development and only one binary vector, BIBAC2 (15), has been developed for large-insert DNA library

*To whom correspondence should be addressed. Tel: +1 409 862 2244; Fax: +1 409 862 4790; Email: hbz7049@pop.tamu.edu

development and plant transformation. Moreover, all of the currently publicly used BAC and BIBAC vectors are derived from the first BAC cloning vector, pBAC108L (8; below). Furthermore, because just one or two copies of BACs are present per cell, the color of BAC colonies is relatively light, which is not well-suited for robot-assisted assembly of ordered libraries. The one to two copies of BACs and PACs per cell also limits the yield of insert DNA and thus, a large number and a large scale reproduction of cloned DNA fragments.

It was shown that the first BAC vector, pBAC108L, is capable of cloning and stably maintaining DNA fragments >300 kb in *E.coli*, but lacks selectable marker for recombinants (8). Thus, the recombinant clones in pBAC108L must be identified by colony hybridization, which is not well-suited for library development. The two widely used BAC vectors, pBeloBAC11 (13) and pECBAC1 (14), are derived from pBAC108L, in which the *lacZ* gene is introduced for recombinant selection. pBeloBAC11 was derived from pBAC108L by replacing the cloning site region of pBAC108L with a DNA fragment from pGEM3Z (Promega, Madison, WI) containing the *lacZ* gene and cloning sites for colony color selection, blue for non-recombinants and colorless for recombinants. pBeloBAC11 has two *Eco*RI sites, one in the *lacZ* gene and the other in the transformant selection marker chloramphenicol (CM^r) resistance gene. To facilitate cloning of large DNA fragments generated with *Eco*RI, pBeloBAC11 was further modified into pECBAC1 by destroying the *Eco*RI site in the chloramphenicol resistance gene of pBeloBAC11 (13). Therefore, pECBAC1 is capable of cloning DNA fragments generated with *Eco*RI, in addition to those generated with *Hin*dIII and *Bam*HI as pBeloBAC11 (14).

In species with large, complex genomes, many traits are controlled by clusters of genes (e.g. the genes for most plant disease resistances; 16–17). Transformation of such gene clusters will be most efficient using a binary vector that is capable of cloning and transforming very large DNA fragments. Such a large DNA fragment cloning and plant binary transformation system will also facilitate positional cloning of genes in plants if the major purpose of the research is to isolate DNA fragments carrying genes of interest. To address this need in plants, a binary BAC vector (BIBAC2) was developed based on pBAC108L for *Agrobacterium*-mediated plant transformation. The BIBAC2 vector was shown to transform tobacco with a 150 kb human DNA fragment (15). However, the BIBAC2 has a single cloning site, *Bam*HI, and the *sacB* gene for recombinant selection. In our hands, ~10% of the clones selected were found to be false recombinants (unpublished results). These limitations have influenced the utility of the BIBAC2 vector for large-insert binary BAC library development.

It is apparent from the above efforts on the development of currently used BAC and binary BAC vectors that there are several theoretical questions in large DNA fragment cloning to be answered which are extremely important for current genomics research. First, all currently used large DNA fragment cloning systems, YAC (1), BAC (8) and PAC (9), are based on artificial chromosome systems, which seems that the construction of artificial chromosomes is essential for large DNA fragment cloning. This concept has seriously limited the development of large DNA fragment cloning vectors as illustrated in the above efforts and thus, development and applications of large-insert DNA libraries in genomics research. The question is if non-artificial chromosome systems, such as conventional plasmids and conventional cosmids that have been already widely used in molecular genetics and molecular biology, can be used for large DNA fragment cloning as the artificial chromosome systems, i.e. BAC and PAC. If the answer is yes, an unlimited number and variety of plasmid-based vectors developed previously for different biological research needs, including binary and expression vectors that are widely used in plant and animal molecular biology, could be directly used for large-insert DNA library development and plant and animal transformation. Development of new/improved vectors, especially binary and expression vectors, for large DNA fragment cloning and biotechnological systems for gene identification, cloning, analysis, transformation and expression would be significantly expedited. Second, many studies (for example 8,9,11–14) have demonstrated that BACs and PACs are stable in their bacterial hosts and suggested that the stability of the BACs and PACs is due to their unique copy per cell (8,9,13). The concept on the relationship between the stability of large-insert BAC and PAC clones and their unique copy per cell has been widely accepted at the present time (11–14,18), however, it has not been experimentally demonstrated yet. Studies are needed to answer this question because this concept is significantly influencing the development of new vectors for large DNA fragment cloning and engineering in plants and animals and production of cloned eukaryotic DNA fragments in bacteria for genomics research. Third, dozens of BAC and PAC libraries (11–14) have been developed since the BAC and PAC systems were established (8,9). All existing BAC and PAC libraries have average insert sizes of ~150 kb and among these libraries the largest BAC or PAC in insert size that has been observed so far is ~350 kb (8,9,11–14). The question is why no BAC or PAC with an insert size >500 kb has been observed as YACs (1,5). Is this due to the limited capacity of the BAC and PAC host, *E.coli*, in stable maintenance of foreign DNA fragments or the limitation of the currently used large DNA fragment cloning techniques? This is the first question needing to be answered for further improvement of current cloning capacity of bacterial cloning systems such as BAC and PAC. Efficacy of genomics research would leap forward if the average insert size of stable large-insert bacterial clones could be increased to 300–500 kb.

In this report, we show that conventional plasmid-based vectors allow cloning and stable maintenance of DNA fragments >300 kb in *E.coli*. The bacterium *E.coli* strain DH10B is able to stably maintain at least 1200 kb of foreign DNA fragments in the form of plasmids in a single cell. We also show that the stability of large-insert clones does not seem contingent on the single copy status of the clones. These results urge a re-evaluation of the dogma for large DNA fragment cloning and suggest a new conceptual and theoretical basis for large DNA fragment cloning and plant and animal engineering beyond current BACs and PACs.

MATERIALS AND METHODS

Megabase DNA preparation

Sorghum bicolor (L.) Moench BTX623 seedlings were grown in a greenhouse and used for DNA preparation. Megabase sorghum DNA was isolated according to Zhang *et al.* (19). Sorghum nuclei were embedded in low melting point agarose microbeads and the DNA was purified in the microbeads.

Large DNA fragment cloning

One conventional plasmid vector, pGEM11 (Promega), one conventional binary plasmid vector, pSLJ1711 (20), and one conventional binary cosmid vector, pCLD04541 (20,21), were used in the experiments. pGEM11 is a widely used commercial plasmid vector for genomic DNA cloning, whereas pSLJ1711 and pCLD04541 are conventional binary plasmid and cosmid vectors previously designed for *Agrobacterium*-mediated plant transformation (20,21). pCLD04541 and pSLJ1711 are ~29 and 27 kb in size, respectively, and derived from the plasmid vector pRK290 of 20 kb, which is a derivative of a native bacterial plasmid, RK2 (22). RK2 belongs to the P1 incompatibility group and has a size of 56 kb. The genes *oriV*, *trfA* and *trfB* constitute the replicons of RK2 and its derivative, pRK290, which exist at 5–8 copies/chromosomal equivalent in *E.coli* (22,23). pGEM11 is ~3 kb in size, has the replicon of ColE1 and exists at 30–40 copies/chromosomal equivalent in *E.coli* (Promega). We selected the binary cosmid and plasmid vectors pCLD04541 and pSLJ1711, because such binary vectors could be directly used for large-insert DNA library development and direct plant transformation via *Agrobacterium*. The currently used BAC and BIBAC cloning vectors, pBeloBAC11 (13), pECBAC1 (14), and BIBAC2 (15), were used as controls. The development and features of these vectors were described in detail in previous studies (8,9,13–15,24).

Vector preparation and large DNA fragment cloning were done as previously described by Zhang *et al.* (12,18) and Tao *et al.* (11) with modifications. The partially digested DNA fragments were size-selected on 1% pulsed-field low melting point agarose gels in 0.5× TBE (45 mM Trizma base, 45 mM boric acid and 1 mM EDTA, pH 8.3). The molar ratio of the vectors to the pulsed-field gel-selected sorghum DNA for ligation was 3:1 for pCLD04541, pSLJ1711 and BIBAC2 and 5:1 for pGEM11, pBeloBAC11 and pECBAC1. The ligated DNA was transformed into *E.coli* ElectroMAX DH10B cells (Gibco BRL, Grand Island, NY) by electroporation using the Cell Porator and Voltage Booster system (Gibco BRL). The Cell Porator settings were 350 V, 330 μF capacitance, low ohm impedance and fast charge rate, and the Voltage Booster was adjusted by setting the resistance to 4000 Ω.

Clone insert size analysis

Colorless clones (potential recombinants) were inoculated in 5 ml of LB broth containing suitable antibiotics and grown at 37°C with shaking at 250 r.p.m. overnight. The cells were harvested and DNA was isolated with the alkaline lysis method as described by Zhang *et al.* (12). DNA was digested with *Not*I and subjected to pulsed-field gel electrophoresis on a 1% agarose gel in 0.5× TBE and at an initial pulse time of 5 s, a final pulse time of 15 s, 120°, 6 V/cm and 10.5°C for 16 h. The gel was stained with ethidium bromide, destained in water for 30 min and photographed. The insert size of each clone was determined by adding up all insert DNA bands. The bands appearing in all lanes were from the cloning vectors.

Clone insert stability

Nine clones of pSLJ1711 and 20 clones of pCLD04541 with insert sizes from 100 to 300 kb were randomly selected and analyzed as described in the analysis of BAC and PAC insert stability (8,9). Individual clones were inoculated in 5 ml of LB broth containing 15 mg/l tetracycline. After growth at 37°C, 250 r.p.m. for 24 h, the cells of each clone were continuously inoculated into 5 ml of fresh LB broth containing 15 mg/l tetracycline with a toothpick and grown, and the remaining cells were harvested. The cells of each clone continued to grow for an additional 4 days with a change of fresh medium every 24 h as above. After 5 days of growth, the cells of each clone were harvested. DNA was isolated from the cells by the alkaline lysis method as described by Zhang *et al.* (12). The DNAs of each clone isolated after 1 and 5 days growth were: (i) digested with *Hin*dIII and subjected to 0.8% agarose gel electrophoresis; (ii) the insert released with *Not*I and subjected to 1% pulsed-field gel electrophoresis; (iii) directly analyzed on a 1% pulsed-field gel. The gels were stained with ethidium bromide, destained in water for 30 min and photographed.

Cloned DNA fragment fingerprinting

Random pCLD04541 maize DNA clones with an average insert size of 152 kb were used to test the feasibility of pCLD04541 clones for physical mapping of genomes by fingerprint analysis. BACs randomly selected from our maize (S.Santos *et al.*, in preparation), rice (12) and apple (B.Vinatzer *et al.*, submitted for publication) libraries and BIBACs from our tomato libraries (C.Hamilton *et al.*, in preparation) were used as controls. The maize DNA clones in pCLD04541 were developed according to the procedure described above in this study (unpublished results). The DNAs of BAC, BIBAC and pCLD04541 clones were purified according to our procedure (Q.Tao, Y.-L.Chang and H.-B.Zhang, submitted for publication) that was developed for rapid development of physical maps of the rice (Q.Tao *et al.*, unpublished results) and *Arabidopsis thaliana* (Y.-L. Chang *et al.*, unpublished results) genomes with large-insert bacterial clones. The DNAs were fingerprinted using our previously developed large-insert bacterial clone fingerprinting kit (the Fpase kit) (H.-B.Zhang and Q.Tao, Invention no. TAMUS1228). Briefly, the DNAs were mixed with the enzyme Fpase I, the Fpase I buffer and [^{32}P]dATP and incubated at 37°C for 2 h. The reactions were stopped with the DNA sequencing gel loading dye (98% v/v deionized formamide, 0.3% bromophenol blue, 0.3% xylene cyanol and 10 mM EDTA, pH 8.0), denatured at 95°C for 5 min and subjected to 4% denaturing polyacrylamide gel electrophoresis. The gel was dried onto 3MM blotting paper and autoradiographed to X-ray film.

RESULTS

Several experiments were conducted to test the cloning capacity of conventional plasmid-based vectors. The existing conventional plasmid vector, pGEM11 (Promega), and binary plasmid and cosmid vectors, pSLJ1711 (20) and pCLD04541 (20,21), were randomly selected and used. The currently used BAC and BIBAC cloning vectors, pBeloBAC11 (13), pECBAC1 (14) and BIBAC2 (15), were used as controls. Megabase DNA isolated from sorghum was used in the experiments. The sorghum megabase DNA was partially digested with *Bam*HI and *Eco*RI, respectively, size-selected and ligated into the *Bam*HI sites of pBeloBAC11, BIBAC2, pGEM11 and pCLD04541 and the *Eco*RI sites of pECBAC1, pCLD04541 and pSLJ1711. The ligations were transformed into *E.coli* by electroporation. From a single transformation with 0.05–1.0 ng of the ligated DNA,

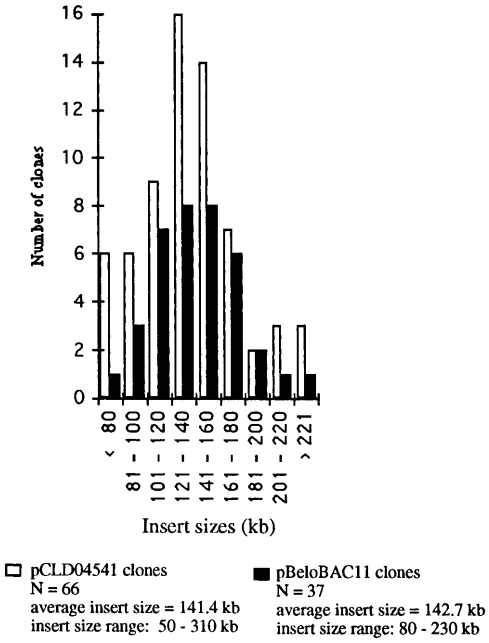

Figure 1. Distributions of the insert sizes of pCLD04541 and pBeloBAC11 clones. Colorless clones on selectable medium were randomly selected, inoculated in 5 ml LB broth containing suitable antibiotics and grown at 37°C, 250 r.p.m. overnight. DNA was isolated, digested with *Not*I, run on a 1% pulsed-field gel, stained with ethidium bromide and photographed. The insert size of each clone was obtained by adding up all bands of sorghum genomic DNA (for example Fig. 2).

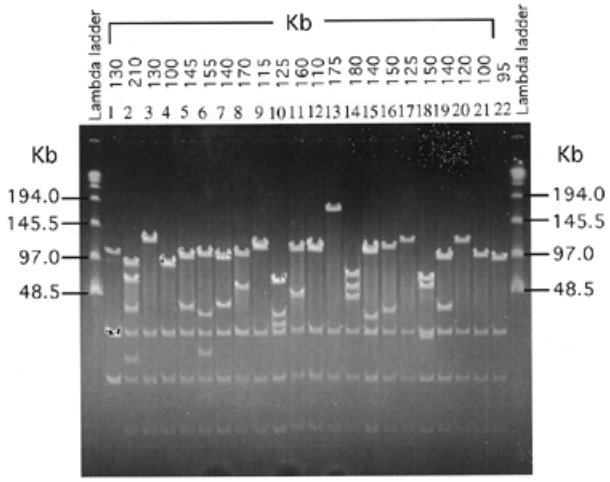

Figure 2. Large-insert sorghum DNA clones cloned in the conventional plant binary cosmid vector pCLD04541. The clone DNA was isolated, digested with *Not*I and subjected to pulsed-field gel electrophoresis. The three bands appearing in all lanes are from the cloning vector pCLD04541 and the remaining bands are from the insert DNA. The insert sizes of the pCLD04541 clones range from 95 to 210 kb, with an average insert size of ~140 kb.

100–1000 recombinant clones were obtained from the ligates of pGEM11, pBeloBAC11, pECBAC1, BIBAC2, pSLJ1711 and pCLD04541 (for example Fig. 6 below).

The insert sizes of five clones of pGEM11 and 20–160 clones of pBeloBAC11, pECBAC1, BIBAC2, pSLJ1711 and pCLD04541 were analyzed by pulsed-field gel electrophoresis. The results showed that the clones of pGEM11 have an insert size range from 40 to 100 kb (not shown) and the clones of pBeloBAC11, pECBAC1, BIBAC2, pSLJ1711 and pCLD04541 have insert size ranges from 40 to 310 kb. Figure 1 shows the distributions of the insert sizes of pCLD04541 clones and pBeloBAC11 clones. A total of 66 random pCLD04541 clones were analyzed for insert size. The insert sizes of these pCLD04541 clones ranged from 50 to 310 kb with an average insert size of 141.4 kb. For comparison, a total of 37 random pBeloBAC11 clones that were developed from the same size-selected sorghum DNA fragments at the same time as the pCLD04541 clones were also analyzed for insert size. The insert sizes of these control pBeloBAC11 clones ranged from 80 to 230 kb with an average insert size of 142.7 kb. Statistical analysis showed that the average insert size (141.4 kb) of pCLD04541 clones had no significant difference from that (142.7 kb) of the pBeloBAC11 clones. Similar results were also obtained for the comparative studies of the insert sizes of pSLJ1711 and pCLD04541 clones with those of the pBeloBAC11, pECBAC1 and BIBAC2 clones (not shown). Figure 2 shows 22 clones of pCLD04541 digested with *Not*I, the insert sizes of which range from 95 to 210 kb. These results strongly support our hypothesis that conventional plasmid-based vectors have the DNA cloning capacity in size as large as the artificial chromosome (BAC) cloning vectors in bacteria.

To test clone insert stability, nine clones of pSLJ1711 and 20 clones of pCLD04541 with an insert size range from 100 to 300 kb were further analyzed by fingerprinting according to the procedure used for the analysis of BAC and PAC insert stability (8,9). The results showed that none of the 29 clones of pSLJ1711 and pCLD04541 changed in DNA fingerprint patterns after 5 days of continuous growth relative to 1 days growth. Figure 3 shows the DNA fingerprints of five pCLD04541 clones and nine pSLJ1711 clones with 1 versus 5 days growth. These results clearly show that the plasmid-based binary vectors, pSLJ1711 and pCLD04541, are capable of stably maintaining DNA fragments of at least 300 kb in bacterial cells.

Previous studies (25) reported the instability of multicopy cosmid clones of Lawrist 16 and Supercos (Stratagene, La Jolla, CA) in *E.coli* strain DH5MCR (Gibco BRL), which disagreed with our results obtained in this study although the same approach was used to test the clone stability. Therefore, we further studied the stability of the large-insert pCLD04541 and pSLJ1711 clones carefully. Nine of the pCLD04541 clones shown in Figure 2, with insert sizes ranging from 115 to 210 kb, were inoculated and continuously grown for 5 days as described in Materials and Methods. The DNAs isolated from 1 and 5 day culture were analyzed in two ways: (i) digesting the DNAs with *Not*I to release the inserts and pair-wise running on a pulsed-field gel (Fig. 4A); (ii) directly running the undigested DNAs on a pulsed-field gel (Fig. 4B). This is because if the clones were unstable, both deleted and intact clones would be present in the cell populations after 5 days continuous growth and could be straightforwardly detected by analysis of the undigested or linearized clones. The results of this experiment showed that the pulsed-field gel analyzed patterns of the DNAs isolated from 1 and 5 day culture are identical; no detectable deletion was observed in these pCLD04541 clones (Fig. 4). These results have further demonstrated that the pCLD04541 and pSLJ1711 vectors are capable of stably

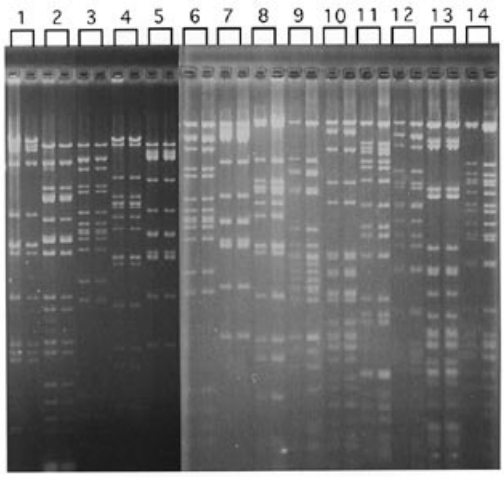

Figure 3. DNA fingerprints of pCLD04541 and pSLJ1711 clones after growth for 1 versus 5 days. Lane pairs 1–5 are for the DNA fingerprints of pCLD04541 clones and lane pairs 6–14 for the DNA fingerprints of pSLJ1711 clones. The insert sizes of these clones range from 100 to 300 kb. DNAs were isolated from the cells cultured for 1 and 5 days, respectively, digested with *Hin*dIII and subjected to 0.8% agarose gel electrophoresis. Note the identity of the paired clone fingerprints. The left lane of each lane pair is for the DNA isolated from the cells cultured for 1 day and the right lane of each lane pair is for the DNA isolated from the cells continuously cultured for 5 days.

maintaining foreign DNA fragments of at least 300 kb in bacterial cells.

Previous studies demonstrated that pCLD04541 and pSLJ1711 exist at 5–8 copies/chromosomal equivalent in *E.coli* (20–23). To test if this remains for the large-insert clones of these two vectors, four random clones of pCLD04541 in *E.coli* strain DH10B with an insert size range from 110 to 200 kb were used, with five random clones of pBeloBAC11 in the same strain of *E.coli* with an insert size range from 130 to 160 kb as controls. Each of these clones was inoculated in equal amounts (5 ml) of LB broth with 15 mg/l tetracycline for pCLD04541 clones and 12.5 mg/l chloramphenicol for pBeloBAC11 clones and grown at 37°C, 250 r.p.m. overnight. The DNAs of the clones were isolated according to Zhang *et al*. (12), dissolved in equal amounts (40 µl) of TE (10 mM Tris–HCl, pH 8.0, 1 mM EDTA, pH 8.0) and were analyzed for the yield of DNA by pulsed-field gel electrophoresis and with a fluorometer. The DNAs were digested with *Not*I, run on a 1% pulsed-field gel, stained with ethidium bromide, destained in water and photographed (Fig. 5). The results of gel analysis showed that the DNA yields of the pCLD04541 clones were estimated at 4- to 5-fold higher than those of pBeloBAC11 clones. Similar results were also observed for pSLJ1711 clones versus pBeloBAC11 clones. The DNA concentrations of the pCLD04541 and pBeloBAC11 clones were also measured with a DyNA Quant 200 Fluorometer (Hoefer) using the DNA-specific binding dye Hoechst 33258 (Sigma, St Louis, MO) to stain the DNAs. The result showed that the mean DNA concentration of the four pCLD04541 clones was 102 µg/ml and that of the five pBeloBAC11 clones was 31 µg/ml, the concentration of the pCLD04541 clones being 3.3-fold higher than that of the pBeloBAC11 clones. This result is consistent with those obtained by the above pulsed-field gel analysis if contamination of the clone DNA with host chromosomal DNA and the insert size of the clones were considered. The comparative analysis of the DNA yields between pCLD04541 clones and pBeloBAC11 clones in the same bacterial strain (DH10B) indicates that the large-insert clones of pCLD04541 are present at at least 4–5 copies/cell if the pBeloBAC11 clones are present at a single copy per cell (8,13).

The color development of pCLD04541 and pSLJ1711 clones was also observed using the pECBAC1 and pBeloBAC11 clones as controls. The pCLD04541 and pECBAC1 clones were constructed from the same size-selected DNA fragments at the same time. The ligations of the two types of clones were transformed into the same *E.coli* strain DH10B by electroporation, grown in SOC medium (26) for 1 h and then plated on LB agar medium containing 60 mg/l X-gal (5-bromo-4-chloro-3-indolyl-β-D-galactoside), 15 mg/l

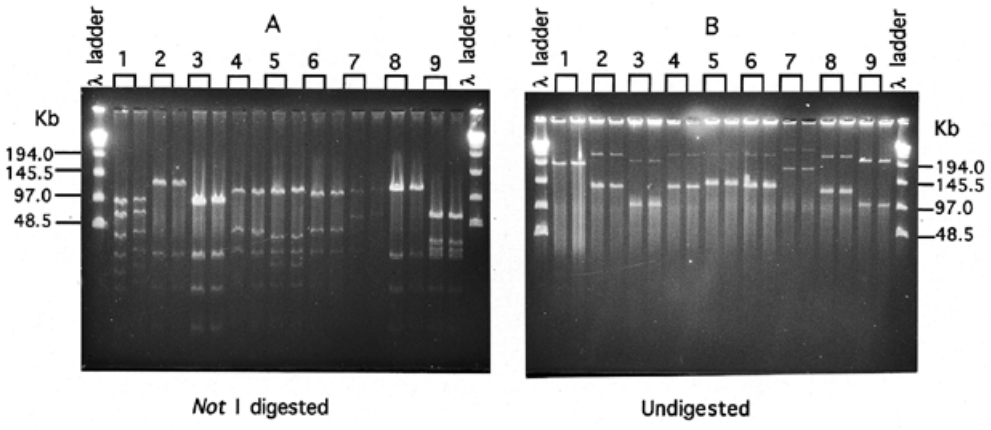

Figure 4. The *Not*I digested DNAs (**A**) and undigested DNAs (**B**) of pCLD04541 clones after growth for 1 versus 5 days. The insert sizes of these clones range from 115 to 210 kb (Fig. 2). (A) As described in Figure 3 DNAs were isolated from the cells cultured for 1 and 5 days, respectively, digested with *Not*I to release the inserts of the clones (note that there is no enzyme available to linearize the plasmids) and subjected to 1% pulsed-field agarose gel electrophoresis. (B) The undigested DNAs of the clones were directly subjected to 1% pulsed-field agarose gel electrophoresis. Note that the top band of lane pair 1 of (B) remained in the wells, which was revealed by further analysis on a separate gel (not shown). The left lane of each lane pair is for the DNA isolated from the cells cultured for 1 day and the right lane of each lane pair is for the DNA isolated from the cells continuously cultured for 5 days. As the results shown in Figure 3, the DNA patterns of each lane pair are identical, which has further demonstrated the stability of these large-insert clones.

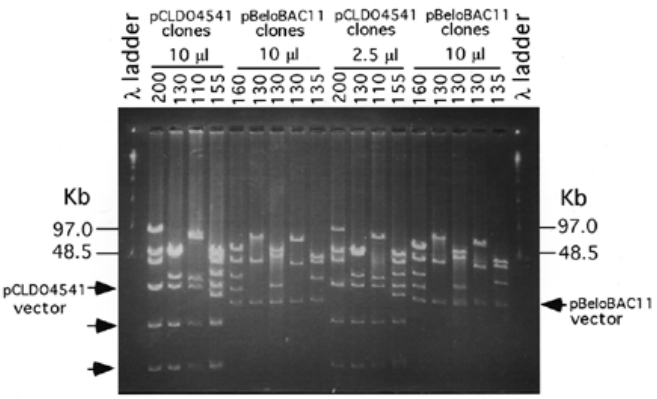

Figure 5. DNA yields of pCLD04541 clones relative to those of pBeloBAC11 clones. A single colony of each clone with similar colony size was inoculated in an equal amount (5 ml) of LB broth with 15 mg/l tetracycline for pCLD04541 clones and 12.5 mg/l chloramphenicol for pBeloBAC11 clones and grown at 37°C, 250 r.p.m. overnight. DNA was isolated and dissolved in an equal amount (40 µl) of TE. An equal amount (10 µl) (lanes 2–10 from left) and different amounts (2.5 versus 10 µl) (lanes 11–19 from left) of the 40 µl DNA of the pCLD04541 clones and the pBeloBAC11 clones were digested with *Not*I, respectively, analyzed on a 1% pulsed-field agarose gel and stained with ethidium bromide. Note the staining intensities of the DNA fragments between pCLD04541 clones and pBeloBAC11 clones. The pCLD04541 and pBeloBAC11 clones are both in *E.coli* strain DH10B.

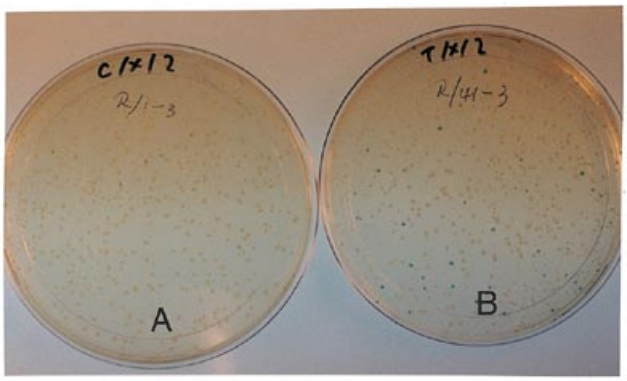

Figure 6. Colony colors of pECBAC1 clones (**A**) and pCLD04541 clones (**B**). The ligations of pECBAC1 and pCLD04541 with the same section of size-selected sorghum DNA fragments were transformed. The transformed cells were plated on LB agar medium containing suitable antibiotics and the same amount of X-gal and IPTG for colony color selection and incubated at 37°C for 24 h. The white (colorless) colonies indicate potential recombinant clones with sorghum DNA inserts (BACs or cosmids) and the blue colonies indicate non-recombinant clones without sorghum DNA inserts. Note that the colony colors of non-recombinant clones of pCLD04541 (B) are much bluer than those of pECBAC1 (A). The pCLD04541 and pECBAC1 clones are both in *E.coli* strain DH10B.

IPTG (isopropylthio-β-D-galactoside) and 15 mg/l tetracycline for pCLD04541 clones and 12.5 mg/l chloramphenicol for pECBAC1 clones. The cells were incubated at 37°C for 24 h. The result is shown in Figure 6, in which the colony color of non-recombinant pCLD04541 clones is much bluer than that of non-recombinant pECBAC1 clones. Similar results were also observed in the colony color difference of non-recombinant pCLD04541 and pSLJ1711 clones from those of non-recombinant pBeloBAC11 and pECBAC1 clones. These results have further confirmed the above copy number analysis of the pCLD04541 and pSLJ1711 clones, assuming that the *lacZ* gene is expressed at the same level in the non-recombinant plasmid-based clones as in the non-recombinant BAC clones in strain DH10B. Importantly, the significant difference in the colony colors between the plasmid-based and BAC non-recombinant clones indicated that the clones of pCLD04541 and pSLJ1711 could be much better suited for robot-assisted assembly of DNA libraries than those of pECBAC1 and pBeloBAC11 clones.

To test the feasibility of the pCLD04541 clones for physical mapping of complex genomes by fingerprint analysis, 64 random maize DNA pCLD04541 clones were fingerprinted according to the procedure that we have developed and used in development of physical maps of the rice and *A.thaliana* genomes (unpublished results), using the tomato BIBAC2, apple pECBAC1 and maize pBeloBAC11 clones as controls. The results are shown in Figure 7. The fingerprint bands of the pCLD04541 clones were clear and sharp, similar to those of the pBeloBAC11, pECBAC1 and BIBAC2 clones; however, no band derived from the vector was observed in the fingerprints with a fragment size range from 10 to 3000 bp for the pCLD04541 clones. In comparison, at least four vector bands of BIBAC2 clones and one to two (depending on the cloning site used) vector bands of pBeloBAC11 and pECBAC1 clones were observed in the fingerprints of the same size range. This result indicates that the pCLD04541 clones are better feasible for fingerprint analysis, in addition to its better facility and speed of cloned DNA purification, than the BAC and BIBAC clones. This feature of the pCLD04541 clones is of particular importance for physical mapping of genomes by fingerprint analysis because the vector bands have to be manually deleted one after one during editing before the fingerprints of large-insert clones are used for contig assembly with the computer package FPC (27).

DISCUSSION

This study has first demonstrated that conventional plasmid-based vectors, such as plasmids and cosmids, are capable of cloning and stably maintaining DNA fragments as large as artificial chromosome vectors, namely BACs and PACs. This finding suggests that many existing plasmid-based vectors, such as plasmids and cosmids, could be directly used for large DNA fragment cloning and transformation in genomics research, as well as templates for improved and/or new vector development for large-insert library generation and plant and animal transformation.

The successful cloning of large DNA fragments in the conventional plasmid-based vectors in this study is largely attributable to the use of electroporation technology in bacterial cell transformation. This suggests that much of the increased DNA cloning capacities from plasmid (~20 kb), cosmid (~50 kb) and bacteriophage P1(~100 kb) (10) to BAC and PAC (~300 kb) resulted from the improvement of transformation technology. Before the advent of the electroporation transformation technology, plasmid constructs were transformed into bacterial cells by DNA uptake through CaCl-treated competent cells (26) and cosmid (26) and bacteriophage P1 constructs (10) were transformed into bacterial cells through *in vitro* bacteriophage λ particle head packaging, both of which have their limitations in transferring large (>100 kb) DNA fragments into bacterial cells (10,26). The application of electroporation technology in transformation has

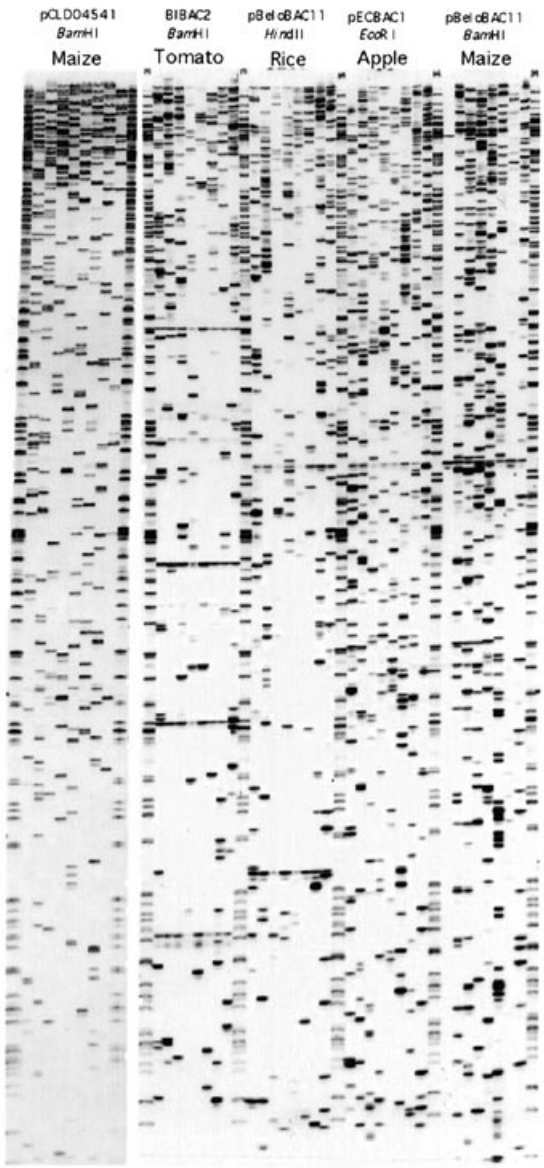

Figure 7. Fingerprints of pCLD04541, BIBAC2, pBeloBAC11 and pECBAC1 clones. From left, lanes 1, 10, 11, 20, 29, 38 and 47 are λ/*Sau*3AI DNA markers and the remaining lanes are plant DNA clones. Eight random clones were fingerprinted for each combination of vector and cloning enzyme used. The bands appearing in all clone lanes are derived from the cloning vectors. Note that no vector band is observed in the fingerprints of pCLD04541 clones, whereas 1–4 vector bands are observed in the fingerprints of BIBAC2, pBeloBAC11 and pECBAC1 clones.

significantly increased the capacity of transforming large DNA fragments into bacterial cells. Our Texas A&M BAC Center has recently constructed 27 very large-insert BAC and cosmid libraries from different plant and animal species (http://hbz.tamu.edu). The collective results indicate that it is difficult to transform DNA fragments >300 kb into *E.coli* by electroporation. Therefore, an improvement in current bacterial transformation technologies is needed to significantly increase the sizes of DNA fragments cloned in bacterial cells with the existing artificial chromosome (e.g. BAC and PAC) or conventional plasmid-based vectors.

The native plasmid RK2 is 56 kb in size and the two binary vectors derived from it, pSLJ1711 and pCLD04541, are 27.4 and 29.1 kb in size, respectively. Both vectors exist at 5–8 copies/chromosomal equivalent in bacterial cells (20–23) and allow cloning and stable maintenance of DNA fragments of at least 300 kb (this study). Comparative analysis showed that the amounts of clonal DNA isolated from 5 ml cultures of pCLD04541 and pSLJ1711 clones in *E.coli* strain DH10B were 4- to 5-fold higher than those of pBeloBAC11, pECBAC1 and BIBAC2 clones isolated from the same amount of culture in the same bacterial strain (Fig. 5). This result is consistent with the 5–8 copies/cell of pCLD04541 and pSLJ1711 if the BACs and BIBACs are present at 2 copies/cell. The result indicates that the large-insert pSLJ1711 and pCLD04541 clones are present at 4–5 copies/bacterial cell if the BACs and BIBACs are present at a single copy per cell. Moreover, since both pCLD04541 and pSLJ1711 are capable of cloning and stably maintaining fragments of eukaryotic DNA >300 kb (this study), *E.coli* can collectively maintain at least 1200–1500 kb of foreign DNA per cell, i.e. ~1/3 of the *E.coli* genome size. This result suggests that it may be possible to increase the current average insert sizes of BACs and PACs (~150 kb) by 5- to 8-fold through improvement of the current bacterial transformation technologies.

This study first reports that the conventional binary cosmid and plasmid vectors, pCLD04541 and pSLJ1711, are capable of cloning fragments of eukaryotic DNA >300 kb in bacterial cells. The pSLJ1711 and pCLD04541 clones of very large inserts (100–300 kb) were as stable as the control BAC and BIBAC clones (Fig. 3). The high stability of pSLJ1711 and pCLD04541 clones seems inconsistent with the dogma that the stability of large-insert clones in BACs and PACs is due to their 1–2 copies/cell (8,9,13). If the stability of large-insert clones were contingent on such low copy number, neither multicopy pSLJ1711 nor pCLD04541 clones would have been stable.

The relationship between clone stability and copy number has not been clear to date, although a few studies have been performed in this respect. One is a study of large-insert PAC stability in bacterial strain DH10B (9), the results of which agree with our results obtained in this study. The PAC vector pCYPAC1 has two replication mechanisms, the single-copy replicon and the multicopy replicon. For PAC cloning and propagation the single-copy replicon is used; however, multicopy PACs can be induced in the presence of the *lac* inducer, IPTG. Fingerprint analysis of the induced multicopy PACs indicated that they were stable in the bacterial strain DH10B (9). Another study is that of Kim *et al.* (25), the results of which disagree with our results obtained in this study. Human genomic DNA was cloned in the multicopy cosmid vectors Lawrist 16 and Supercos (Stratagene) and the single-copy fosmid vector pFOS1, transfected into *E.coli* strain DH5MCR (Gibco BRL) and studied for their stability with the same fingerprinting approach as those used in the studies of BAC (8) and PAC (9) stability and in this study. Instability of the clones of all three vectors was detected although the fosmid clones were more stable than the cosmid clones. In comparison, the BAC vectors pBAC108L (8), pBeloBAC11 (13) and pECBAC1 (14) have the same backbone and replicon as the fosmid vector pFOS1 (25); however, no unstable BACs have been reported to date although the BACs have much larger insert sizes than the fosmid clones and the BAC stability has been

extensively studied (for example 8,11–14). A reasonable explanation of the clone stability to these studies may be due to the host strains and the cloning systems. The multicopy pCLD04541 and pSLJ1711 (this study) and the single-copy PAC (9) and BAC (8,11–14) clones were all stable in the *E.coli* strain DH10B even though they have very large insert sizes and the BACs have the same replication mechanism as the fosmids (25). Nevertheless, neither the single-copy fosmid nor the multicopy cosmids (25) were as stable in the *E.coli* strain DH5MCR as the single-copy BACs (8, 11–14) and the multicopy pCLD04541 and pSLJ1711 clones (this study) were in the *E.coli* strain DH10B, in spite of their smaller insert sizes and the same replication mechanism of the fosmid (25) as the BACs. On the other hand, the vectors used in these studies are different in replication mechanisms. The BAC (8,13,14) and fosmid (25) vectors are based on the bacterial F-factor plasmid, the PAC vector is based on the bacteriophage P1 cloning system (9), the pCLD04541 and pSLJ1711 vectors are based on the bacterial P-1 plasmid (20–23) and the Lawrist and Supercos vectors are based on the pBR cloning system (25). Further investigations are needed to understand the relationships among clone copy number, insert size, stability and host. If large DNA fragment cloning vectors present at a few copies, such as pSLJ1711 and pCLD04541, can be used or developed for large-insert DNA library development, the yield of the cloned DNA fragments could be increased by a few fold. The increased yield of the cloned DNA fragments will significantly facilitate cloned DNA reproduction and thus, the utility of large-insert bacterial clones in genome physical mapping and large-scale genome sequencing.

This study indicates that the existing plant binary cosmid and plasmid vectors, pCLD04541 and pSLJ1711, which were initially designed mainly for plant transformation through *Agrobacterium* (20,21), are well-suited for very large-insert library construction. Further investigation is needed to test if they can be used to transfer such large DNA fragments into plants by *Agrobacterium*-mediated transformation while they have been successfully used to transfer DNA fragments of ~20 kb into plants (21). If so, they could complement BIBAC2, the recently released and sole vector for large-insert library development and plant transformation (15). Furthermore, pCLD04541 and pSLJ1711 have a few additional advantages for large-insert library development over BIBAC2, pBeloBAC11 (13) and pECBAC1 (14). First, pCLD04541 and pSLJ1711 have 5–8 copies/chromosomal equivalent; therefore, the DNAs of their clones are much easier to purify than those of BIBAC2, pBeloBAC11 and pECBAC1, which are at 1–2 copies/cell (8,24; Fig. 5). Second, both of these cosmid and plasmid vectors have cloning sites for *Xho*I, *Cla*I, *Hin*dIII, *Eco*RI and *Bam*HI, which give a few more choices of restriction enzymes for large-insert DNA library construction than BIBAC2, pBeloBAC11 and pECBAC1. BIBAC2 has a single cloning site (*Bam*HI), pBeloBAC11 has two cloning sites (*Hin*dIII and *Bam*HI) and pECBAC1 has three cloning sites (*Hin*dIII, *Bam*HI and *Eco*RI). Third, pCLD04541 and pSLJ1711 have the *lacZ* gene as the selectable marker for recombinant clones (blue and white color selection), whereas the BIBAC vector has the *sacB* gene as the selectable marker for recombinants. In our hands, the *sacB* gene is not good as the *lacZ* gene for recombinant selection because the clones selected with the *sacB* gene have false recombinants at ~10% frequency. Although pBeloBAC11 and pECBAC1 have the *lacZ* gene as the recombinant selection marker, the colony color of the two vector clones cannot develop as well as that of pCLD04541 and pSLJ1711 clones (Fig. 6). Additionally, pBeloBAC11 and pECBAC1 can only be used for large-insert DNA library development, whereas pCLD04541 and pSLJ1711 are binary vectors that may also be directly used in plant transformation of large-insert clones through *Agrobacterium* (20,21).

Partial sorghum *Bam*HI and *Eco*RI libraries were developed using pCLD04541 as the vector in this study (Figs 1 and 6). The average insert size of the sorghum *Bam*HI library is 141 kb, based on the insert sizes of 66 random clones (Fig. 1), and that of the sorghum *Eco*RI library is 138 kb, based on the insert sizes of 44 random clones (not shown). Because of the success in development of the two partial sorghum DNA libraries in this study, several large-insert DNA libraries thereafter have been developed using pCLD04541 as the vector in our laboratories. These libraries are two for *A.thaliana* (K.Meksem *et al*., unpublished results; Y.-L.Chang *et al*., unpublished results), two for soybean (K.Meksem *et al*., in preparation), one for lablab (C.Liu *et al*., submitted for publication), one for *Triticum tauschii* (O.Moullet *et al*., in preparation) and one for maize (unpublished results; http://hbz.tamu.edu . Note that V41 = pCLD04541 at the Web site). The average insert sizes of these libraries range from 120 kb for the lablab library to 152 kb for the maize library, which have no significant difference in average insert sizes from the existing PAC, BAC and BIBAC libraries (8,9,11–14; http://hbz.tamu.edu).

This study has established the following new concepts for cloning of large eukaryotic DNA fragments in bacteria which are of extreme importance for current genomics research. The existing conventional plasmid-based vectors other than BACs and PACs could be well-suited for large-insert DNA library construction, although they need to be tested individually before use. The vectors present at a few copies per cell are capable of stably maintaining fragments of eukaryotic DNA as BACs and PACs present at 1–2 copies/cell. These findings will provide quick inroads into developing new large DNA fragment cloning systems, by direct usage of existing vectors, as well as using them as templates to develop new vectors. For example, this result may expedite the development of new large DNA fragment cloning vectors for not only library development and plant and animal transformation, but also site-specific recombination. This study also demonstrates that *E.coli* is able to stably maintain foreign DNA fragments at least 1200 kb/cell in the form of plasmids. This has provided a theoretical basis for significant improvement in the current insert sizes of BAC, PAC and plasmid-based vector clones.

ACKNOWLEDGEMENTS

We thank Drs J.D.G.Jones and C.Dean, H.Shizuya and M.Simon, R.W.Michelmore and C.M.Hamilton for kindly providing the pCLD04541, pSLJ1711, pBeloBAC11, pECBAC1 and BIBAC2 vectors. The authors also like to thank Dr David M.Stelly, Texas A&M University, for his critical reading of and valuable suggestions on the manuscript. The research was supported by the Rockefeller Foundation and Texas Higher Education Coordinating Board–the Advanced Technology Program (grant 999902-042) to H.-B.Z.

REFERENCES

1 Burke,D.T., Carle,G.F. and Olson,M.V. (1987) *Science*, **236**, 805–811.
2 Coulson,A., Waterston,R., Kiff,J., Sulston,J. and Kohara,Y. (1988) *Nature*, **335**, 184–186.
3 Martin,G.B., Brommonschenkel,S.H., Chunwongse,J., Frary,A., Ganal,M.W., Spivey,R., Wu,T., Earle,E.D. and Tanksley,S.D. (1993) *Science*, **252**, 1432–1438.
4 Zhang,H.-B., Martin,G.B., Tanksley,S.D. and Wing,R.A. (1994) *Mol. Gen. Genet*., **244**, 513–521.
5 Chumakov,I.M., Rigault,P., Gall,I.L., Bellanne-Chantelot,C., Billault,A., Guillou,S., Soularue,P., Guasconi,G., Poullier,E., Gros,I. *et al.* (1995) *Nature*, **377**, 157–298.
6 Anderson,C. (1993) *Science*, **259**, 1684–1687.
7 Zhang,H.-B. and Wing,R.A. (1997) *Plant Mol. Biol*., **35**, 115–127.
8 Shizuya,H., Birren,B., Kim,U.-J., Mancino,V., Slepak,T., Tachiiri,Y. and Simon,M. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 8794–8797.
9 Ioannou,P.A., Amemiya,C.T., Granes,J., Kroisel,P.M., Shyzuya,H., Chen,C., Batzer,M.A. and de Jong,P.J. (1994) *Nature Genet*., **6**, 84–89.
10 Sternberg,N. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 103–107.
11 Tao,Q., Zhao,H., Qiu,L. and Hong,G. (1994) *Cell Res*., **4**, 127–133.
12 Zhang,H.-B., Choi,S., Woo,S.-S., Li,Z. and Wing,R.A. (1996) *Mol. Breed*., **2**, 11–24.
13 Kim,U.-J., Birren,B.W., Slepak,T., Mancino,V., Boysen,C., Kang,H.-L., Simon,M.I. and Shyzuya,H. (1996) *Genomics*, **34**, 213–218.
14 Frijter,A.C.J., Zhang,Z., van Damme,M., Wang,G.-L., Ronald,P.C. and Michelmore,R.W. (1997) *Theor. Appl. Genet*., **94**, 390–399.
15 Hamilton,C.M., Frary,A., Lewis,C. and Tanksley,S.D. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 9975–9979.
16 Kanazin,V., Marek,L.F. and Shoemaker,R.C. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 11746–11750.
17 Yu,Y.G., Buss,G.R. and Saghai Maroof,M.A. (1996) *Proc. Natl Acad. Sci. USA*, **93**, 11751–11756.
18 Zhang,H.-B., Woo,S.-S. and Wing,R.A. (1996) In Foster,G.D. and Twell,D. (eds), *Plant Gene Isolation: Principles and Practice*. John Wiley & Sons, Chichester, UK, pp. 75–99.
19 Zhang,H.-B., Zhao,X.P., Ding,X.L., Paterson,A.H. and Wing,R.A. (1995) *Plant J*., **7**, 175–184.
20 Jones,J.D.G., Shlumukov,L., Carland,F., English,J., Scofield,S.R., Bishop,G.J. and Harrison,K. (1992) *Transgenic Res*., **1**, 285–297.
21 Bent,A.F., Kunkel,B.N., Dahlbeck,D., Brown,K.L., Schmidt,R., Giraudat,J., Leung,J. and Staskawicz,B.J. (1994) *Science*, **265**, 1856–1860.
22 Ditta,D., Stanfield,S., Corbin,D. and Helinski,D.R. (1980) *Proc. Natl Acad. Sci. USA*, **77**, 7347–7351.
23 Figurski,D.H., Meyer,R.J. and Helinski,D.R. (1979) *J. Mol. Biol*., **133**, 295–318.
24 O'Connor,M., Peifer,M. and Bender,W. (1989) *Science*, **244**, 1307–1312.
25 Kim,U.-J., Shizuya,H., de Jong,P.J., Birren,B. and Simon,M.I. (1992) *Nucleic Acids Res*., **20**, 1083–1085.
26 Sambrook,J., Frisch,E.M. and Manitatis,T. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
27 Soderlund,C., Longden,I. and Mott,R. (1997) *CABIOS*, **13**, 523–535.