Journal of Oil Palm Research Vol. 15 No. 1, June 2003, p. 12-20

# PRELIMINARY ATTEMPTS AT THE CONSTRUCTION OF LARGE INSERT DNA LIBRARIES FOR OIL PALM (*Elaeis guineensis* Jacq.)

**RAJINDER SINGH*; PHILIP A LESSARD**; TAN SOON GUAN+; JOTHI M PANANDAM++; ANTHONY SINSKEY** and SUAN-CHOO CHEAH***

**ABSTRACT**

*Oil palm is one of the most important oil bearing crops, being by far the highest oil yielder per unit land area in the world. In order to facilitate oil palm genome analysis leading to physical mapping, identification of molecular markers associated with quantitative trait loci (QTLs) and map-based cloning, we attempted to develop the tools and techniques needed to construct a bacterial artificial chromosome (BAC) library for oil palm. A suitable method to purify and prepare the single copy vector (pBeloBACII) for BAC transformation was established. The proper partial digestion conditions for oil palm megabase DNA for BAC library construction were also determined. Several BAC clones were successfully identified. Hybridization of these BAC clones with oil palm DNA as probe confirmed the presence of oil palm DNA in those clones.*

**Keywords:** *Elaeis guineensis*, high molecular weight (HMW), DNA, BAC.

**Date received:** 25 October 2002; **Date approved:** 6 November 2002; **Date revised:** 27 June 2003

## INTRODUCTION

The development of pulsed-field gel electrophoresis (PFGE) (Schwartz and Cantor, 1984) has enabled the manipulation of DNA in megabase (Mb) scale. In plant molecular biology, high molecular weight (HMW) DNA analysis by PFGE has already found widespread applications in genome mapping. The technique can be applied to large scale physical mapping of specific chromosome regions, construction of yeast or bacterial artificial chromosome (YAC or BAC) libraries and map-based cloning of economically important genes (Woo *et al*., 1994). An essential element in all these studies was the availability of good quality HMW DNA of several million base pairs.

The most common method for plant HMW DNA extraction is isolation from protoplasts or nuclei. Recently, isolation from plant nuclei is the method of choice (Zhang *et al*., 1995). This is because protoplast isolation on a large scale can be costly and tedious. Furthermore, HMW DNA from protoplast can contain significant amounts of chloroplast and mitocondrial DNA, which could complicate efforts in chromosome walking studies (Zhang *et al*., 1995).

In this study, we attempted to isolate HMW DNA from the nuclei of oil palm leaf. The analysis of a complex genome such as the oil palm, whose haploid genome is estimated as 1700 Mb (Rival *et al*., 1997), can be greatly aided by the availability of HMW DNA amenable to digestion by rare and frequent cutting restriction enzymes. The feasibility of using the DNA for the construction of large insert libraries, such as BAC, was tested. The BAC cloning system developed in the early 1990s (Shizuya *et al*., 1992)

* Malaysian Palm Oil Board, P. O. Box 10620, 50720 Kuala Lumpur, Malaysia. E-mail: rajinder@mpob.gov.my
** Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Ave., Cambridge MA 02139, USA.
+ Department of Biology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
++ Department of Animal Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.

offered several advantages over earlier systems used to generate large insert DNA libraries, such as YAC. The advantages include stability, efficient transformation and easy handling (Marek and Shoemaker, 1997).

Preliminary attempts were made to use the HMW DNA isolated for BAC cloning. The availability of large insert libraries, such as BACs, in oil palm will enable applications such as map-based cloning and physical mapping. Such studies are important for enhancing knowledge of the oil palm genome.

## MATERIALS AND METHOD

### Isolation of High Molecular Weight (HMW) DNA from Oil Palm Nuclei

HMW DNA was isolated from the nuclei of *tenera* oil palm according to Liu and Whitter (1994). The nuclei from 20 g (fresh weight) leaf tissue were embedded in agarose plugs. The DNA was analysed on a PFGE system with 1% (w/v) agarose in 0.5 M TBE buffer at 6 V $cm^{-1}$, 14ºC with 80-100 s ramped pulse for 20 hr.

To purify undegraded megabase DNA at around 2 Mb, the DNA was run on a 1% low melt point (LMP) agarose (NEB, USA) using the PFGE conditions above. After the electrophoresis, the lanes containing the yeast chromosome size standards (NEB, USA) were removed from the gel and stained with ethidium bromide. The region around 2Mb was marked and used to excise DNA fragments of the same size from the gel. The purified megabase DNA was then stored in TE buffer at 4ºC. The integrity of the purified DNA was also tested on 1% (w/v) agarose gel in 0.5 M TBE, using PFGE as described above.

### Restriction Digestion of High Molecular Weight (HMW) DNA

Prior to digestion, the agarose plugs were washed at 50ºC for 8 hr in several volumes of $T_{10}E_{10}$ buffer (10 mM Tris-HCl pH 8.0 and 10 mM EDTA pH 8.0) containing 1% (w/v) phenylmethylsulfonyl fluoride (PMSF). This was followed by washing in several changes of $T_{10}E_{10}$ buffer. The DNA samples were then equilibrated for 1 hr in 1 *m*l of digestion buffer. Digestion was then performed separately for independent plugs in 250 µl volumes with 70 units of *Eco*RI, *Bam*HI and *Not*I, respectively, at 37ºC overnight. After digestion, the plugs were directly transferred to the gel and subjected to electrophoresis. Southern blotting, probe labelling and hybridization were carried out as described by Cheung and Gale (1990).

### Preparation of *Hin*dIII Partially Digested Oil Palm High Molecular Weight (HMW) DNA

Partial digestion of HMW DNA embedded in LMP agarose was essentially done according to Woo *et al.* (1994). Agarose plugs containing megabase DNA were pre-incubated in 1 x *Hin*dIII buffer (6 mM Tris-HCl, 6 mM $MgCl_2$, 50 mM NaCl, 1 mM DTT, pH 7.5) in a final volume of 800 µl on ice for 30 min. The plugs were then separated into four different Eppendorf tubes and incubated further in 1 x *Hin*dIII buffer, 100 µg $ml^{-1}$ BSA and 4 mM spermidine in a final volume of 250 µl on ice for 30 min. A series of dilutions of *Hin*dIII (NEB, USA) (1 unit, 2.5 units, 5 units and 7.5 units) were added to the different tubes containing the agarose plugs and incubated on ice for another 30 min to allow the enzyme to diffuse into the plugs. The reaction was incubated at 37ºC for 30 min and stopped by the addition of 1/10 volume 0.5 M EDTA pH 8.0 on ice.

The product of the partial digestion was analysed on PFGE using the CHEF DRIII (BioRad, USA) system. The gel was prepared with 1% (w/v) agarose in 0.5 M TBE. The lambda ladder PFG marker (NEB, USA) was used as the molecular weight marker. The gel was subjected to PFGE in 0.5 M TBE buffer at 6 V $cm^{-1}$, 14ºC, with a 20 s pulse for 16 hr. The enzyme concentration which resulted in a majority of the fragments being between 100-400 kb was selected for the large scale partial digestion of HMW DNA to be used for the BAC library construction.

The products of the subsequent digestion with the optimal enzyme concentration were also analysed on the CHEF DRIII (BioRad, USA) electrophoresis system. The gel was prepared with 1% (w/v) LMP agarose in 0.5 M TBE. After electrophoresis, the lanes containing the lambda DNA ladder were removed and stained with ethidium bromide. The regions around 50 to 250 kb were marked and used to excise DNA fragments of the same size from the gel. The DNA fragments in the LMP agarose block were stored in TE buffer at 4ºC.

### Bacterial Artificial Chromosome (BAC) Vector Preparation

The pBeloBACII (7.5 kb) in *E. coli* strain DH10B was obtained from New England Biolabs (NEB), USA. The pBeloBACII was streaked out on LB agar containing 12.5 µg $ml^{-1}$ chloramphenicol, and grown overnight at 37ºC. A single colony was used to inoculate 5 ml LB medium containing 12.5 µg $ml^{-1}$ chloramphenicol, which was incubated by shaking at 37ºC for 6 hr. One millilitre of this culture was used as seed to inoculate three flasks, each containing

1 litre LB medium with 12.5 μg $ml^{-1}$ chloramphenicol, and grown overnight at 37ºC.

The vector DNA was isolated from each flask using different methods: i) Wizard Maxiprep Columns (Promega, USA), ii) alkaline lysis followed by cesium chloride gradient centrifugation, and iii) alkaline lysis followed by polyethylene glycol (PEG) purification. Methods 2 and 3 were as described in Sambrook *et al.* (1989). The vector DNA in all three cases was precipitated with ethanol and dissolved in sterile water.

The pBeloBACII was digested to completion with *Hin*dIII at 37ºC overnight. Complete digestion was verified on a 1% (w/v) TAE agarose gel and the DNA extracted with phenol/chloroform, precipitated with ethanol and resuspended in TE buffer. The vector was dephosphorylated by adding one unit of Shrimp alkaline phosphatase (Roche Molecular Biochemicals, USA) for 1 μg of vector and incubating at 37ºC for 30 min. The phosphatase was inactivated by heating for 10 min at 65ºC, followed by phenol/chloroform extraction and ethanol precipitation. The vector was then resuspended in TE buffer as decribed. The extent of dephosphorylation was assayed by a self ligation test. The vector DNA was stored in 5 μl aliquots at -80ºC.

**Ligation of DNA Fragments into the Bacterial Artificial Chromosome (BAC) Vector**

The LMP agarose gel slices containing the size selected HMW DNA fragments were dialyzed twice with TE, 30 min each time. Each slice was then melted at 65ºC for 10 min, transferred to a 42ºC water bath, digested with β-agarase (NEB, USA) for 2 hr as specified by the manufacturer.

When the gel was completely digested, the DNA was ethanol precipitated and redissolved in TE. Ligation of the DNA to the vector was carried out as described by Marek and Shoemaker (1997).

**Transformation of Recombinant DNA Molecules (BAC) into *E. coli* by Electroporation**

*E. coli* strain DH 10B (Invitrogen, USA) was used as the host. The ligated DNA was transformed into Electromax DH 10B cells (Invitrogen, USA) by electroporation using a GenePulser (BioRad, USA). Ligation mix (5 μl) were added to 24 μl of the cells for a single electroporation at 1.9 kV, 25 μFD and 1000 ohms. The transformed cells were diluted immediately with 1 ml SOC medium (2% Bacto-tryptone, 0.5% Bacto-yeast, 10 mM NaCl, 2.5 mM KCl, 5 mM $MgCl_2$, 5 mM $MgSO_4$ and 20 mM glucose), transferred to 15 ml blue cap tubes, incubated at 37ºC for 1 hr before plating on a selective medium (LB plates with 12.5 μg $ml^{-1}$ chloramphenicol, 0.55 mM IPTG and 80 μg $ml^{-1}$ X-Gal). The clones (BACs) could be identified as blue (non-recombinant) or white (recombinant) phenotype.

**Individual Bacterial Artificial Chromosome (BAC) Analysis**

Individual BAC clones were inoculated into 5 ml of LB broth containing 12.5 μg $ml^{-1}$ chloramphenicol and grown at 37ºC with shaking at 250 rpm overnight. The circular BAC DNA was isolated by the alkaline lysis method (Sambrook *et al*., 1989). The final DNA pellet was dissolved in 15 μl TE and typically 10 μl used for restriction digestion analysis with *Not*I. The digested BAC DNA was analysed by PFGE using a CHEF DRIII (BioRad, USA) electrophoresis system at 4.4 V $cm^{-1}$, 10 s pulse time for 16 hr. The gel was stained with ethidium bromide and photographed.

For Southern blotting, the BAC DNA in the gel was UV treated and blotted onto a positively charged nylon membrane (Roche, USA). The blot was hybridized with genomic DNA from oil palm using the DIG High Prime kit (Roche Molecular Biochemicals, USA) as recommended by the manufacturer.

## RESULTS AND DISCUSSION

**Isolation of High Molecular Weight (HMW) DNA**

The isolation of good quality HMW DNA is a pre-requisite for the construction of large insert libraries, such as BAC. The most common method for isolation of plant HMW DNA involves isolation of protoplasts or nuclei, embedding the protoplasts or nuclei in agarose plugs, followed by cell lysis or degradation (Woo *et al*., 1994). In this study, we attempted to isolate HMW DNA from oil palm leaf nuclei.

In terrestrial plant systems, the rigid cell wall makes it difficult to prepare HMW DNA when compared to mammalian or yeast cells. In this study, the oil palm leaf tissue was homogenized by grinding in liquid nitrogen with a mortar and pestle to break the cell walls physically. Nuclei were isolated from the homogenate and embedded in LMP agarose. The embedded nuclei were then lysed by incubation in a solution containing a detergent (sodium N-lauroyl sarcosine), a protease (proteinase-K) and a high concentration of chelating agent, EDTA. The large DNA molecules remained trapped within the agarose matrix and protected against shearing by the mechanical stress, while the degraded cell material diffused out.

*Figure 1* shows the HMW DNA isolated from oil palm leaf analysed by PFGE using *Sacchoromyces*

*cerevisiae* chromosome as molecular weight marker. The HMW DNA band was higher than that of the 1.9 Mb yeast chromosome. Besides the HMW DNA band, a smear was observed with a size range of 2 Mb to 100 kb. This smear probably contained organelle and sheared genomic DNA. Some sheared genomic DNA was unavoidable due to the grinding of the leaf. However, the amount of sheared DNA was reduced considerably by purifying the undegraded DNA at around 2 Mb as shown in *Figure 2*. The DNA isolated was deemed to be in the megabase range and suitable for construction of large insert libraries.

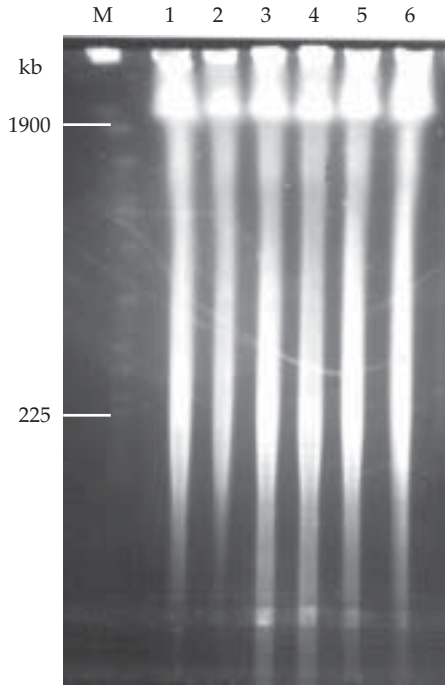


*Figure 1. The pulsed-field gel electrophoresis (PFGE) of megabase DNA from oil palm. The gel (1% agarose, 0.5 x TBE) was run with a CHEF-DRIII apparatus (Biorad) at 6 V cm$^{-1}$, 14°C and ramping times of 80 s and 100 s for 20 hr. Lanes 1 to 6 indicate megabase DNA prepared in agarose plugs. M is the yeast chromosome size standard.*

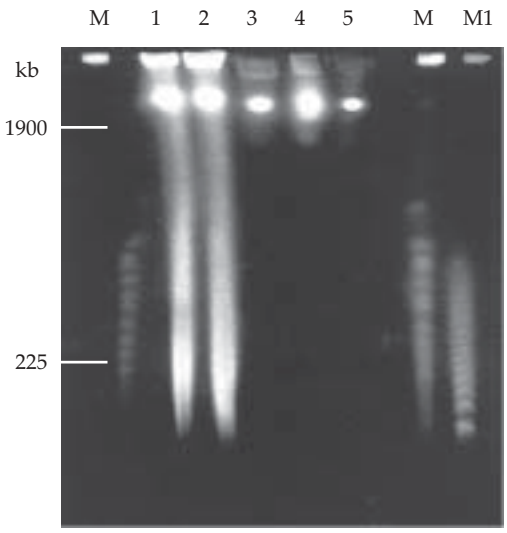


*Figure 2. Gel analysis of purified megabase DNA from oil palm. Lanes 1-2 show unpurified DNA, lanes 3-5 purified DNA. Lane M is yeast chromosome used as size marker and M1 is a Lambda ladder PFG marker.*

In order for the HMW DNA to be suitable for the construction of large insert libraries or physical mapping, it must also be amenable to digestion by restriction enzymes. The ability of the oil palm DNA to be digested by a series of restriction enzymes - *Not*I (8 bp, rare cutter), *Eco*RI (6 bp cutter) and *Bam*HI (6 bp cutter) - was thus determined.

Prior to restriction digestion, the proteinase-K and EDTA were removed from the plugs through diffusion by placing them in large volumes of $T_{10}E_{10}$ buffer. Since proteinase-K is a very stable enzyme, the protease inhibitor, PMSF, was also added to inhibit any remaining activity of the enzyme. Digestion was then carried out using the three restriction enzymes. *Figure 3* shows that by using *Eco*RI and *Bam*HI, the HMW DNA was completely digested, producing fragments of <200 kb. This indicated that all the DNA was accessible to the restriction enzymes and that no local impurities were inhibiting the digestion. As for *Not*I, a significant portion of the *Not*I digest was not resolved and remained in the limited mobility region under the

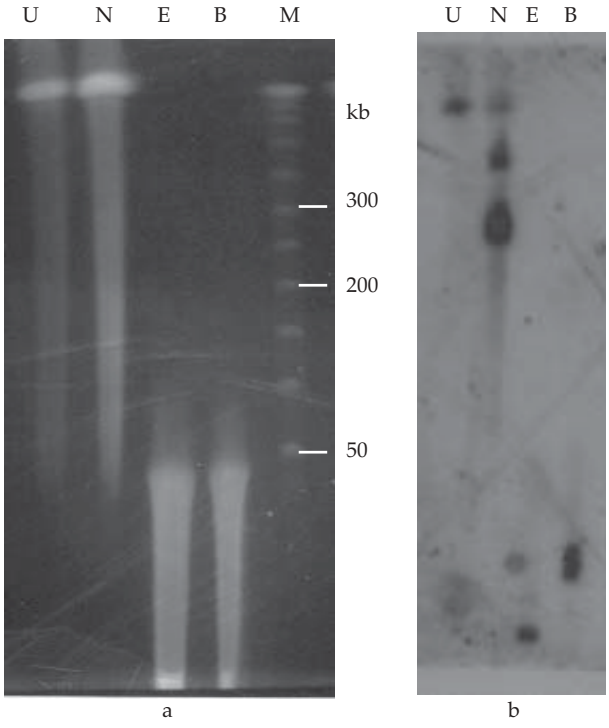


*Figure 3. (a) Digestion of megabase DNA with restriction enzymes as indicated. N:* NotI*; E:* Eco*RI, B:* Bam*HI, and U: undigested. Lane M is a lambda ladder PFG marker. (b) Southern hybridization of gel in (a) hybridized with probe G142 (single copy probe from oil palm).*

pulse conditions used for the gel. Subsequent blotting and hybridization using the probe G142 [a single copy RFLP probe from oil palm, used in genetic mapping studies (Rajinder *et al*., 2001)] revealed the presence of distinct high molecular weight fragments in the *Not*I (500 and 450 kb), *Eco*RI (20 kb) and *Bam*HI (45 kb) digests. Incubation with more enzyme did not alter the *Not*I, *Eco*RI and *Bam*HI hybridization patterns, suggesting that the presence of the two G142-containing fragments was not due to partial digestion. These results indicated that the *in situ* digestion by *Eco*R1, *Bam*HI and the rare cutting *Not*1 proceeded to completion and should be highly useful in generating restriction fragments of plant DNA in the 20-500 kb range.

Since the HMW DNA was amenable to digestion with restriction enzymes, it would also be suitable for constructing large insert libraries for physical mapping.

**Partial Digestion of High Molecular Weight (HMW) DNA**

The construction of a large insert BAC library requires megabase size DNA that can be partially digested with the restriction enzyme, *Hin*dIII. Previous studies have shown it necessary to first determine the optimal conditions for partial digestion. Partial digestion conditions that yield a majority of fragments between 100-400 kb are suitable for the construction of a BAC library (Zhang *et al*., 1996).

As mentioned in the Materials and Methods, various concentrations of *Hin*dIII (1 U, 2.5 U, 5 U and 7.5 U) were used to digest the DNA. The partial digests were analysed by PFGE. *Figure 4* shows that increasing the amount of enzyme increased the digestion of the DNA as was to be expected. Generally, five units of enzyme/agarose plug resulted in DNA fragments of the right size range (100 to 400 kb) for the BAC library construction.

For BAC library construction, digested DNA across several size ranges - 50-100 kb, 100-200 kb and 200-250 kb - was purified from the gel. The fragments were analysed on gel and *Figure 5* shows that all the three correct size ranges were obtained.

**Preparation of the pBeloBACII Vector**

The single copy nature of the vector pBeloBACII made its purification laborious. Three methods were attempted and the results summarized in *Table 1*. The Wizard columns (Promega, USA) resulted in poor quality plasmid contaminated with genomic DNA. Alkaline lysis followed by two days of cesium chloride gradient ultra centrifugation gave very good quality plasmid but with an extremely poor yield. Further, the method was not feasible for

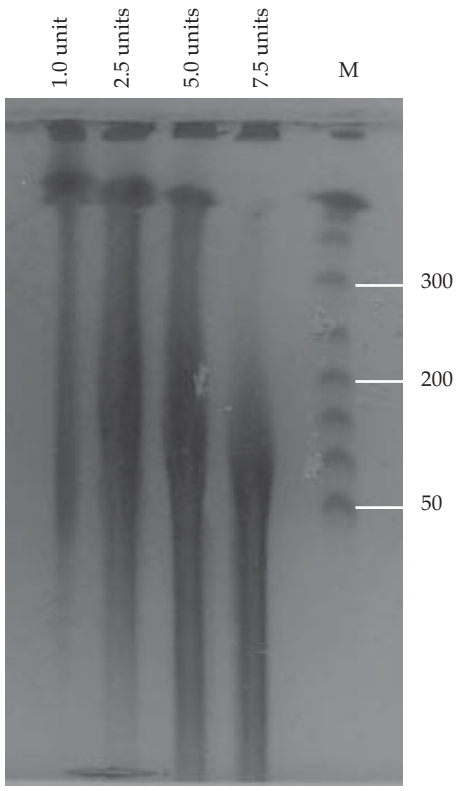


*Figure 4. Partial digestion of oil palm DNA. The megabase size DNA in agarose plugs was partially digested with various concentrations of* Hin*dIII as indicated at the top of each lane. Lane M is a PFG lambda ladder size marker.*

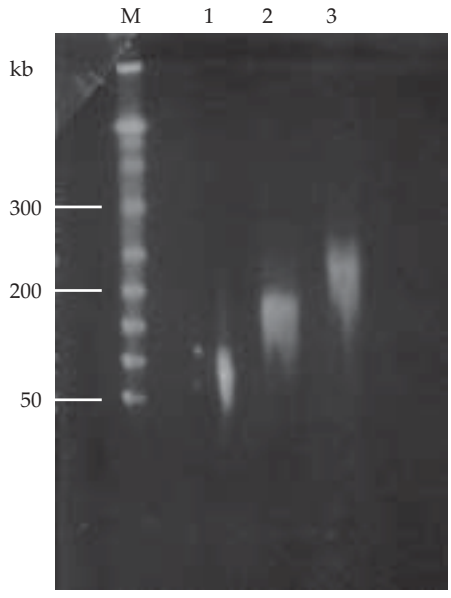


*Figure 5. The DNA fragments purified from low melt agarose gel and examined on a pulsed-field gel electrophoresis (PFGE). Lane 1: 50-100 kb, lane 2: 100-200 kb, lane 3: 200-250 kb and lane M: PFG lambda ladder.*

routine and large scale purification as it was extremely tedious, slow and required large amounts of ethidium bromide, a carcinogen. The alkaline lysis method followed by PEG purification resulted in reasonable yield, and furthermore the quality of the vector was largely satisfactory. Since about 20 ng vector DNA were required per ligation reaction, the alkaline lysis-PEG method resulted in sufficient vector for several ligation reactions.

The concentration and quality of the pBeloBACII vector was determined by electrophoresing an aliquot of the vector on a 0.8% (w/v) TAE agarose gel. The vector also produced bands of the expected sizes (4200 and 3300 kb) when digested with *Kpn*I (data not shown), indicating that it was amenable to digestion by restriction enzymes. The purified vector was then digested with the cloning enzyme, *Hin*dIII, to produce a linear fragment of around 7.5 kb. The digested vector was subsequently treated with alkaline phosphotase to remove the 5′ phosphate (P) groups which was important to avoid re-ligation of the vector. To ensure that the dephosphorylation was successful, an aliquot of the vector was treated with the enzyme, T4 DNA ligase. As shown in *Figure 6* (lane 2) the vector remained linear at 7.5 kb and did not re-ligate to its circular form. Dephosphorylation can also damage the restriction ends and reduce the ligation efficiency to genomic DNA. To ensure that the cohesive ends of the vector were intact, an aliquot of the vector was also treated with T4 kinase to restore the 5′ P groups. Subsequent treatment with T4 DNA ligase resulted in the circular form of the plasmid (*Figure 6*, lane 3), indicating that the dephosphorylation was successful and did not damage the cohesive ends of the vector.

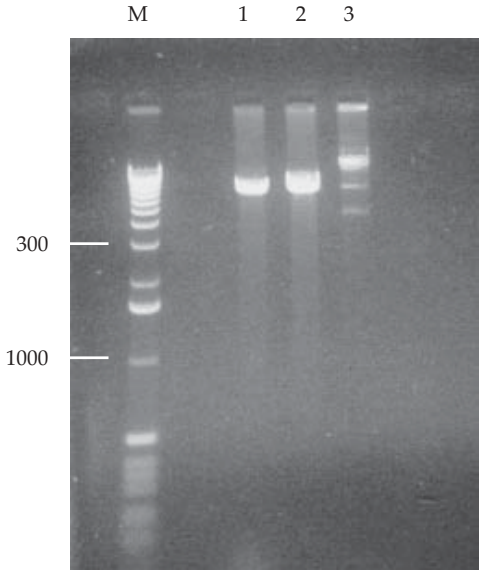


*Figure 6. Preparation of pBeloBACII plasmid for ligation with genomic DNA. Lane 1 indicates the plasmid digested by the cloning enzyme,* Hin*dIII. Lane 2: Plasmid digested (*Hin*dIII), dephosphorylated and treated with T4 DNA ligase. Lane 3: Plasmid digested (*Hin*dIII), dephosphorylated and treated with kinase (to restore 5′ phosphate groups) and subsequently T4 DNA ligase. Lane M is a 1 kb ladder (Gibco BRL, USA).*

**Construction and Characterization of Oil Palm Bacterial Artificial Chromosome (BAC) Clones**

BAC library construction was attempted for *tenera* oil palm, the commercial planting material. The optimal partial digestion conditions giving DNA in the 50-250 kb range using *Hin*dIII were determined. DNA between 50 to 250 kb was isolated from the gel and ligated to pBeloBACII. The ligation was transformed into the *E. coli* strain DH10B, which produced an average of 13 white colonies (recombinant BAC) per ligation mixture. This was, however, much lower than the 200 recombinants reported for *Arabidopsis* (Zhang *et al*., 1996) or the up to 1500 recombinants per transformation for sorghum (Woo *et al*., 1994). As expected, DNA in the lower size range produced more recombinant colonies (*Table 2*).

As shown in *Table 2*, it was obvious that a high percentage of non-recombinant (blue) colonies was produced in each transformation. The blue colonies probably represented part of the vector DNA not adequately dephosphorylated, and, as such, re-ligated into the circular form and transformed into *E. coli* cells. Some amount of background non-recombinant clones can be expected although literature search could not ascertain the likely

**TABLE 1. THE THREE METHODS TESTED FOR pBeloBACII PURIFICATION**

| **Method** | **Yield/litre** (μg) | **Quality** |
|---|---|---|
| Wizard minipreps | 0.8-1.2 | Poor |
| Alkaline lysis - cesium chloride | 0.4 | Good |
| Alkaline lysis- polyethylene glycol | 3.0 | Satisfactory |

Note: *Quality of the vector was assessed by examining an aliquot on a 0.8% TAE gel.

**TABLE 2. ANALYSIS OF TRANSFORMATION EFFICIENCY OF BACTERIAL ARTIFICIAL CHROMOSOME (BAC) CLONES**

| Size of DNA used in ligation (kb) | No. of blue colonies | No. of white colonies | No. of white colonies analysed | No. of clones with inserts | Average size of inserts (kb) |
|---|---|---|---|---|---|
| 50-100 | 140 | 40 | 20 | 8 | 30 |
| 100-200 | 154 | 18 | 10 | 4 | 40 |
| 200-250 | 70 | 11 | 10 | 3 | 40 |
| Total | 294 | 69 | 40 | 15 | 35 |

percentage of such colonies. Nevertheless, the number of such colonies (averaging 80% per transformation) was considerable. As such, further improvement in the vector dephosphorylation and/or ligation of the DNA/Vector (in terms of molar ratio) will need to be carried out.

BAC DNA was isolated from approximately 40 clones and analysed for the presence and average size of inserts. The BAC DNA was digested with *Not*I to release the oil palm DNA fragments from the cloning vector and analysed by PFGE. *Figure 7a* shows an example of the BAC clones examined. The common band in all the lanes is the BAC vector. The remaining bands are oil palm DNA fragments. Some of the clones appeared to have internal *Not*I sites as shown by the multiple bands. The insert size of each clone was determined by adding up the sizes of all the fragments in each lane except the 7.4 kb BAC vector band. The average insert size was about 35 kb, with a range from 15 to 40 kb. This was much smaller than the average insert size of about 100-130 kb reported for BAC clones in other plants (Zhang *et al*., 1996). Smaller DNA fragments trapped within the DNA zone excised for ligation may explain the smaller than expected insert sizes. Smaller fractions tend to ligate more efficiently than larger fragments. In order to increase the average insert size, a second size selection will need to be performed on the digested DNA. The second size selection has been shown to increase the average insert size of BAC DNA (Woo *et al*., 1994). Alternatively, the partial digestion conditions can be re-optimized to obtain DNA fragments >300 kb (between 300-500 kb) for ligation reactions.

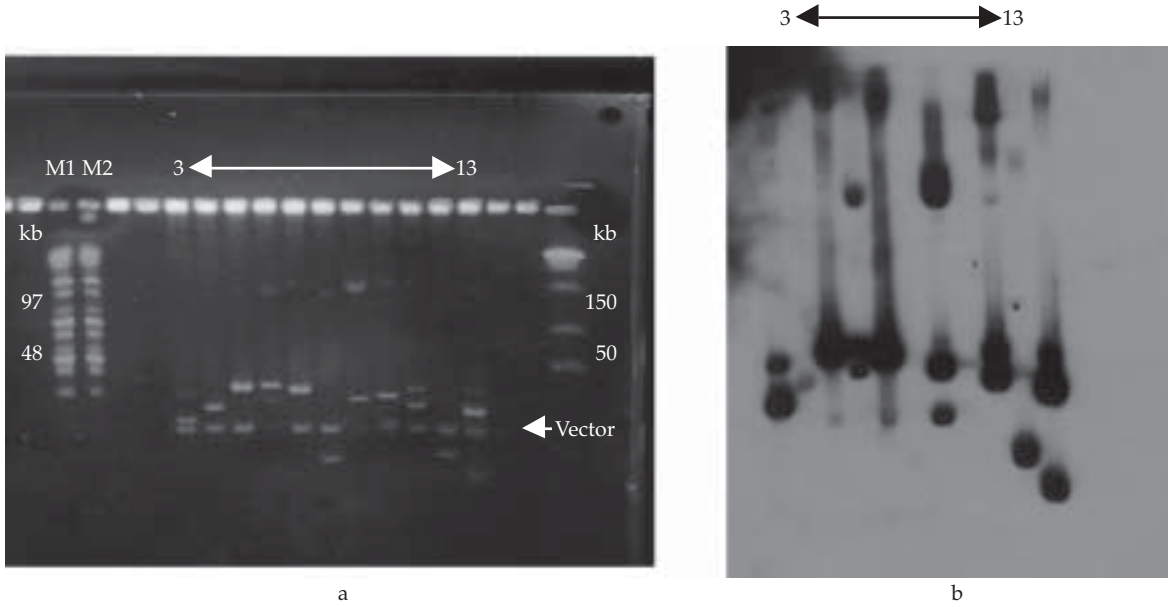


*Figure 7. Analysis of oil palm bacterial artificial chromosome (BAC) clones by pulsed-field gel electrophoresis (PFGE). (a) Ethidium bromide stained CHEF gel. Lanes M1 and M are a mid range PFG marker and PFG lambda ladder, respectively (NEB, USA). Lanes 3-13 are alkaline lysis mini preparations of randomly picked recombinant BAC clones derived from size selected DNA and digested with* Not1*. (b) Southern hybridization of gel in (a) hybridized with radioactively labelled total oil palm DNA.*

From *Table 2*, it is also obvious that only 40% of the white colonies on PFGE after *Not*I digestion appeared to carry oil palm DNA. One possible explanation is that some of these clones had extremely small inserts not visible on PFGE. As such, a second size selection, or using DNA fragments >300 kb, is essential to improve the BAC cloning.

To determine if the inserts detected were in fact oil palm DNA and the type of insert (either genomic low copy or repetitive), the gel in *Figure 7a* was blotted and hybridized with labelled total oil palm genomic DNA. The results in *Figure 7b* show that all the 11 clones examined showed hybridization to oil palm DNA. This was a clear indication that the BAC clones were indeed carrying oil palm DNA. Eight of the 11 clones showed a very strong hybridization signal, indicating that they contained highly repetitive DNA. The remaining three clones showed weak signals, indicating BACs containing low copy DNA.

The hybridization with total genomic DNA could also be due to the presence of chloroplast DNA sequences (especially for those clones with strong hybridization signals). However, the DNA used in making the BAC clones was derived from the nuclei, and as such would have been devoid of chloroplast DNA. Nevertheless, some plant BAC libraries constructed from nuclei have been reported to contain some chloroplast inserts (Woo *et al*., 1994; Zhang *et al*., 1996). Since the presence of the highly repetitive chloroplast sequence can complicate efforts in map-based cloning, the level of chloroplast sequence in the oil palm BAC clones will need to be determined. If it is found to be too high, then improvements will have to be made on the DNA isolation protocol from oil palm nuclei.

## CONCLUSION

Large insert DNA libraries are important for map-based gene cloning, the assembly of physical maps and simple screening for specific genomic sequences in an organism such as the oil palm which has a large genome. Large insert libraries (such as BAC) are important to keep the number of clones required for screening at a manageable level. The DNA libraries usually contain considerable redundancy of cloned DNA. The number of DNA clones needed (N) for a certain probability of finding a target clone can be calculated by:

$$N = \frac{\ln(1\text{-probability required})}{\ln(1\text{- DNA insert size / haploid genome size})}$$

(Paterson, 1986)

As a rule of thumb, a library of DNA inserts which collectively add up to three times (3x) the amount of DNA in a single gamete of the organism will afford about 95% confidence that any DNA element in the genome is represented at least once in the library. A library with *5 genome equivalent* will afford about 99% confidence of having the target gene element. As such, for oil palm (haploid genome size about 1700Mb) (Rival *et al*., 1997), only about 75 000 BAC clones (of 120 kb average size) are required to provide five times (5x) genome coverage, compared to almost 600 000 clones which would have been needed if a normal genomic library (average 15 kb insert size) is used. Thus, a primary objective of this work was to evaluate the utility of the BAC cloning system for the construction of large insert DNA libraries for oil palm. In this study, a general BAC cloning protocol was established for oil palm. The megabase size DNA from leaf nuclei was found to be reasonably suitable for BAC cloning. Several oil palm BAC clones were successfully identified. Hybridization of these BAC clones with oil palm DNA as probe in Southern blot analysis confirmed the presence of oil palm sequence in these clones. Both repetitive and low copy oil palm sequences were cloned. However, some improvement still has to be made to the BAC cloning efficiency to increase the number of clones for effective genome coverage. The average insert size for the oil palm BAC clones obtained was approximately 35 kb. This was much lower than the cloning capacity of the vector (which is at least 300 kb) or the average reported for other plants (around 120 kb). Several strategies have been identified to improve on the insert size and these include: i) a second size selection of the digested DNA to eliminate the smaller trapped DNA fragments; and ii) re-optimization of the partial digestion conditions to obtain most of the DNA in the 300-500 kb range. Nevertheless, in the work done, the method for HMW DNA isolation and the basic parameters for BAC library construction for oil palm were successfully established.

## ACKNOWLEDGEMENT

We thank the Director-General of MPOB and the National Biotechnology Directorate (BIOTEK), Ministry of Science, Technology and the Environment, Malaysia for permission to publish this paper. The work reported was funded under the Malaysia-MIT Biotechnology Partnership Programme (MMBPP).

## REFERENCES

CHEUNG, Y and GALE, M D (1990). The isolation of high molecular weight DNA from wheat, barley and rye for analysis by pulse-field gel electrophoresis. *Plant Mol. Biol*., *14*: 881-888.

LIU, Y G and WHITTIER, R F (1994). Rapid preparation of megabase plant DNA from nuclei in agarose plugs and microbeads. *Nucleic Acids Research*, *22*: 2168-2169.

MAREK, F L and SHOEMAKER, R C (1997). BAC contig development by fingerprint analysis in soyabean. *Genome*, *40*: 420-427.

PATERSON, A H (1986). Physical mapping and map-based cloning: bridging the gap between DNA markers and genes. *Genome Mapping in Plants*. Academic Press, Inc. p. 7-21.

RAJINDER, S; CHEAH, S C; MADON, M; OOI, L C L and RAHIMAH, A R (2001). Genomic strategies for enhancing the value of the oil palm. *Proc. of the 2001 PIPOC International Palm Oil Congress – Agriculture Conference*. MPOB, Bangi. p. 3-17.

RIVAL, A; BEULE, T; BARRE, P; HAMON, S; DUVAL, Y and NOIROT, M (1997). Comparative flow cytometric estimation of nuclear DNA content in oil palm (*Elaeis guineensis* Jacq.) tissue culture seed derived plants. *Plant Cell Reports*, *16*: 884-887.

SAMBROOK, J; FRITSCH, E F and MANIATIS, T (1989). *Molecular Cloning: A Laboratory Manual*. Second edition. Cold Spring Harbour Laboratory Press, Cold Spring Harbour, NY.

SCHWARTZ, D C and CANTOR, C R (1984). Separation of yeast chromosome-sized DNA by pulsed-field gradient gel electrophoresis. *Cell*, *37*: 67-75.

SHIZUYA, H; BIRREN, B; KIM, U J; MANCIANO, V; SLEPAK, T; TACHIIRI, Y and SIMON, M (1992). Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor based vector. *Proc. of the National Academy of Science, USA*, *89*: 8794-8797.

WOO, S S; JIANG, J; GILL, B S; PATERSON, A H and WING, R A (1994). Construction and characterization of a bacterial artificial chromosome library of *Sorghum bicolor*. *Nucleic Acids Research*, *22*: 4922-4931.

ZHANG, H B; CHOI, S; WOO, S S; LI, Z and WING, R A (1996). Construction and characterization of two rice bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. *Molecular Breeding, 2*: 11-24.

ZHANG, H B; ZHAO, X; DING, X; PATERSON, A H and WING, R A (1995). Preparation of megabase-size DNA from plant nuclei. *Plant Journal*, *7*: 175-184.