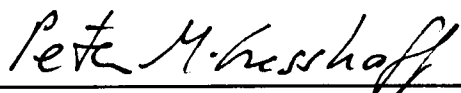


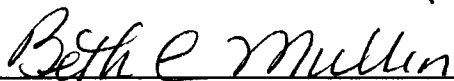
To the Graduate Council:

I am submitting herewith a thesis written by Sanjeev Vasu Pillai entitled "Construction and Characterization of High Molecular Weight DNA Clones of Legumes for Positional Cloning". I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Science, with a major in Life Sciences.



Dr. Peter M. Gresshoff, Major Professor

We have read this thesis and
recommend its acceptance:



Dr. Beth C. Mullin



Dr. Jeffrey M. Becker

Accepted for the Council:



Associate Vice Chancellor and
Dean of The Graduate School

**CONSTRUCTION AND CHARACTERIZATION OF HIGH
MOLECULAR WEIGHT DNA CLONES OF LEGUMES FOR
POSITIONAL CLONING**

A THESIS
PRESENTED FOR THE
MASTER OF SCIENCE
DEGREE
THE UNIVERSITY OF TENNESSEE
KNOXVILLE

SANJEEV VASU PILLAI

MAY 1998

ACKNOWLEDGMENTS

It has been a long road to reach thus far, but at the end of it all I have a great sense of achievement. I want to thank my major advisor Peter Gresshoff for all that he has done for me over the last three years in terms of advice, encouragement, appreciation and support. I also wish to express my gratitude to my other thesis committee members for their advice. It goes without saying that this work would not have been possible without the cooperation and support of the PMG group. I really enjoyed experiencing the diversity in the group. I wish to thank Roel Funke for introducing me to the field of high molecular weight DNA, Debbie Ellis for her constant cooperation and friendship and Artem Men for joining me in the final stages of my project. Finally, I want to thank everyone in the PMG lab, especially Jiri Stiller and Ru-Ju Chian, for the wonderful work environment that I had in the lab all this time.

ABSTRACT

The purpose of this study was to construct and characterize high molecular weight DNA (HMW DNA) clones for further use in the positional cloning of legume nodulation genes. Bacterial and Yeast Artificial Chromosomes (BACs and YACs), two of the most powerful and popular HMW DNA tools were constructed and characterized. Later work mostly concentrated on BACs. This work forms a part of a larger project aimed at studying genes that control nodulation, a natural process by which leguminous plants, in association with specific soil-inhabiting bacteria form specialized root structures called nodules. For this purpose, this thesis is focussed on two plant species, one, the crop plant soybean (*Glycine max* L. Merr) and the other, a model legume *Lotus japonicus* (Regal) Larsen. The specific phenotype that was followed up for positional cloning in soybean was nitrate tolerant symbiosis (*nts*), a mutant that exhibits supernodulation - a condition in which the autoregulatory control of nodulation is diminished.

Conditions were standardized for the construction of appropriate sized HMW DNA clones of soybean and *Lotus*. A HMW DNA library of *Lotus* was also constructed. Clones were characterized for insert size and stability. Pulsed Field Gradient Electrophoresis (PFGE) was used to prepare HMW DNA for cloning. Difficulties encountered in the construction of clones are described along with solutions. A comparison of BACs and YACs as the HMW DNA clones of choice was done. Screening of BAC clones identified candidates for an RFLP marker pUTG-132a that is tightly linked to the *nts-1* gene. End clones were isolated from these candidates to facilitate construction of contigs and chromosome walking. The general utilization of HMW DNA clones in genome analysis is described.

TABLE OF CONTENTS

CHAPTER	PAGE
I	INTRODUCTION AND LITERATURE REVIEW
Positional Cloning.	3
Plant genes isolated by positional cloning.	5
Pulsed Field Gel Electrophoresis (PFGE).	10
Yeast Artificial Chromosomes (YAC).	13
Plant YAC libraries.	20
Bacterial Artificial Chromosomes (BAC).	24
Plant BAC libraries and their applications.	28
Nodulation and Symbiotic Nitrogen Fixation	32
Nitrate Tolerant Symbiosis and Supernodulation.	36
<i>Lotus japonicus</i> - A model legume.	38
Objective-Positional cloning as a strategy to study nodulation	39
II	MATERIALS AND METHODS
Preparation of High Molecular Weight (HMW) DNA.	42
Estimation of enzyme levels for optimum partial digestion.	44
Partial digestion and size selection by PFGE	46
Isolation and preparation of the vectors.	47
Ligation of HMW DNA.	48
Transformation.	49
Insert size characterization.	51
Screening of BAC pools for candidate clones	54
Confirmation of candidate BAC clones.	56
Isolation of BAC end clones	57
Sequencing of BAC end clones.	58

	PCR amplification using end clone primers . . .	58
III	RESULTS	
	Construction of HMW DNA clones	
	Isolation of HMW DNA	60
	Partial digestion and size-selection by PFGE . . .	61
	Isolation and preparation of vectors	64
	Ligation and Transformation	66
	Insert size characterization of HMW DNA clones. .	70
	Utilization of BACs	
	Screening of BAC pools for pUTG-132a candidate clones.	75
	Confirmation of candidate BAC clones.	80
	Isolation of end clones.	83
IV	DISCUSSION AND CONCLUSIONS	
	Technical aspects of HMW DNA library construction. .	91
	BACs Vs YACs - a comparison based on technical feasibility.	96
	Identification of candidate BAC clones.	98
	Future perspectives.	103
	REFERENCES.	106
	VITA	130

LIST OF FIGURES

Figure		Page
1-1	Comparison of cloning strategies	7
1-2	Contour-Clamped Homogenous Electric Field (CHEF)	12
1-3	pYAC4 vector	14
1-4	pBeloBAC11 vector	25
1-5	The arrangement of pUTG-132a	38
3-1	PFGE of partially digested DNA	54
3-2	Plasmid-Safe DNase enzyme treatment of the BAC vector	65
3-3	Preparation of the 7.4 kb BAC vector	66
3-4	Preparation of the YAC vector	67
3-5	PFGE insert size check of <i>Lotus</i> BACs	71
3-6	PFGE insert size check of a set of soybean BACs	71
3-7	An initial set of <i>Lotus</i> BACs	72
3-8	PFGE separation of <i>Lotus</i> YACs	74
3-9	Hybridization of the PCR products obtained on amplifying pool 14	78
3-10	Hybridization of the PCR products obtained on amplifying pool 40	79
3-11	PFGE of the candidate BACs	81
3-12	Nested PCR amplification of the 8 candidate BACs	82
3-13	Endclone analysis	83
3-14	Hybridization of endclones	84
3-15	PCR amplification of BACs using 5RE primers	87
3-16	PCR of soybean and <i>Lotus</i> genomic DNA using 5RE primers	89
3-17	PCR amplification of a 1.1 kb pUTG-132a fragment	90

ABBREVIATIONS

BAC	Bacterial Artificial Chromosome
CHEF	Contour-Clamped Homogenous Electric Field
EST	Expressed Sequence Tag
FISH	Fluorescent <i>In-situ</i> Hybridization
GFP	Green Fluorescent Protein
HMW	High Molecular Weight
PCR	Polymerase Chain Reaction
PFGE	Pulsed Field Gradient Electrophoresis
RFLP	Restriction Fragment Length Polymorphism
STS	Sequence Tagged Site
YAC	Yeast Artificial Chromosome

CHAPTER I

INTRODUCTION AND LITERATURE REVIEW

In the last two decades, molecular tools have revolutionized biological sciences. One of the most striking developments in this respect has been the emphasis on genome analysis. Genome analysis refers to the in-depth study of the arrangement of DNA in the chromosomes of an organism. Such a genomic study, coupled with phenotypic analysis, is expected to give a complete molecular level understanding of the development and functioning of the organism. The term "Genomics" was coined in 1986 by Thomas Roderick to describe the scientific discipline of mapping, sequencing and analyzing genomes (Hieter and Boguski, 1997). Genomics is now undergoing an expansion from the mapping and sequencing of genomes to an emphasis on genome function. Hence, genome analysis may now be divided into "structural genomics" and "functional genomics". Structural genomics refers to the initial phase of genome analysis - the construction of high-resolution genetic, physical and transcript maps of an organism. The ultimate physical map of an organism is of course its complete DNA sequence. Functional genomics refers to the development and application of experimental approaches to assess gene function by making use of the information provided by structural genomics. The fundamental strategy is to expand the scope of biological investigation from studying single genes or proteins to studying all genes and proteins at once in a systematic fashion. This approach promises to narrow the gap between sequence and function and to yield new insights into the behavior of biological systems.

The scientific and commercial implications of such an approach can never be overemphasized. Scientific implications range from understanding the basic processes taking place in a living cell to the evolutionary relationships between the millions of organisms that inhabit our planet. Applied aspects range from increased food production to more efficient drugs for combating diseases to a cleaner environment. Realizing this potential, genome analysis projects have been launched in a wide range of organisms in the last decade (Table 1). The best known of these is of course the Human Genome Project (Olson, 1993), an effort aimed at deciphering the arrangement of the 3 billion base pairs of

the human genome. Similar projects have been done or are underway in mouse, the common bakers yeast *Saccharomyces cerevisiae*, the bacterium *Escherichia coli*, the fruit fly *Drosophila melanogaster*, the nematode *Caenorhabditis elegans*, and a whole range of pathogenic microorganisms.

Table1
Current status of genome projects

Organism	Genome size(Mb)	Sequenced(Mb)	Percentage done
Microbial genomes (About 11)	0.6-4.2	0.6-4.2	100
<i>E.coli</i>	4.6	4.6	100
Yeast (<i>S. cerevisiae</i>)	13	13	100
<i>C. elegans</i>	100	71	71
<i>Drosophila</i>	130	8	6
Mouse	3000	6	0.2
Human	3000	60	2

Plant genome analysis has centered around the model plant *Arabidopsis thaliana*, commonly known as the mouse ear cress (reviewed by Meyerowitz, 1987). However, substantial efforts have been put into the genome analysis of commonly used crop plants like rice, maize, sorghum, soybean and several fruit and vegetable crops (i.e. tomato, lettuce etc.) Plant genome analysis is complicated by some factors, most importantly the abundance of repetitive DNA in the genome of many crop plants. The presence of repetitive DNA sequences confuse gene identification experiments, since the probes generated could be repetitive sequences and could lead to any other part of the genome where these sequences are present (Dunford et al. 1993). Here again different plant species have varying degrees of repetitive DNA in their genomes. Relatively small genomes like those of *Arabidopsis* and tomato have relatively small amounts of repetitive DNA (14% and 15-20% respectively; Lapitan, 1992). Larger genomes, like that of wheat and maize have >60% of repetitive sequences (Lapitan, 1992) while that of soybean consists of 30-40% repetitive

DNA (Gurley et al., 1979) There also appears to be a relationship between the genome size and the stretch of continuous unique sequences. Smaller genomes have long stretches of unique DNA while larger genomes have short stretches of unique DNA interspersed between stretches of repetitive DNA (Lapitan, 1992). The average length of unique sequences in the *Arabidopsis* genome is around 120 kb (Lapitan, 1992), while in the soybean genome about 60-70% of single copy sequences have an average length of 1.11-1.36 kb interspersed with 0.3-0.4 kb repetitive elements (Gurley et al., 1979). The complications arising from these variations in genome size and characteristics forced plant genome researchers to look for a representative model plant. As a model plant *Arabidopsis* has many advantages but it lacks some important biological phenomena like symbiotic nitrogen fixation. So, for scientists studying this phenomenon it became imperative to study leguminous plants including crop plants like soybean which have much higher genome sizes. However, the concept of model legume plants is emerging, with *Lotus japonicus* (Handberg & Stougaard, 1992) and *Medicago truncatula* (Barker et al., 1990) being introduced to facilitate easier molecular level understanding of this natural phenomenon. Despite all these, an *Arabidopsis* genome sequencing project has been launched and is expected to be complete by the year 2005 (Delseny et al, 1996). This will give plant researchers valuable information about the molecular genetic functioning of a plant, just as the human and mouse genome projects would do for human genome researchers.

Positional Cloning

With the increased emphasis on genome analysis a variety of methodologies have been developed to facilitate such studies. One such technique is positional or map-based cloning. By definition it refers to the cloning of a gene that is known only phenotypically, by relating it with the position of linked markers on the genome (Wicking & Williamson, 1991). The following are the steps involved in positional cloning:

- 1) Detection of the phenotype - Since the only clue available in positional cloning is the expression of the phenotype, this step is the basic step in such a venture. In case of humans

this phenotype could be the expression of a disease, while in plants this could be an abnormal morphological character.

2) Determination of inheritance - Once the phenotype has been detected the next step is to determine the segregation pattern of the phenotype over generations. This is done either by tracing familial records, in case of human beings, or by following segregating experimental breeding populations, in case of plants.

3) Genetic and physical mapping - After the inheritance has been determined, the position of the gene in the genetic map is located. This is done by linking the inheritance with a genetic marker that is already present in the genome. For this it is necessary to have a well-characterized and saturated genetic linkage map of that region of the genome. In case of well-characterized maps it might even be able to locate the gene between two closely linked markers. Once associated with the marker, the genetic distance of the gene from the marker is determined, from which it will be possible to determine the physical distance.

4) Isolation of HMW DNA clones - It is at this step that it becomes imperative to have a HMW DNA library of the organism under study. A clone containing the linked marker(s) is isolated. Depending on the size of the clone and the physical distance between the marker and the gene, a set of regional HMW DNA clones are isolated until the gene of interest is reached or covered. This process is easier if a flanking marker is available on the other side of the gene of interest. It is even possible, in some cases, to isolate one HMW DNA clone with the gene and the marker(s).

5) Location of the gene - This step can be carried out in different ways. One way is to search for conserved sequences in the isolated HMW DNA clones, and isolate RNAs encoded in those conserved sequences from the affected tissue. Once the RNAs are available, the mutated gene responsible for the phenotype can be found by sequence comparison with the normal phenotype. This is usually done in case of human diseases. In plants, functional complementation analysis can be done by transforming the mutant phenotype with the isolated piece of HMW DNA to restore the wild type phenotype. The

transforming DNA sequence is sub-cloned into smaller fragments and the transformations repeated. This finally allows the isolation of a single piece of DNA that contains the gene of interest. RNA analysis is done to confirm the findings. Molecular characterization, by sequencing, is the last step in the process.

Many human and mammalian genes have been isolated by positional cloning. Examples include the Cystic Fibrosis gene (CFTR; Harris et al., 1993) and the Duchenne and Becker Muscular Dystrophy gene (Hyser, 1989). Industry has also acknowledged the potential of positional cloning and many pharmaceutical and biotech companies are pursuing this as a strategy to isolate genes involved in the development of diseases. A classic example is a genomics company - Myriad Genetics, Salt Lake City, UT - that utilises the extensive familial records available with certain Utah families to isolate human disease genes.

Plants genes isolated by positional cloning

Positional cloning is an invaluable tool for isolating plant genes. Many of the plant genes of interest, like disease resistance genes and genes involved in developmental pathways, cannot be cloned by conventional procedures. Either the gene product or the time and location of expression of these genes are not known. However, most of them have clearly detectable phenotypic expression. Hence, provided required resources like enriched genetic maps and HMW DNA libraries, positional cloning can be pursued as an effective strategy in isolating plant genes. Given these facts it is not surprising that most of the genes cloned this way have come from *Arabidopsis*, tomato (*Lycopersicum esculentum*) and rice (*Oryza sativa*). These plants have small genomes that are relatively devoid of repetitive sequences. This in turn, facilitated the early development of genome analysis tools. Detailed enough genetic maps have been constructed in *Arabidopsis* (Chang et al., 1988; Nam et al., 1989) and tomato (Tanksley et al., 1992) that most of the genes lie within few hundred kilobases of the nearest marker. HMW DNA libraries are also available for both *Arabidopsis* (Grill and Somerville, 1990; Ward and Jen, 1990; Wang et al, 1996; Creusot et al, 1995) and tomato (Martin et al, 1992). Rice is fast becoming the model for cereal genetics (Havukkala, 1996). Completion of the physical map is eagerly awaited by

researchers. The most significant advances have been the progress towards positional cloning of genes and identification of quantitative trait loci (QTLs) from detailed RFLP maps.

The obvious advantage with plants is the ability to functionally complement, by transformation of the mutant phenotype with the isolated wild-type gene. This is facilitated by an efficient transformation system. High frequency *Agrobacterium*-mediated transformation systems have been established in many plants especially the model ones like *Arabidopsis* (Koncz et al., 1992). All the genes that have been isolated by positional cloning have been tested thus for activity.

Scientists have also attempted to solve problems associated with large plant genomes. The original concept behind positional cloning was to find a DNA marker linked to a gene and then walk to the gene by way of overlapping clones. This is difficult to apply for most plant species with large genomes. A paradigm of chromosome "landing" instead of walking has been proposed for larger genomes (Tanksley et al., 1995). This strategy suggests that one first isolates one or more DNA marker(s) at a physical distance from the targeted gene that is less than the average insert size of the genomic library being used for clone isolation. The DNA marker is then used to screen the library and isolate (or "land on") the clone containing the gene, without any need for chromosome walking and its associated problems. This is likely to be the main strategy by which positional cloning is applied to isolate genes from plants with large genomes (Fig. 1-1)

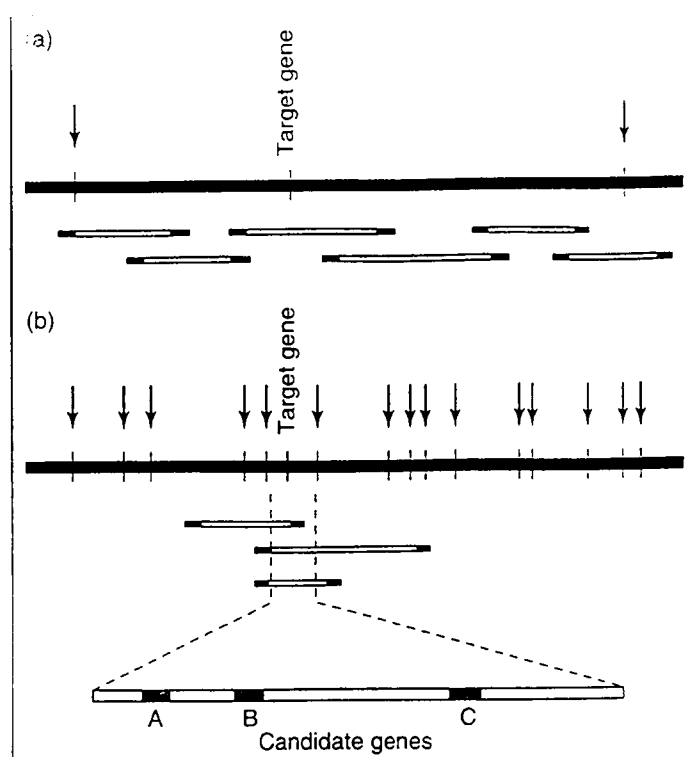


Figure 1-1: Comparison of cloning strategies. Comparison of strategies using (a) chromosome walking and (b) chromosome landing. Arrows indicate positions of linked DNA markers. In (a) there are only two markers available, at considerable distances from the gene. To isolate the gene, one has to laboriously build contigs connecting the two markers. On the other hand in (b) the region around the gene is enriched for markers. Isolation of a few overlapping clones containing markers close to the gene can result in the isolation of the gene. The dark boxes at the ends of the clones indicate overlapping chromosomal regions. (Figure adapted from Tanksley et al., 1995)

The most recent example of a plant gene isolated by positional cloning is the sugarbeet nematode resistance gene *Hs1* isolated from the sugarbeet *Beta vulgaris* (Cai et al, 1997). The *Hs1* gene was cloned with the use of genome-specific satellite markers and chromosomal break point analysis. Expression of the corresponding cDNA in a susceptible variety conferred resistance to infection with the cyst nematode. A well-characterized HMW YAC library was used in this case. Another commercially important gene, the rice blast resistance *Xa-21*, was isolated about two years back (Song et al, 1995) from a resistant line of rice. This work also facilitated the isolation of a whole cluster of a putative disease resistance gene family. BAC and cosmid libraries were used to build a contig of clones from a tightly linked marker to the gene. Sub-cloning and transformation revealed a receptor kinase-like protein, that might be activating a signaling cascade that confers disease resistance.

The first plant gene to be isolated by positional cloning was *fad3*, the gene controlling fatty acid desaturation of membrane and storage lipids (Arondel et al., 1992), from *Arabidopsis*. This was done by scoring the fatty acid composition of F₂ progeny from a cross between mutant and wild-type lines of *Arabidopsis*. The gene was located 0.4 cM to one side of a pair of RFLP markers. YACs hybridizing to these markers were identified. A contig of 170 kb was built. This contig was found to be enough to encompass the *fad3* since 1 cM in that region of the genome was found to be equal to 105 kb. A 120 kb YAC suspected to contain the *fad3* gene was hybridized to a cDNA library. Candidate clones were picked and one of them was confirmed as being *fad3* by functional complementation.

The *Arabidopsis* ethylene response *ETR1* gene required more extensive chromosome walking using YAC and cosmid libraries (Chang et al., 1993). After localizing the gene between two morphological markers, one RFLP marker was found that co-segregated completely with the mutation. Two YACs spanning about 230 kb were isolated, using which, a contig of cosmids was built. Polymorphic DNA fragments from the chromosome walk were used to locate recombination breakpoints in the F₂ progeny and the gene was zeroed in on by segregation analysis.

The tomato *Pto* locus, which confers resistance to bacterial speck disease was isolated by Martin et al in 1993. A lot of preliminary work had to be done to identify markers linked tightly to the gene, because the tomato genome was much less well-characterized than the *Arabidopsis* genome. The resistance gene was naturally introduced into *L. esculentum* from a wild tomato species *Lycopersicon pimpinellifolium*. Thus susceptible and resistant near-isogenic lines of *L. esculentum* were available with the resistant being polymorphic to *L. esculentum*. This polymorphism was exploited to identify molecular markers linked to the *Pto* gene (Martin et al., 1991). The closest that cosegregated with *Pto* was found to lie within 0.6 cM of the gene. This marker was used to isolate a 400 kb YAC and the gene located just inside the right arm of this YAC.

Another *Arabidopsis* gene to be isolated by positional cloning was the ABA (abscisic acid) response *ABI1* gene (Leung et al., 1994; Mayer et al, 1994). The methodology followed was in principle the same as above. A YAC contig of 250 kb was assembled and a maximal segment of about 150 kb was isolated that encompassed the gene. However, the reverse order of gene identification was followed from that point. Subclones of this contig were randomly used to transform the mutant phenotype to confer the wild-type ABA resistant phenotype. Several rounds of subcloning and transformation converged on a short segment of DNA that was used to screen a cDNA library and isolate the *ABI1* gene.

Other plant genes isolated by positional cloning are the *Arabidopsis* *RPS2* gene (Bent et al., 1994), that confers resistance to bacterial speck disease caused by *Pseudomonas syringae* and the tomato *Cf-2* locus (Dixon et al., 1996) that confers resistance to leaf mold caused by *Cladosporium fulvum*.

The near future will see a large number of plant genes, with a wide range of functions, isolated by positional cloning. The main reasons for this include the development of the necessary tools, development of *Arabidopsis* as a model plant, and the development of comparative genetics, through which it will be possible to isolate genes

from one organism using information available about related genes from another closely related organism. Advances in transformation technology, especially the standardization of techniques to transform plants with *Agrobacterium* carrying DNA molecules as large as 150 kb (Hamilton et al., 1996), will aid a lot in doing functional complementation analysis of entire multi-gene families. All this points to a strong role for positional cloning as a tool in plant genome analysis.

Pulsed Field Gel Electrophoresis (PFGE)

One of the basic requirements for positional or map-based cloning of genes is the isolation of large clonable fragments of DNA. This would facilitate the construction of large insert libraries, which are an absolute prerequisite in chromosome walking experiments. Conventional electrophoretic methods can separate DNA only up to about 20 kb in size beyond which, all the DNA molecules tend to migrate the same. DNA molecules separated in agarose under a constant electric field are retarded by the agarose matrix. This retardation is inversely proportional to the size of the molecules. However, beyond a certain size, DNA molecules tend to attain a rod-like shape, with more or less the same cross-sectional area. This results in uniform migration, regardless of the size of the molecule. Under specific conditions, using static field electrophoresis can separate DNA up to 50 kb in size, but not any bigger. This became a major impediment in genome analysis experiments since it is necessary to have large pieces of contiguous DNA.

A big breakthrough in separating large pieces of DNA came with the development of Pulsed Field Gel Electrophoresis (PFGE; Schwartz and Cantor, 1984). PFGE is based on the size-dependent relaxation of DNA molecules when subjected to differential electric currents. Periodic changing of the orientation of the electric field causes the DNA in the gel matrix to relax on the removal of the first electric field and elongate to align with the new field. This process of relaxation and elongation essentially depends on the size of the molecules, with the smaller ones adjusting much faster to the differential fields than the larger ones. When the electric field changes, the DNA molecules have to re-orient themselves before moving in the new direction. The time required for this reorientation is

very sensitive to the length of the molecule, as described above. Large DNA molecules take longer time to realign than the smaller ones because of the physical barrier of the agarose matrix. Hence molecules of increasing size have to spend a larger portion of each switching cycle reorienting before they can begin to migrate through the gel. As long as the alternating fields are equal with respect to length and voltage, the DNA will migrate in a straight path down the gel that reflects the sum of the many short zigzag steps actually taken.

In essence, any electrophoresis that uses more than one electric field in which the electric fields are activated alternately can be called a Pulsed Field. The amount of time for which each alternating field is active is called the switch interval or pulse time. The resolution of DNA is primarily determined by this parameter. Depending on the pulse time DNA molecules above a specific size move together in what is called a compression zone. The molecules below this specific size limit are separated out. The angle between the two alternating electric fields is called the reorientation angle. This angle determines the direction in which the DNA moves. Another parameter is the Voltage Gradient, which is the electrical potential applied to the gel, measured in volts/cm.

There are different kinds of PFGE, of which the one most commonly used is the Contour-clamped Homogenous Electric Field (CHEF; Chu et al., 1986; Fig. 1-2). This type of PFGE was developed primarily to obtain straight line migration of DNA. Homogenous electric fields are generated by using multiple electrodes in a closed contour. Twenty four electrodes are arranged in a hexagonal array, four on each side of the array. The original CHEF apparatus could generate homogenous fields using only a fixed reorientation angle (120°). However, newer versions of CHEF can use different reorientation angles to separate very large fragments. The main advantage of CHEF is its ability to separate a large number of DNA fragments in straight lines.

Since PFGE was developed in 1984, the technique has undergone a lot of modifications both in equipment and protocols. PFGE is extensively used in all labs working on genome analysis. A comprehensive lab manual has been published (Birren and Lai, 1993) covering all aspects of PFGE in detail. Many companies now produce PFGE

equipment of different dimensions and all of these are commercially available for the use of researchers.

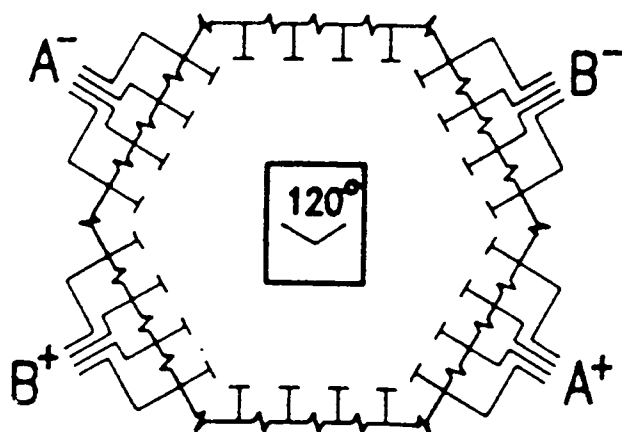


Figure 1-2: Contour-clamped Homogenous Electric Field (CHEF). The arrangement of the 24 electrodes is shown. The switch of current takes place between the A and B sets of electrodes. The angle between the two sets is 120 degrees. (Figure adapted from Birren and Lai, "PFGE- A Practical guide")

The most important application of PFGE is in bridging the gap between physical and genetic maps. Genetic and physical maps yield qualitatively different information - genetic maps allow one to get an overall view of the entire chromosome with low resolution of details, whereas physical maps provide very high resolution in local regions of a chromosome without necessarily linking distant regions. The resolution in genetic analysis usually is in centimorgans, corresponding to hundreds of kilobases. Inability to separate clonable DNA beyond 50 kb had hampered relating physical and genetic maps, especially of eukaryotes. PFGE integration of physical and genetic maps is of utmost importance in positional cloning of genes. Because of the high resolution power of PFGE a detailed genetic map around a gene can be linked with the physical map and the genomic region encompassing the gene can be cloned in HMW DNA clones. PFGE led to the development of vectors capable of cloning very large pieces of DNA. YAC and BAC (described later) vectors were developed to clone the large fragments generated by PFGE.

Physical mapping of chromosomal regions with low level of recombination has been restricted by the shortage of markers in those areas. Molecular markers for these areas have been found by analysis of large DNA fragments resolved by PFGE (Warburton and Willard, 1990; Kolchinsky and Gresshoff, 1994).

When molecular markers detect a mixture of polymorphic and monomorphic bands in rare-cutter enzyme digests, it presents a difficulty to the researcher to identify the correct polymorphic band and to determine whether some of the bands are partial digestion products. PFGE helps in solving this problem and some physical maps have been constructed (Walter et al., 1990; Warburton and Willard, 1990; Funke et al., 1994). The DNA was first separated by PFGE and the entire lane of separated DNA digested, with the enzyme. A second gel was then run perpendicular in direction of the first one. Hybridization of the second run assigned markers to the individual bands.

Another application of PFGE is to produce electrophoretic karyotypes of DNA from organisms with small genomes (Verdoes et al., 1994). Single chromosomes of bacteria or other prokaryotes can be restricted by rare cutting enzymes (or partial digests) and the resulting fragments separated by PFGE. This makes it possible to construct physical maps of the genomic area of interest. Resolution of individual chromosomes of lower eukaryotes is more challenging because of the size of the chromosomes and so the PFGE might take days to complete. As an alternative to hybridization, PCR can also be used for physical mapping by PFGE (Denton et al., 1994).

Yeast Artificial Chromosomes

After the development of technologies to separate large pieces of DNA, the next major step was to design a stable system to maintain these large fragments of DNA. As stated before, one of the major requirements for positional cloning is a representative HMW DNA library of the organism.

A major impediment for genome analysis using plasmid or phage cloning systems was the low average insert size, as it is very difficult to clone and stably maintain pieces of DNA larger than ~30 kb in such systems. An alternative came with the development of yeast artificial chromosomes (YACs; Burke et al., 1987). This technology, as the name suggests, consists of cloning large pieces of DNA into modified vectors that have all the

short but essential sequences of, and hence will mimic, native yeast chromosomes (Figure 1). Since the native yeast chromosomes are 225 - 2000 kb in size, it was expected that it should be possible to maintain foreign DNA of many hundred kilobases cloned into artificial chromosome vectors. The most commonly used vector is pYAC4 (Burke et al., 1987; Fig. 1-3).

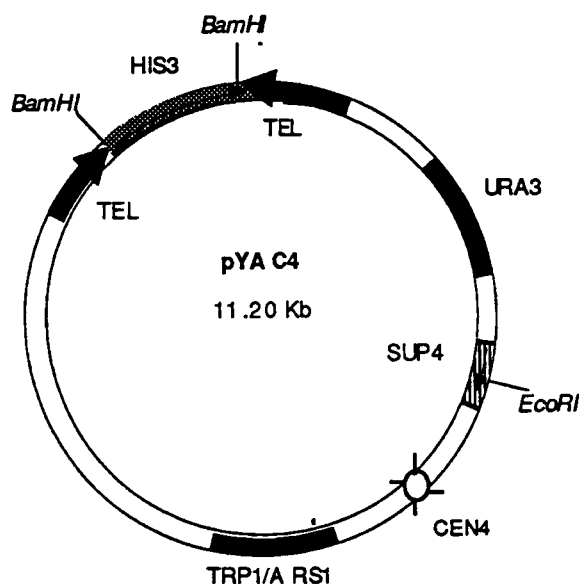


Figure 1-3: pYAC4 vector (see text below for description)

The vector carries sequences necessary for an artificial chromosome - a centromere (*CEN4*) and an autonomous replicating sequence (*ARS*; with properties expected of replication origins), which had previously been cloned from the baker's yeast *S. cerevisiae* and identified for cell function. The telomeres (*TEL*) were derived from the simple ciliate eukaryote *Tetrahymena*. As the telomeric sequences are homologous between the two species, the yeast system recognizes it. The telomere sequences are separated by a spacer fragment (*HIS*) with *Bam* HI cloning sites at either end. The vector also consists of an amber suppresser gene (*SUP4*) which contains an *Eco*RI cloning site. This gene complements an adenine mutation in the host (that gives the yeast an amber color). Transformation with *SUP4* restores the normal white color of the yeast. Hence the

interruption of the *SUP4* gene, which results in a recombinant molecule, is phenotypically visible with a second change of color of the yeast from white back to amber.

Another important feature is that it is a shuttle vector, that can be maintained both in yeast and as a high copy plasmid in *E. coli*. It has a bacterial origin of replication and an ampicillin-resistance gene as well. This facilitates isolation of large quantities of the vector from bacterial cultures for use as yeast cloning vectors.

Double digestion with *EcoRI* and *BamHI* will result in a linear vector ready to ligate with a piece of clonable DNA. The *EcoRI* digestion exposes the cloning site, while the *BamHI* digestion will remove the spacer fragment thereby exposing functional telomeres at either end of the vector.

Isolation of large pieces of clonable DNA is done by partial digestion of HMW DNA in agarose plugs and size-selecting digested DNA molecules by PFGE. These recombinant molecules are transformed into an appropriate yeast host. The yeast strain most commonly used as the YAC host is AB1380 (Genotype *MATa ade2-1 ura3 can1-100 lys2-1 trp1 his5 ψ⁺*). This YAC strain transforms with very good efficiency, and allows one to take full advantage of all the selective features built into the YAC vector. Once transformed into suitable yeast cells, these clones are maintained and carried over as native chromosomes. The application of techniques allowing the manipulation of cloned insert DNA usually requires a change in the yeast genetic background, to introduce desirable markers. Hence, techniques have been described to transfer YAC clones from AB1380 into other yeast strains (Spencer et al., 1994). However, AB1380 is the strain of choice for library construction and the need for other strains comes only when a YAC of interest has been identified.

Following the development of YAC vectors, YAC libraries have been constructed in a whole range of organisms. These include human (Anand et al., 1989; Albertsen et al., 1990; Dausset et al, 1992), mouse (Dietrich et al, 1995; Larin et al, 1993), rat (Cai et al, 1997), bovine (Libert et al, 1993) and porcine (Leeb et al, 1995) YAC libraries. Model

organisms like *C. elegans* (Coulson et al, 1988) and *Drosophila* (Garza et al, 1989) also have well characterized YAC libraries. Even pathogens like *Bacillus subtilis* (Sorokin et al, 1996) and parasites like *Plasmodium falciparum* (Camargo et al, 1997) have been analyzed using YACs. The biggest advantage is the almost limitless cloning capacity (Albertsen et al 1992, Guzman & Ecker 1988, Larin et al 1991, Heiter et al 1990). This facilitates the construction of physical maps of organisms (Burke, 1991). High resolution physical maps have been constructed in the regions of human chromosomes that are of importance because of the location of disease genes (Anand et al, 1991; Clark et al, 1995). The large size of the YAC clones has led to the construction of contigs that now cover the entire human genome map (Short, 1995). Improvised techniques have resulted in the construction of chromosome-specific YAC libraries (Korenberg et al, 1995; Green et al, 1995) to facilitate individual chromosome analysis in genome projects. Most recently YAC-STS (Sequence Tagged Sites; Green and Green, 1991) maps with a resolution less than 100 kb have been constructed in human chromosomes X (Nagaraja et al, 1997) and number 7 (Bouffard et al, 1997). This literally means that there is a land mark (STS) every 80 kb or so in these chromosomes. This was made possible mainly because of large insert YAC libraries. The positioning of Expressed Sequence Tags (ESTs), an important step towards assembling a functional map of any genome, has also been made possible by the use of YAC contigs (Gianfrancesco et al, 1997).

Several strategies can be applied to either link a YAC to a contig or to extract a refined physical map from a YAC. For chromosome walking experiments, two methods have been used to isolate extremities from the DNA insert. The first, plasmid rescue, takes advantage of the genetic markers in the YAC vector that may be selected for in bacteria. Unfortunately, with pYAC4, only the extremity next to the centromere can be rescued this way. The second method is based on PCR assay and several variants have been described. One is designed so that one primer recognizes the vector sequence next to the cloning site and the other primer recognizes a highly repeated sequence specific to the source genome (Nelson et al, 1989). However, this technique is not universally applicable since it relies entirely on the presence of a highly repeated within a short distance of the end. Another method uses a specialized adapter, the vectorette, that is ligated to the restriction digested

insert (Riley et al, 1990). By design and choice of primers, only fragments containing part of the vector (one primer) and the vectorette (the other primer) will give rise to a PCR product. Yet another technique relies on inverse PCR (IPCR). In this method (Silverman et al, 1989), the YAC is digested with a frequently cutting restriction enzyme, and individual fragments circularized. Junction fragments containing the cloning sites and a part of both the vector and the insert can be amplified with PCR using amplimers recognizing the vector. This again relies on the presence of two restriction sites within the size range of PCR amplification.

With the increasing use of YACs as genome analysis tools, techniques have been perfected for isolation of cDNAs and genes from YACs. This includes direct cDNA selection (Jacob et al, 1996; Snell et al, 1993; Parimoo et al, 1993) and exon trapping (Gibson et al, 1994). The former method consists of hybridizing the identified YAC with cDNA libraries. The latter method consists of selectively removing introns by using the eukaryotic splicing mechanism, resulting in a sort of exon amplification.. These techniques have greatly helped in isolating localized genes from large insert clones thereby avoiding the scanning of an entire chromosome or genome.

With the increasing use of YAC libraries the demand for efficient methods of library screening also increased. The biggest problem in YAC screening was the difficulty in obtaining sufficient amounts of pure YAC DNA in a relatively quick manner, from the thousands of clones. A solution came about with the development of automated procedures for DNA isolation from YAC clones (MacMurray et al, 1991). This method consists of isolating DNA from YAC clones arrayed in 96 well titer plates without taking them out of the plates. Filter, rather than plastic, bottoms are used for dialyzing in the reagents. The DNA thus isolated is suitable for PFGE, PCR and restriction enzyme digestion. A variety of methods are now available for screening of YAC libraries (Wang et al., 1994; Strauss et al. 1992; Green and Olson, 1990). Basically, however, YAC library screening is a three step process with slight variations (Albertsen et al, 1992). The first step is the PCR amplification of the library pools. Pools are made by mixing a large number of clones on the basis of their location on the plates. This first step helps to exclude a large part of the

library. The second step is the colony hybridization of the positive pools, after plating out the pools and transferring the clones to filters. Hybridization with the probe reveals the position of the positive clone. Alternatively, after plating out the pools, the individual clones can be picked into 96 well plates, further sub-pooled and amplified by PCR. The positive clone can be zeroed in by aligning the axes of the pools that give a PCR product of the expected size. The final step consists of verification of the candidate clone(s) by hybridization of the isolated YAC DNA to the probe.

YACs have proved their biological usefulness in transfection experiments (Peterson et al., 1993). YACs were successfully introduced into mammalian cell lines to complement enzymatic deficiencies. This application holds great promise for YACs. In the not too distant future, YACs may be used as vectors for gene therapy. As we shall see later, the applications of YAC transfer back to host cells holds exciting possibilities in plant genetics and might become a powerful tool in plant breeding.

As we have seen above the applications of YACs are many. However, there are some glaring disadvantages too. Technically, YAC library construction is cumbersome and time consuming. It is difficult to efficiently transform yeast especially with such large recombinant molecules. It will take a large number of transformations to complete an ample coverage of the genome so as to have a reasonable chance of isolating a given sequence. Also, a preferential transformation of smaller DNA fragments has been observed in YAC transformation (Connelly et al., 1991). For example a 330 kb YAC is 40-fold less efficient in transforming yeast than a 110 kb molecule. Addition of polyamines to transformation mixtures have been found to reduce this extreme size bias. Whatever be, these problems are purely technical and can be overcome by putting in more resources towards the library construction.

The serious problems with YACs are mainly two - chimerism and instability. Chimerism refers to the presence of fragments from different regions of the target genome (Libert et al 1993, Green et al 1991). This is especially true in very large insert (>500 kb) libraries (Anderson, 1993). This problem is detrimental to chromosome walking, since the

walk relies on accurate progress through contiguous segments of the genome. A mislead chromosome walk will complicate genome analysis to the extent that the gene of interest can be impossible to isolate. A large portion of the very large insert libraries are chimeric, some times up to a third of the total. This is very problematic, especially in case of complex genomes like those of plants, for which one cannot find many libraries with which to work. The other major problem, instability, refers to the failure of the host yeast to maintain the whole insert DNA intact. Deletions of large parts of the insert have been noticed in many YAC libraries (Anderson et al., 1993). These problems could be due to the inherent mitotic recombination mechanisms present in yeast. Parts of the insert could recombine with regions in native yeast chromosomes and get lost or added on to another part of the yeast genome. This could give rise to inter- and intra-fragment rearrangements. Another reason could be the ligation of different fragments of DNA to themselves before ligating to the vector. These problems not only complicate chromosome walks, but also are detrimental to isolation of cDNAs or genes from the insert segment. A lot of labor and time could be wasted.

Solutions have been proposed for the inherent problems with YACs. Most of the solutions lie in using recombination-deficient strains of yeast as hosts (Kohno et al., 1997; Haldi et al, 1994). Recombination-deficient *rad52* mutants and *rad51/rad52* double mutants have been used to reduce levels of recombination and thus chimerism in YAC cloning. It has also been proposed that transformation of intact yeast cells instead of the conventional spheroplasts can greatly reduce chimerism (Heale et al., 1994). Understandably chromosome specific libraries will have a lower level of chimerism since the chances of interference of other parts of the genome are far less.

Despite all the problems YACs remain the most powerful tool, especially in human genome analysis. This is mainly because of the availability of a number of libraries. The greater the number of libraries available, the lesser will be the problems caused by the inherent disadvantages. The advantages outweigh the disadvantages by far. However, in case of many other organisms, where multiple libraries are not available, scientists are moving towards other HMW DNA cloning vectors.

Plant YAC libraries

YACs found applications in plant genome analysis slightly later than in other organisms. The complexity of the genomes and the presence of large regions of repeated DNA sequences probably deterred researchers from making an early attempt. The first plant YAC library was made in 1990 (Ward and Jen, 1990) of the model plant *Arabidopsis thaliana*. The presence of long stretches of single copy sequences was one of the rationales behind this. Since then at least three YAC libraries have been constructed, in *Arabidopsis* (Ecker, 1990; Grill and Somerville, 1991; Creusot et al., 1995). In addition, YAC libraries have also been constructed in rice (Umehara et al., 1995), tomato (Martin et al., 1992; Bonnema et al., 1996), maize (Edwards et al., 1992), sugarbeet (Eyers et al., 1992; Klein et al., 1994), and barley (Kleine et al., 1993). However, the average insert size of most plant YAC libraries is lower than that of human or other mammalian YAC libraries. This might be because of the difficulty in isolating good quality HMW DNA to start with. In this regard, two plant YAC libraries stand apart from the rest. The *Arabidopsis* YAC library constructed at INRA, France (Creusot et al., 1995) has an average insert size of 420 kb, the highest among all the plant YAC libraries. This will be an invaluable resource for genome analysis of *Arabidopsis*. The second one is the rice YAC library (Umehara et al., 1995) which has an average size of 350 kb. The remarkable progress made in the genome analysis of these two plants is due to the availability of such high quality libraries. The average insert size of most other plant YAC libraries is around or less than 200 kb. It should be pointed out here that the average insert size of most human and mouse libraries is greater than 500 kb.

Some of these YAC libraries have been successfully used for map-based cloning of genes and physical mapping of genomes. Needless to say, most of the achievements have occurred in *Arabidopsis* and rice. Examples of plant genes cloned thus have already been described in a previous section dealing with the positional cloning of plant genes. In most of the cases YAC libraries were used.

Physical mapping has progressed enormously. Thirty-one YAC contigs, covering the majority of *Arabidopsis thaliana* chromosome 5, have been described (Schmidt et al., 1997). In order to generate a physical map of chromosome 5, 142 molecular markers mapping to chromosome 5 were used to screen four *Arabidopsis* YAC libraries. This screening gave 634 YAC clones, which were arranged into 31 contigs. Genetic mapping allowed positioning of these contigs along chromosome 5. However, a high proportion of clones were found to contain chimeric inserts. An 800 kb contiguous region, encompassing a 3.5 cM region of *Arabidopsis* chromosome 1, has been isolated using YAC clones (Vijayaraghavan et al., 1995). The *apetalal* mutation that affects floral meristem identity is in this region of the chromosome. An RFLP marker closest to the mutation was used as a starting point for contig generation. The contig was built over 19 characterized YAC clones. This contig was used to correlate the physical and genetic maps for the region and it was concluded that 1 cM corresponds to approximately 200 kb. Repeated DNA sequences have been analyzed in two YAC libraries of *Arabidopsis* (Schmidt et al., 1994). Ribosomal, chloroplast and the paracentromeric repeat sequences were found to be differentially represented in the two libraries. One of the libraries was found to be highly chimeric because individual clones carried unique and repetitive sequences from different parts of the genome. It was also found that YAC clones carrying tandemly repeated sequences were highly unstable. These observations have implications for the construction of physical maps. A 1,700 kb contiguous region on *Arabidopsis* chromosome 5, which includes a 300 kb region containing the flowering-time gene *CO*, has been isolated (Putterill et al., 1993). This required the detailed analysis of 67 YACs, from which 87 end probes were isolated and examined in chromosome walking experiments. DNA fragments isolated from YACs were used as RFLP markers to localize *CO* to a 300 kb region within the cloned DNA.

One third of the *Arabidopsis* genome has been covered in YAC clones with respect to map position using the first three YAC libraries (Hwang et al., 1991). YAC clones corresponding to 125 *Arabidopsis* RFLP markers have been identified. At least one YAC clone has been isolated for each of the RFLP markers tested. This is the first step towards assembly of an overlapping YAC library of the *Arabidopsis* genome.

The other plant species to be analyzed extensively using YACs is rice. The extensive analysis is possible because of a well characterized large insert YAC library. Chromosome numbers 2, 4, 7, 8 and 9 have been extensively characterized. Most of the work has been done as part of the Rice Genome Research Program. For chromosome 2, a total of 128 RFLP markers and 4 sequence tagged-sites (STS) markers located on the high-density genetic map were used for YAC clone landing (Umehara et al., 1997). A total of 239 individual YACs were selected to form 43 contigs encompassing about 50% of the length of chromosome 2. Similar procedures were followed for the other chromosomes. A 17.1 cM region of chromosome 4 was covered in 16 contigs encompassing 62 of the 258 YACs selected (Koike et al., 1997). The other YACs were arranged to 23 positions. On chromosome 7, 203 individual YACs were landed on 109 DNA markers. Sixty-four out of the 203 YACs formed 15 contigs which covered a total length of 21.8 cM and 139 YACs were localized to 26 positions (Koike et al., 1997). In chromosomes 4 and 7 also, 50% of the total length has been covered in YACs. In chromosomes 8 and 9, a total of 252 YACs were confirmed to contain the mapped marker DNA fragments on both chromosomes (Antonio et al., 1997). A contig map was constructed by ordering these YACs and about 53% and 43% genome coverage was obtained for chromosomes 8 and 9, respectively.

Plant YACs have also found applications in cytogenetic analysis and chromosome mapping, mainly by way of Fluorescent *In situ* Hybridization (FISH). *In situ* localization of short and low copy sequences is difficult in plants, mainly because of the complexity of the genome and difficulty in obtaining good chromosome preparations and probes. Because of their large size YACs are excellent probes for chromosome mapping using FISH. A partial soybean YAC library was screened for low copy sequences and four YACs were chosen for FISH (Zhu et al., 1996). By using degenerate oligonucleotide-primed PCR (DOP-PCR), relatively small amounts of YAC DNA were used as probes for FISH. Three of the four YACs gave simple hybridization patterns consistent with a single homologous locus. YACs have also been *in situ*-localized in tomato and potato metaphase chromosomes (Fuchs et al., 1996). Two YACs specific for a single copy marker on the long arm of the NOR-chromosome 2 of tomato were selected as probes. Hybridization of these to tomato

and potato chromosomes showed clear single locus signals. These experiments can be invaluable in the chromosome level mapping of complex plant genomes.

As in case of other YACs, the ultimate use of plant YACs will also be in the ability to be transformed back into the host plant cells for phenotypic complementation. Complex biosynthetic pathways in plants, yielding products of commercial value, are often present as multi-gene pathways. Individual steps in these pathways are often controlled by gene families that are clustered together in large fragments of DNA. Transformation of a plant with the entire gene family would mean transforming with large fragments of DNA. Conventional transformation procedures can transform only small (<25 kb) fragments of DNA. The utility of plant YAC libraries would be dramatically increased if these YACs could be used directly for plant transformation. Attempts are underway to make this feasible (Adam et al., 1997). A pair of vectors that allow clones from YAC libraries to be modified for plant transformation by direct DNA transfer methods such as particle bombardment or electroporation, has been developed. Modification of the YAC has been achieved in two sequential yeast transformation steps taking advantage of the homologous recombination system in yeast. Two plant selectable marker genes and sequence elements required for copy number amplification in yeast were introduced into YACs. This resulted in the retrofitting of YACs ranging in size from 80-700 kb. The many fold increase in copy number of these modified YACs facilitated the isolation of YAC DNA sufficient for direct transformation. This procedure should help to bridge the gap between existing mapping and plant transformation procedures.

Plant YACs have found applications, especially in selected species like *Arabidopsis* and rice. The inherent problems with YACs have also been found in plant YAC libraries. The availability of many or high quality libraries have contributed to the success of *Arabidopsis* and rice genome analysis. However, researchers are starting to analyze a greater number of plant species, especially crop species with large and complex genomes. The option to put resources into the construction of a YAC library is no longer very attractive not only because of the inherent problems but also because of the time and labor consuming procedures. Hence the importance of developing alternative HMW DNA

cloning vectors is increasing. As we will see shortly many of the crop plants now have libraries constructed in other vectors. However, for sheer insert size and utility in chromosome walking, YACs are still the most useful tools available.

Bacterial Artificial Chromosomes

The problems that complicate genome analysis with YACs prompted researchers to look for alternate HMW DNA cloning systems. The biggest breakthrough in this direction came about with the designing of Bacterial Artificial Chromosome (BAC) vectors (Shizuya et al, 1992). The main problem with the maintenance of large foreign DNA in bacteria molecules is the inaccurate replication mechanism of *E. coli*. This inaccuracy is again dependent on the copy number of the plasmid. It is difficult for the bacteria to faithfully replicate a large number of plasmids within its short generation time. The higher the copy number of the plasmid, the greater are the risks of inaccurate replication of the insert molecules. Shizuya and colleagues took advantage of the bacterial F plasmid. This is a sex plasmid that controls conjugation, the bacterial mode of sexual reproduction. Because of its important function, this plasmid, which is maintained as single copies in the cell, is replicated faithfully by the bacteria. These properties are ideal for cloning large DNA molecules into bacteria. Initial experiments revealed that this system could maintain human genomic DNA fragments as large as 300 kb. Individual clones were maintained with a high degree of structural stability even after 100 generations of serial growth. The cloning efficiency was found to be almost ten-fold that of YACs. Moreover, it was found that the cloned piece of DNA could be treated as any other plasmid DNA, which makes isolation and manipulation of the cloned piece of DNA much easier than in YACs. All these properties made researchers take notice of this cloning system. Soon BACs became a system of choice for researchers interested in genome analysis.

The method of transformation of *E. coli* also facilitates large scale library construction using the BAC system. Electroporation, the method of transforming host cells with foreign molecules using high-voltage electric current, has already been established in *E. coli* (Dower et al., 1988; Taketo 1988). It has been found very effective in

transformation of large DNA molecules as well (Sheng et al., 1995). DNAs ranging in size from 7-240 kb were used, as well as BAC ligation mixtures containing a range of different sized molecules. Even though the efficiency of transformation was found to be strongly dependent on the strain of *E. coli* used, even for host strains which transform relatively well with large DNA, transformation efficiencies dropped dramatically with increasing size of DNA. Molecules of 240 kb transformed approximately 30-fold less well, on a molar basis, than molecules of 80 kb. Maximum transformation of large DNA occurred with different voltage gradients and time constants than the ones that are optimal for smaller DNA. This provides the opportunity to increase the yield of transformants which have taken up large DNA. This shows that conditions may be selected which increase the average size of BAC clones and give good transformation. The method of transformation by electroporation is easy and quick and can give as many as a thousand BAC transformants per transformation and an entire library per ligation mixture, for organisms with relatively small genome sizes.

The BAC vector of choice is pBeloBAC11 (Fig.1-4). Based on the F-plasmid, additional features have been introduced to facilitate easier cloning.

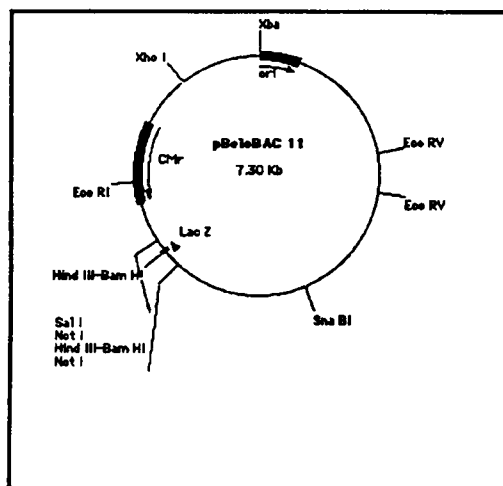


Figure 1-4: pBeloBAC11 vector. (see text below for description)

The cloning site of this 7.4 kb vector is either a *Bam*HI or a *Hind*III restriction site. The cloning site is in the middle of a *lacZ* reporter gene that enables use of the bacterial blue-white color selection for recombinant clones. Immediately flanking the cloning site there are two *Not*I sites.

*Not*I, being a rare cutter, is expected to digest the entire insert in either a single piece or a few pieces. This digested DNA, when subjected to PFGE can easily reveal the size of the insert. A selectable marker gene, the antibiotic chloramphenicol, has also been introduced into the vector. A few other restriction sites are

also present, enabling operations such as end rescue and inverse PCR for obtaining end sequences.

As in case of YACs, a number of BAC libraries have been constructed in different organisms. In case of the human genome the availability of a number of well-characterized YAC libraries delayed the construction of a BAC library of the entire genome. However, well-characterized BAC libraries of the human genome are currently available (Kim et al., 1996; Asakawa et al., 1997). As the human genome project split into the detailed study of individual chromosomes, the construction of chromosome-specific BAC libraries was necessitated (Wang et al., 1994, Kim et al., 1995). Bovine (Cai et al., 1995), murine (George et al., 1997) and chicken (Zimmer and Gibbins, 1997) BAC libraries have also been constructed. While most of these BAC libraries have average insert sizes around 150 kb, the chicken BAC library has an astonishing average size of 390 kb. Clones in this library range in size from 25-725 kb. This is the first report, not only of such high average insert size, but also of BAC clones harboring inserts in excess of 500 kb. This clearly indicates that the BAC system is capable of harboring DNA inserts in the YAC size range.

The applications of BACs are in the same areas where YACs have found applications. Physical mapping is obviously the most important application of such large insert libraries. A pilot study for the application of BACs to the generation of contiguous physical maps was conducted in the human ryanodine receptor (*RYR1*) gene region (Rouquier et al., 1994). The total genomic BAC library was screened with a complex cosmid probe specific for the gene. One 150 kb BAC was found positive. This BAC hybridized to a 150 kb overlapping set of nine chromosome-19 specific cosmids at the 3' terminus of the gene as well as to two established cosmid contigs. This suggested that the 150 kb BAC was non-chimeric and can be used effectively in mapping strategies.

A BAC-based framework contig map has been constructed for human chromosome 22q (Kim et al., 1996). The map features 613 BAC clones. The BAC-based map provides immediate access to clones that are stable and convenient for direct genome analysis. The approach to rapidly developing marker-specific BAC clones is relatively straightforward

and can be extended to generate scaffold BAC contig maps of the rest of the chromosomes. These contigs will provide substrates for sequencing the entire human genome. A 750 kb contig has been constructed using BACs in the region of human chromosome 21 containing the progressive myoclonus epilepsy region (Stone et al., 1996). The restriction sites for a number of restriction enzymes have also been determined resulting in a high-resolution restriction map of the region.

A BAC contig covering the 3.5 Mb of the Down syndrome congenital heart disease region (DS-CHD) in human chromosome 21 has also been constructed (Hubert et al., 1997). A BAC library was screened using several YACs spanning the region. FISH confirmed the location of all BAC clones to that particular region of the chromosome. In another important work, a 1.1 Mb human alpha/delta T-cell receptor locus was analyzed with BACs (Boysen et al., 1997). The region was mapped and partially sequenced with BAC clones. Seventeen BAC clones were isolated covering the region. Of these five were sequenced directly using the shotgun approach. The ends of the 17 BAC clones were distributed evenly across the region. This work led to the proposal of a new approach to sequence the human genome.

Other applications of BACs include use as probes in FISH studies (Cai et al., 1995; Hanson et al., 1995) and in transformation studies (Mejia and Monaco, 1997; Baker and Cotton 1997). As in case of YACs, the functional characterization of a gene contained within a BAC require the transferring of the BAC into the original eukaryotic cell, either for transient or long term expression. The *NotI* restriction site was used to introduce the reporter gene, *lacZ*, together with a selectable marker (Mejia and Monaco, 1997). In another modification, Baker and Cotton (1997), used two series of BAC vectors bearing marker genes for luciferase and Green Fluorescent Protein (GFP). BACs as large as 170 kb were successfully transfected into mammalian cells. The easy isolation of pure BAC DNA coupled with modified transformation vectors should make BACs even more useful in gene expression studies.

As stated before, the availability of well-characterized YAC libraries has limited the use of BACs in human genome analysis. In many cases BACs are used as a back-up to YAC-based analysis. In this regard also BACs have proved their worth. The potential of BACs is just beginning to be realized. In a few years time a lot of genome analysis and sequencing work will be based on BACs.

Plant BAC libraries and their applications

Bacterial Artificial Chromosomes have made a large impact in plant genome analysis. The difficulty in constructing and the problems associated with YAC libraries have forced many plant researchers to look at alternative cloning systems. Moreover, the availability of multiple libraries is a remote possibility for most plants, with the exception of *Arabidopsis*. BACs have come in handy for this purpose because of the ease in construction and insert sizes comparable to that of YACs. In fact it is not wrong to say that BACs have made a greater impact in plant genome analysis than in mammalian genome analysis.

BAC libraries are now available or are under construction in a whole range of plant species. The first plant BAC library to be constructed was in *Sorghum bicolor* (Woo et al., 1994). The library, made from sorghum protoplast DNA contains 13,440 clones with an average insert size of 157 kb. Sorghum inserts of up to 315 kb were isolated and shown to be stable when grown for over 100 generations in liquid media. No chimeric clones were detected as determined by FISH of ten BAC clones to metaphase and interphase *S.bicolor* nuclei. The library was screened with six sorghum probes and three maize probes and all but one sorghum probe hybridized to at least one BAC clone. It was also estimated that there is a 91% chance of isolating a specific genomic region from this library.

A rice BAC library has been constructed from nuclear DNA (Wang et al., 1995). The library consists of 11,000 clones with an average insert size of 125 kb. The library was evaluated for the presence of organellar, repeated and telomeric sequences. Since it was made from nuclei, the library consists of a very low percentage of chloroplast and mitochondrial DNA clones. Eighteen BACs were identified that hybridized with an

Arabidopsis telomere repeat. Sixteen BACs hybridized with a genome-specific repetitive sequence. Twelve clones were isolated that hybridized with 3 DNA markers linked to the *Xa-21* disease-resistance locus. Later, these BACs were used in the isolation of the rice blast resistance-gene *Xa-21* (Song et al., 1995). This BAC library was also used to map the rice genome with BACs using FISH (Jiang et al., 1995). This is also a classic case of the use of FISH in plants. The application of FISH has been limited in plant species, especially for the mapping of single or low copy sequences, due to the inconsistent signal production in plant chromosome preparations. Jiang et al. showed that repetitive sequences in BAC clones can be suppressed efficiently by using rice genomic DNA as a competitor in the hybridization mixture. BAC clones as small as 40 kb were successfully mapped.

Two other cultivar-specific BAC libraries have also been constructed in rice (Zhang et al., 1996). The libraries were constructed from the parents of a permanent recombinant inbred mapping population (Lemont and Teqing). The Lemont library consists of 7,296 clones with an average insert size of 150 kb, while the Teqing library consists of 14,208 clones with an average insert size of 130 kb. The libraries cover 2.6 and 4.4 rice genome equivalents respectively. Three single copy DNA probes were used to screen the libraries and at least two overlapping BAC clones were isolated for each probe from each library, with sizes ranging from 45-260 kb. The simultaneous construction of two libraries is the proof for the ease of construction of BAC libraries.

An 800 kb contig has been constructed in the rice blast resistance gene *Pi-ta* region (Nakamura et al., 1997). A rice BAC library was constructed from green leaf protoplasts of the cultivar Shimokita, which harbors the *Pi-ta* gene. The average insert size of 155 kb and the library size of seven genome equivalents makes it one of the most comprehensive BAC libraries available. By probing with markers close to *Pi-ta2* (closely linked to *Pi-ta*) and *Pi-b* a contig of about 800 kb consisting of 19 clones was constructed in the *Pi-ta* region. This region has a high frequency of repetitive sequences. To circumvent this difficulty, a two-step walking method was devised, one for the repetitive genomic regions and the other for the normal genomic regions. FISH with a BAC clone from the *Pi-b* region (non-repetitive) yielded clear signals on the long arm of chromosome 2, while a clone from the *Pi-ta2*

region (repetitive) showed various cross-hybridizing signals near the centromeric region of all chromosomes.

BACs have been used to demonstrate large regions of genomic colinearity among grass species (Chen et al., 1997). Rice and sorghum genes homologous to the *sh2* locus maize were cloned into BACs and observed that a homologue of the maize *al* gene was also present on each of those BACs. In sorghum, a direct duplication of the *al* homologues was found, separated by about 10 kb. In maize, the *sh2* and *al* are approximately 140 kb apart and transcribed in the same direction, with *sh2* upstream of *al*. This arrangement was found fully conserved in rice and sorghum. However, these two loci were found separated by only 19 kb in both rice and sorghum. These results indicated the conserved gene order, but extensive sequential difference between these genes, among many cereal crops.

Genome analysis in legumes has been slow, when compared to cereals. The lack of large scale international collaborative projects, like the *Arabidopsis* and rice genome projects, and the relatively fewer number of researchers has hampered rapid progress. However, a lot of progress has been made in the last couple of years in acquiring sufficient resources for positional cloning. The first BAC library in legumes was constructed in soybean (Marek and Shoemaker, 1996). This library, made from the cultivar Williams 82, consists of 40,000 clones with an average insert size of 150 kb and thus represents 4-5 genome equivalents. This was followed quickly by two more BAC libraries in soybean. To isolate nematode resistance genes and study plant-pathogen interactions, a soybean BAC library was constructed from the cyst nematode-resistant variety "Faribault" (Danesh et al., 1997). This library covers the genome 3 times with an average insert size of 120 kb. Detailed analysis of a small number of clones indicates that the library will be useful for physical mapping and chromosome walking. Yet another soybean BAC library has been completed (Salimath and Bhattacharya, 1997). The library was constructed from HMW nuclear DNA of the near-isogenic line Williams 82, which carries a root and stem rot disease-resistance gene *Rps1-k*. The average insert size is comparatively low (105 kb), but the library covers the genome 4 times. The library was screened for 3 RFLP markers linked to *Rps1-k*, one *PI-PLC1* specific cDNA and two barley photosynthetic genes, using a

newly described method based upon the RFLP analysis of BAC pools. This method allowed the identification of BAC pools carrying the *Rps1-k* linked polymorphic copy of the middle repetitive sequences. This method can also be applied for the identification of a specific gene of a multigene family if a gene-specific probe is available. This led to the identification of one BAC pool that carried the *PLC1* gene of the *PI-PLC* multigene family. Additionally the library was screened for the presence of ESTs and clones carrying chloroplast genes.

Arabidopsis has well-characterized YAC libraries available and this has obviated the need for a BAC library. Still, a BAC library was constructed recently in *Arabidopsis*, with an average insert size of 65 kb (Wang et al., 1996). Because of the low insert size this library will not find much use in direct genome analysis, but it will still be a valuable resource to complement the YAC libraries and fill gaps that have not been covered in YACs. With an international effort to sequence the entire genome of *Arabidopsis* on the way, we can expect to see more BAC libraries that can be used as substrates in sequencing.

In addition to those that have been completed, BAC library projects have been initiated in a number of crop plants. There have been reports of BAC library projects in wheat, common bean (*Phaseolus vulgaris*), the model legume *Medicago truncatula*, fruit crops like apple, citrus and grapes, lettuce and cotton (Plant Genome IV, The International Plant Genome Conference). This signifies the increasing emphasis on BACs for plant genome analysis.

A striking development in plant transformation using BAC-cloned DNA has increased the utility of BACs to the point of direct functional complementation studies (Hamilton et al., 1997). A new binary BAC (BIBAC) vector has been developed based on the BAC system and on the binary vector system for *Agrobacterium*-mediated plant transformation. It has been shown that this vector is capable of transferring at least 150 kb of foreign DNA into a plant nuclear genome. The transferred DNA appeared to be intact in the majority of transformed tobacco plants analyzed and was shown to be faithfully inherited in the progeny. This kind of HMW DNA transformation has far reaching

implications on the genetic engineering of plants with multi-gene clusters which are scattered around hundreds of kilobases of DNA. This could give rise to a whole range of new commercial products from plants. The key for all this will be the cloning capacity of BACs.

The high genomic complexity of most cultivated plants and limited availability of time and resources makes it very difficult to achieve optimum coverage of these genomes in YACs. In the years to come BACs will be the major HMW DNA cloning system in the genome analysis of most of these plants.

Nodulation and Symbiotic Nitrogen Fixation

Even though nitrogen is the most abundant element in the atmosphere (78%), most of it is inert and cannot be used directly for any biological purpose. The phenomenon of symbiotic nitrogen fixation is unique in that it is the only process by which a living organism utilizes atmospheric nitrogen directly by converting it into compounds that can be absorbed. This occurs by way of a symbiotic relationship between certain groups of bacteria and plants. In the most commonly found case of symbiosis, eubacteria of the family Rhizobiaceae (*Rhizobium* and *Bradyrhizobium*) interact with plants belonging to the legume family Fabaceae. The bacteria live in specialized plant root structures called nodules. There they metabolize atmospheric nitrogen. The plant supplies the bacteria with photosynthates and nutrients needed for their growth and development. In return the bacteria supplies the plant with ammonia, the product of biological nitrogen fixation, the supply of which would otherwise be growth-limiting.

The basic processes in legume nodulation have been deciphered. The legume plant roots secrete diffusible chemical signals, generally flavonoids but also betaines (Phillips et al., 1992), that attract specific bacteria and induce rhizobial *nod* genes. The bacteria then releases a signal, Nod factor, which is a substituted lipooligosaccharide consisting of β -N-acetyl glucosamine subunits. The nod factors induce a response in the plant roots. The response begins with root hair curling or deformation and cortical cell division. Subsequent

nodule development in legumes follows one of two possible paths - determinate nodule forming and indeterminate nodule forming.

In determinate nodule-forming hosts, like *Phaseolus*, *Glycine* and *Lotus*, cell divisions occur first in the outer cortex, just below the epidermis. Later, the inner cortical cells divide. The two meristematic regions eventually coalesce, forming the nodule primordium. The nodule meristem stops its activity early in nodule development and for the most part, new cells already contain bacteria because they are formed by division of infected cells (Rolfe and Shine, 1984). In indeterminate nodule forming hosts like *Medicago*, *Vicia* and *Pisum*, cell divisions occur in the inner cortex. A nodule primordium is established from the derivatives of these inner cortical cell divisions. A nodule meristem is initiated apically or to one side of the nodule primordium. Cell divisions in this meristem continuously add new cells to the different nodule tissues. These cells differentiate and some of them subsequently become infected by *Rhizobium*. As a result a developmental gradient of infected cells is established along the nodule axis (Vasse et al., 1990).

Rhizobia attach to the deformed root hairs and if compatible with their host, invade the root hair cell via an inward invagination of the host plasma membrane followed by new cell wall synthesis. This infection thread proceeds from cell to cell allowing the spread of the rhizobia to the cells of the nodule primordium. The rhizobia are then released from the thread into an infection droplet. Later the bacteria become individualized by pinching off of the host plasma membrane (Verma, 1992). The response finally results in the formation of a nodule housing the bacteria. Variations to this generalized theme, especially in the mode of bacterial entry, abound in legumes.

The mechanistic and molecular aspects on the bacterial side of this process have been extensively studied and significantly unveiled in recent years. However, the plant side of the process is not very well understood. Although legume research has progressed, the molecular mechanisms triggering and controlling the differentiation of nitrogen-fixing nodules remain rather obscure, especially those of the signal-response pathways that lead from the first nodulation signal to a differentiated nodule. What is understood is that the

legume host has a major role in regulating the overall process of nodulation and nitrogen fixation (Caetano-Anolles and Gresshoff, 1991). In certain species of legumes it was discovered that some plants form nodules in the absence of *Rhizobium*. This demonstrates that the entire genetic complement needed for nodule development is plant-based. The plant determines the nodule morphology and other associated physiological processes (Gresshoff, 1993). However, systemic feedback regulation of nodule formation, called autoregulation, is also partly determined by the plant root (Gresshoff, 1993). Another regulatory mechanism restricts the number of nodules in the presence of nitrate. All of this convincingly proves that the nodule forming mechanism is completely within the plant system.

Bacterial signal molecules, the Nod factors, can cause nodulation-like responses in plant hosts even in the absence of the producing bacteria. In addition to altered root hair growth and initiation of cell division these responses also include expression of a class of nodule-specific plant genes, the early nodulins or *ENODs* (Hirsch, 1992; Spaink, 1995). Most recent work relates to three enduring research themes - the production of new or upregulated transcripts (*ENODs*) in the plant during early nodulation, the involvement of plant hormones, and the pattern of cell division and cell rearrangement early in nodulation response.

For about two decades, the molecular analysis of legume nodules has been focused on the cloning and characterization of nodulins. Several nodulin cDNAs have been isolated, representing genes that are induced during early and/or late stages of nodule development and functioning. The early nodulins are involved in the infection process and in establishing the nodule structure, whereas late nodulins have specific roles in nodule functioning. The temporal and spatial expression patterns of early nodulins suggest potential functions. The earliest expressed nodulin is detected 72 hours after inoculation but is transcribed in stem tissues as well (Kouchi and Hata, 1993). In contrast, cell-cycle related genes such as those for cyclin and p34 kinase show increased mRNA 24 hours after inoculation (Deckert et al., 1994). Late nodulins are usually nodule-specific duplicates of genes with known metabolic functions in other plant organs. These include uricase and

glutamine synthetase. Plant genes encoding products involved in the interaction with the bacteria have also been identified. These include nodulin-26, which encodes a protein that is involved in metabolite exchange (Weaver et al., 1994). Moreover, studies on the oxygen transport protein leghemoglobin, the first nodulin to be characterized biochemically (Appleby, 1984) have provided insight into the phylogeny of nodulated plants. It has been found that the transcription factors responsible for the expression of leghemoglobin in the infected cells are conserved between legumes and non-legume nodulating plants like *Casuarina*. The isolation and characterization of even earlier nodulins by novel methods such as differential display, may ultimately elucidate the relationship between bacterial nod factors and activation of plant early nodulins.

There is promising evidence to suggest that early events in nodulation are determined by an interplay of plant hormones. It has been shown that early nodulation can be induced by auxin transport inhibitors (Verma et al., 1992). It has also been shown that hormones can supplement mutant non-nodulating rhizobacteria and enable them to induce nodules (Cooper and Long, 1994). Further evidence can be inferred from the fact that some nodule-specific genes like the hydroxy-proline rich cell wall protein ENOD2 are induced by cytokinin. Alternatively, *Rhizobium* signals could mimic a class of native plant hormones. The signals, with their chitin backbone, could have co-opted a plant defense response pathway that was originally responding to chitinaceous elicitors. All these evidences have prompted researchers to look deeper in this direction.

Early signal transduction events in epidermal or other cells may account for the coordination of diverse host responses. Little is known of the early events in root hairs, but several studies have explored ion movements as an indicator of cellular activity related to signal perception (Ehrhardt, 1992). Root hairs were found to display a characteristic depolarization of the cytoplasmic membrane in response to the application of Nod factors. Later, this depolarization was found to be because of the movement of Ca^{2+} ions (calcium spiking). Mutants that lack both root hair curling and cell division, also showed no calcium spiking. Thus the movement of calcium ions is a major point of investigation in the study of early cell responses in nodulation.

To get more insight into the processes controlling nodule formation it is essential to integrate a genetic approach. In the past two decades, legume mutants showing a changed nodulation phenotype have been isolated from a variety of species (Caetano-Anolles and Gresshoff, 1991). These studies have demonstrated that mutants affected in every step of nodule development can be obtained. Unfortunately most of these mutations have been generated in legume species that are not easily amenable to positional cloning strategies. With the development of model legumes and of better resources for positional cloning, many of these mutations can be put to use for isolating the corresponding genes.

Nitrate tolerant symbiosis and supernodulation

As mentioned before it has been noted that there is systemic feedback regulation of nodule formation. Yet another regulatory mechanism controls the number of nodules formed when nitrate is present in the soil. The plant senses that there is enough nitrate available and "decides" that there is no need to put forth nodules. A study of this regulatory mechanism can give deep insight into the general mechanism of nodule initiation. Mutants that are deficient in this mechanism have been isolated (Carroll et al., 1985a). *Glycine max* cv. Bragg seeds were mutagenized chemically using ethyl methanesulfonate. The first generation after mutagenesis (M_2) of these seeds was screened for increased nodulation under high nitrate conditions. Fifteen independent nitrate-tolerant symbiotic (*nts*) mutants were obtained from 2,500 M_2 families. When cultured on sand with KNO_3 , nodule mass and nodule number in the mutant lines were found to be several fold higher than those of the wild type cultured under the same conditions. Inheritance of the *nts* character in the subsequent generations was tested in 10 mutant lines. Mutant *nts*382 was also found to nodulate more than the wild type in the absence of nitrate. However it was also found that both the mutant and the wild type were same in nitrogen assimilation. This convincingly proves that the *nts* mutation is affecting a regulatory gene involved in nodule development.

The nodulation characteristics of the *nts*382 have also been described (Carroll et al., 1985b). The mutant was found to nodulate significantly more than the parent cultivar Bragg in the presence or absence of several combined nitrogen sources. Mutant plants were small

in size. It was also found that even though *nts382* is a supernodulator, inoculation with *Bradyrhizobium* was necessary to induce nodule formation. Segregation of M_3 plants indicated that the mutant character is inherited as a Mendelian recessive.

Ever since the *nts-1* mutant was isolated, a lot of work has been done on the molecular genetic characterization of the mutation. It was noticed that the recessive *nts* mutation segregates as a single Mendelian locus and is unlinked to two separately isolated non-nodulation (*nod*) controlling loci (Mathews et al., 1990). The non-nodulation loci *nod49* and *nod139* were found to epistatically suppress supernodulation in doubly recessive mutant lines. The genetic locus containing the *nts* mutation was mapped tightly to a molecular marker, pA-132 (Landau-Ellis et al., 1991a). The mapping was done using an RFLP detected by a subclone, pUTG-132a. The large number of RFLPs between *nts382*, which is homozygous recessive for the *nts-1* allele, allowed the detection of co-segregation of several diagnostic markers with the supernodulation locus in the F_2 families. The *nts-1* locus was detected on the RFLP linkage group E, directly next to pA-132. This very close linkage of the marker and the *nts-1* locus allows the application of this clone as an entry point for the isolation of the *nts-1* gene.

Further analysis of the pA-132 probe revealed that the probe, which was generated as a random *Pst*I fragment, was not unique (Landau-Ellis and Gresshoff, 1994). It was found to contain three *Pst*I fragments. It was determined that the largest insert (1.8 kb) was the fragment detecting the linked polymorphism. When this insert (pUTG-132a) was used as a probe a simple pattern was detected showing no recombination and therefore 100% linkage in the *nts* F_2 segregants. Further evidence for this came when linkage analysis of another supernodulating mutant *nts1007* showed no recombination with the same pUTG-132a marker (Landau-Ellis and Gresshoff, 1992). This suggests that the *nts* locus is either very close to pUTG-132a or that this region of the genome is characterized by low recombination. Molecular analysis of two nearby RFLP markers, after PFGE of genomic soybean DNA, suggest that both markers are on the same 400 kb fragment. Another marker, which is 1.8 cM away, could be detected on the same 900 kb fragment as the above two markers. So, this general region may not have a restriction in recombination, as

0.5% recombination could not have been detected in the original establishment of map distances.

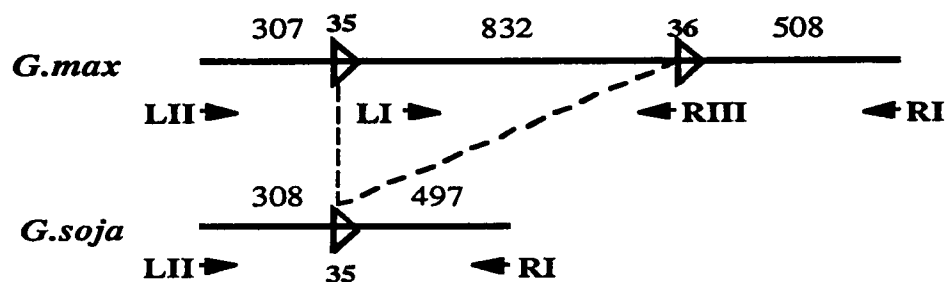


Figure 1-5: The arrangement of pUTG-132a.

The arrangement of pUTG-132a differs in *G. max* and *G. soja* (the numbers LI, LII, RI and RIII indicate the primers designed to amplify the sequence)

Genetic distance between pUTG-132a and *nts-1* has been estimated (Kolchinsky et al., 1996). From that work it is quite reasonable to assume that pUTG-132a is equal to or less than 1 cM away from *nts-1*. Physical mapping in a nearby region suggests that one centimorgan represents about 500 kb (Funke et al., 1993). Hence pUTG-132a could be lesser than 500 kb away from the *nts* gene. Further characterization of the 1.72 kb *Pst*I pUTG-132a fragment revealed that it consists of a 823 bp insertion in *G. max* relative to its wild-type progenitor *G. soja* (Kolchinsky et al., 1996; Fig 1-5). The insertion was found to be flanked by a 35 bp direct duplication. Data suggests that the pUTG-132a sequence is found only once in the genome and this is compatible with the recessive nature of *nts-1*. All this suggests that pUTG-132a is a valuable resource for map-based cloning.

***Lotus japonicus* - A model legume**

Arabidopsis thaliana has been adopted as a model plant for studying the molecular level functioning of plants. However it lacks biological nitrogen fixation and hence cannot be used in the study of nodulation. The importance of legumes, not only as sources of food all over the world but also as forage and rotational crops, has prompted a need to study the molecular details of nodulation and nitrogen fixation. Unfortunately most of the cultivated legumes are not amenable to molecular dissection, because of their poor genetics

(tetraploidy or large genome size). Thus sprung up the need for a "model legume". Two legumes, *Lotus japonicus* (Handberg and Stougaard, 1992) and *Medicago truncatula* (Barker et al., 1990) are the front runners in this quest. As a model plant *L. japonicus* ecotype "Gifu" B129 has many advantages (Handberg and Stougaard, 1992; Jiang and Gresshoff, 1993). It has a haploid genomic size of approximately 400 megabase pairs ($x=n=6$), which corresponds to about 0.5 pg of DNA per cell. The genome size equals about 4 times the size of the *Arabidopsis* genome, but is still lesser than that of rice which has developed into an excellent "model cereal". Isolation of DNA is easy. It is a perennial and has many large flowers which make emasculation easy. It has got a very high seed number of about 6000 per plant. It is self-fertile and plants can be generated from cell culture. Transformation with *Agrobacterium tumefaciens* and *A. rhizogenes* is easy and efficient (Handberg et al 1994). Determinate root nodules appear within one week after inoculation with *Rhizobium loti* strains NZP 2213 and NZP 2037. Nitrogen fixation rates are good. Plants can be grown in closed tubes or plastic growth pouches. Improvised methods for high frequency transformation and regeneration are available to produce large numbers of transgenic lines (Stiller et al., 1997). This is especially significant since the molecular analysis of plant genes involved in nodulation has been slowed by the inability to produce high numbers of transgenic lines. All these attributes make *L. japonicus* a legume of choice for the molecular genetic analysis of determinate nodulation.

Objective - Positional cloning as a strategy to study nodulation

Almost all currently known nodulation genes are genes expressed in already committed cells, necessary for morphological development of symbiotic tissue and onset of nitrogen fixation. Genes that actually trigger or regulate nodule development could escape detection in differential screens or may be expressed in another part of the plant (Caetano-Anolles and Gresshoff, 1991). These could be genes known only by phenotype. Positional cloning is certainly the most widely used method to isolate genes known only by phenotype and could very well fit in as a strategy to isolate nodulation genes.

several molecular maps have been constructed. The genome size of 1,000 Mb is reasonable to be covered in HMW DNA libraries. It has been proven that large pieces of soybean DNA can be cloned into YACs and physical mapping using PFGE has been accomplished (Funke et al., 1993). BAC libraries of impressive insert sizes are available. Also, a rich collection of nodulation-related mutants have been generated in soybean. The universal nature of such mutations is supported by the existence of similar mutations in other legumes.

Since the *nts* gene is known only by its phenotype and genetics and nothing is known about its nature or precise site of expression it is a classical candidate for isolation by positional cloning. The first step in this direction would be to isolate a YAC or BAC clone that harbors the nearest mapped marker, pUTG-132a. As we have seen before pUTG-132a is less than 1 cM away from *nts*. Based on available genomic analysis this distance could be anywhere between 130-600 kb. This gives a reasonable chance of isolating a YAC or BAC clone that can be used as a starting point to build contigs. Once the clone harboring pUTG-132a is reached, the end clones can be isolated to start a chromosome walk. Given the average insert sizes of YAC/BAC libraries it is quite possible that the gene be reached with 2-3 clones of contig. All the resources are currently available to go ahead with this approach and hence this strategy seems to be the most logical to isolate the *nts* gene.

The model legume *Lotus japonicus* will indeed be a target of extensive molecular genetic study to unveil the molecular mechanisms underlying nodulation. It has been proven recently that *Lotus* is amenable to molecular genetic analysis (Jiang and Gresshoff, 1997). This work reports the construction of the first molecular linkage groups of this legume. The mapping population has been advanced to generate a set of immortal recombinant inbred lines useful for sharing in mapping experiments. Morphological loci for different characters such as stem shape, leaf color and flowering period have been found to be inherited as single Mendelian loci. In addition symbiotic mutants and nodule specific ESTs have been isolated (Szczyglowski et al., 1995). Already, over 20 mutations are available conferring non-nodulation, supernodulation and non-fixation phenotypes. This

suggests that the genetic map will get more saturated in the near future. HMW DNA libraries in YACs and BACs are not currently available. It would be worthwhile to construct YAC or BAC libraries of *Lotus*, as this seems to be the next logical step in positional cloning. Technologies and resources are available for the same and are the subject of this study.

CHAPTER II

MATERIALS AND METHODS

Preparation of High Molecular Weight (HMW) DNA

Two methods of HMW DNA preparation were followed. *Lotus* HMW DNA was extracted by protoplast preparation while soybean HMW DNA was extracted by nuclei isolation. The merits and demerits of the 2 methods are discussed later.

Lotus japonicus ecotype Gifu B129 and *Glycine max* v. Bragg plants were grown in horticultural potting mix in the greenhouse. Alternatively, plants were grown in a growth chamber (Perceival Scientific, USA) under the following conditions: 16 hours light/8 hours darkness, 21°C day temperature/15°C night temperature. Plants were watered once a week with a nutrient solution (1 liter of 100x stock containing 40 g NH_4NO_3 , 30 g KNO_3 , 30 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 10 g $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$, 2 g Fe-EDTA, 120 mg $\text{MnSO}_4 \cdot \text{H}_2\text{O}$, 12 mg H_3BO_3 , 40 mg $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 40 mg $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 8 mg $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$ and pH 6.8 - diluted to 1x with deionized water before watering).

Thirty to forty young trifoliates of around one month old *Lotus* plants were picked and surface sterilized in 70% ethanol for 2 minutes. Then they were transferred into 8 cm Petri dishes containing about 20 ml of protoplast buffer (36.4 g mannitol, 735 mg calcium chloride, 487 mg MES, 440 mg ascorbic acid, pH 5.6 in 500 ml water).

A protoplast digestion mixture of the following composition was prepared: 200 mg Driselase (Sigma O-0915); 100 mg Cellulase Onozuka R-10 (Serva 16419); 20 mg Pectolyase (Sigma P-3026) per 10 ml protoplast buffer, stirred for 30 minutes and adjusted to pH 5.6. This enzyme mix was briefly centrifuged and filter-sterilized to remove the particulate matter. The leaves washed in ethanol and protoplast buffer were cut with sterile razor blades into 1-2 mm strips in ice-cold protoplast buffer. The strips were incubated in 2-3 changes of cold protoplast buffer in 8 cm Petri dishes. The buffer was then replaced with the prepared digestion mixture (about 15 ml per dish), the Petri dish sealed with Parafilm® and placed on a rotating shaker at 30 rpm/room temperature for about 3-4 hours.

The appearance of protoplasts was observed with a Phase Contrast microscope (Olympus, USA). When the protoplasts looked perfectly round with very little cell debris and breakage, the incubation was stopped. The suspension was then filtered through a 43 μm Nylon filter into a 50 mL plastic centrifuge tube. The volume was brought up to 50 mL with protoplast buffer and the protoplasts were spun down at 800 rpm for 10 minutes in a Labofuge-B (American Scientific Products, USA). The protoplasts were washed in the protoplast buffer 1-2 times. The protoplasts were then counted using a haemocytometer and the concentration was brought to 10^8 per ml by centrifuging and resuspending in the desired volume of protoplast buffer. An equal volume of 1% Low Melting Point (LMP) agarose (GIBCO-BRL, USA) in protoplast buffer was added to the protoplast solution after the agarose had cooled down to 50°C. After resuspending uniformly, the mixture was then pipetted using a cut pipette tip, into prechilled plug molds. Once solidified, the plugs were extruded into 5 cm Petri dishes and incubated at 50°C in lysis buffer (0.5 M EDTA pH 9.3; 1% Sarkosyl, 1 mg/ml Proteinase K, 10 ml per dish) for 48 hours with one change of the lysis buffer at 24 hours. Then the lysis buffer was replaced with the storage buffer (0.5 M EDTA pH 9.3, 1% Sarkosyl) and the plugs were stored at 4°C till use. Each plug thus prepared (about 120 μl in volume) contained about 4 μg DNA.

For HMW DNA preparation from nuclei the procedure described by Zhang et al (1994) was followed with some modifications. About 3 week old soybean seedlings, with the first trifoliates beginning to unfold, were kept in darkness for 48 hours to lower the starch content. At the end of this period, about 20 g of partly unfolded light green primary trifoliates were harvested. This material was either used immediately for isolating nuclei or wrapped in aluminum foils, frozen in liquid nitrogen and stored at -80°C till use.

The leaves were ground in liquid nitrogen into a fine powder using a mortar and a pestle. The powder was transferred immediately into an ice cold 500 ml beaker containing 200 ml of isolation buffer (10 mM Tris-base, 80 mM KCl, 10 mM EDTA, 1mM spermine, 1mM spermidine, pH 9.4) and 400 μl of β -mercaptoethanol. The contents were stirred on ice for 10 minutes and then filtered into a cold 500 ml flask through two layers of cheese cloth and one layer of mira cloth. 5 ml of 20% triton X-100 (in isolation buffer) was added

and the flask gently swirled for 2 minutes. The solution was divided into 50 ml centrifuge tubes and centrifuged at 1,400 g (2,500 rpm) in a Beckman J2-21 centrifuge (Beckman, USA) in a swing bucket rotor at 4°C for 10 minutes. The supernatant was discarded and the pelleted nuclei in each tube were resuspended in 2 ml of wash buffer (isolation buffer with 0.5 % triton X-100 and 0.15 % β -mercaptoethanol). The nuclei solutions were pooled into one 50 ml centrifuge tube and the volume brought up to 45 ml with wash buffer. The tube was then centrifuged at 57 g (600 rpm) for 2 minutes to pellet and remove the cell debris. The supernatant was then filtered into a new tube through a 80 micron sieve and the nuclei pelleted by centrifuging at 2,500 rpm for 10 minutes. The washing was repeated two more times, after which the nuclear pellet was resuspended in 1 ml of isolation buffer. The total volume of the mixture was estimated using a cut blue pipette tip and an equal volume of 1% LMP Agarose was added. Plugs were prepared and treated the same way as the protoplast plugs.

Estimation of enzyme levels for optimum partial digestion

The amount of enzyme needed for optimum partial digestion differs from one HMW DNA preparation to the other. Hence it was important to estimate the optimum for each preparation. Different methods of digestion were followed for protoplasts and nuclei plugs.

Protoplast plugs were washed once with TE (10 mM tris-HCl and 1 mM EDTA, pH 8.0) at room temperature for 15 min., twice with TE plus 1 mM PMSF (phenylmethylsulphonyl fluoride) at 50°C for 30 minutes each and then once more against TE for 15 min at room temperature. Each plug was then cut into three smaller plugs of approximately 40 μ l each. The plugs were then equilibrated with either 100 μ l of 1x *Hind*III buffer (NEB #2, USA) for BAC cloning or 1x *Eco*RI buffer (NEB, USA) for YAC cloning. The equilibration was carried out in Eppendorf tubes on ice for one hour with one change of buffer at 30 minutes. At the end of this period the old buffer was removed and 50 μ l of fresh buffer was added to each tube. A different amount of enzyme was added to each tube - 0.05, 0.5, 5, 15 units etc. of *Eco*RI (NEB, USA) and 2, 10, 20, 50 units etc. of *Hind*III (NEB, USA). The tubes were allowed to sit on ice for 30 more minutes for the enzyme to

diffuse and then incubated at 37°C for 1 hour for the digestion to proceed. After 1 hour the reaction was stopped by adding one-tenth the volume of 0.5 M EDTA. The tubes were then put on ice.

Nuclei plugs were washed the same way as protoplast plugs. However, the plugs were incubated in the buffer for one hour at room temperature. Then the excess buffer was removed and the plugs melted at 65°C for 15 minutes. Once melted, the plugs were transferred quickly to 37°C and incubated for 5 minutes for cooling down. Then a different amount of enzyme was added to each tube as before and the digestion allowed to work for 30 minutes. The reaction was stopped as before.

Digested protoplast plugs were loaded into a 1% agarose (Amresco, USA) gel along with a 50-1,000 kb PULSE MARKER (Sigma, USA). In case of the nuclei, the melted digested plugs were loaded immediately using a cut pipette tip. The gels were then subjected to Pulsed Field Gel Electrophoresis (PFGE). Different electrophoresis conditions were used as detailed below:

150 V, 8-10 sec switch time, 14°C for 16 hours in a CHEF DR II system (BioRad, USA).

150 V, 20-70 sec switch time, 14°C for 20 hours in a CHEF DR II system.

6 V/cm, 50-90 sec switch time, 14°C for 22 hours in a CHEF-MAPPER system (BioRad, USA).

The gel was then stained with ethidium bromide (1 µg/mL) and viewed under UV light. An estimate of the optimum enzyme concentration was made based on the varying degrees of concentration of digested DNA, in the compression zones at the top of the gel (>200 kb) for DNA ran under the first PFGE conditions and in the 250-500 kb range of the other two PFGE conditions, as opposed to a control undigested sample (Figure 2). For each batch of HMW DNA preparation the enzyme concentration that gave the maximum amount of DNA in the above size ranges was selected as optimum.

Partial digestion and size selection by PFGE

Once the optimum partial digestion conditions are estimated the next step is to do a preparative partial digestion.

Two to three protoplast plugs were washed, cut into smaller plugs and then equilibrated with the respective 1x buffer as above. Then the plugs were incubated with the estimated optimum units of *EcoRI* (usually around 0.05-0.1 units) or *HindIII* (usually around 6-10 units). Digestion was carried out as above, and the plugs were loaded into a 1% LMP agarose gel with a 25 mm long middle well. The appropriate markers were also loaded. The gel was then subjected to PFGE under the compression zone conditions as described above. The sides of the gel including the marker lanes were cut off, stained with ethidium bromide and viewed under UV to locate the compression zone of HMW DNA. Notches were made at the sites of the compression zones. The notched pieces were then realigned with the rest of the gel and a thin slice of gel (not more than 5 mm in width) corresponding to the compression zone, was cut off. This slice contained the large DNA fragments needed for ligation.

The nuclei plugs were treated as before and incubated with the estimated optimum amount of enzyme. After partial digestion the digested, viscous solution was loaded immediately into the 25 mm long middle well of a 1% LMP gel. The gel was then subjected to PFGE under separating conditions as mentioned above. The marker lanes were stained as before and gel slices corresponding to the 250-450 kb range were cut off. At times, a second size selection was done by subjecting these cut off slices to PFGE under the following conditions: 140 V, 5 sec switch, 10 hours at 14°C in a CHEF -DRII or 4V/cm, 5 sec switch, 9 hours at 14°C in a CHEF-MAPPER. The compressed DNA was cut off as before.

The gel slices containing the HMW DNA were put into a 0.05 M EDTA (pH 8.0) solution and stored at 4°C till ready to use. For BAC ligation the following was done: The gel slice was equilibrated in TE for about 2 hours at 4°C. The slice was then put into a 1.5 ml Eppendorf tube and melted at 68°C for 15 min, transferred to 40°C. After 5 min β -

agarase enzyme (NEB, USA) was added at 1u per 100 µl agarose slice and the tube was incubated for about 3 hours. Agarase digestion was checked by incubating the tubes on ice for 10 minutes and seeing whether the contents stayed as a solution. This DNA solution was used directly for ligation. For YAC ligation the gel slices were used as such without digesting the agarose.

Isolation and preparation of the vectors

The BAC vector pBeloBAC11 (Fig 1-4) - This is a single copy plasmid and hence had to be prepared from a large culture volume. A single colony of *E. coli* DH10B with the pBeloBAC11 vector was inoculated into 10 ml LB medium containing 12.5 µg/ml chloramphenicol and was grown overnight at 30°C shaking at 250 rpm. This culture was used to inoculate 3 L of LB media supplemented with 12.5 µg/ml chloramphenicol. This culture was grown for about 20 hours at 30°C shaking at 250 rpm. The plasmid was isolated using Qiagen's Maxi plasmid isolation kit (Qiagen, USA) according to the manufacturer's specifications. The isolated vector was purified by Plasmid-Safe DNase enzyme (Epicentre Technologies, Wisconsin, USA) treatment and by cesium chloride-ethidium bromide equilibrium centrifugation (Sambrook et al 1989) at 48,000 rpm for 48 hours using a fixed angle rotor 70.1 (Beckman, Illinois, USA). The precipitated plasmid was resuspended in about 100 µl of TE. The plasmid was then digested to completion with *Hind*III in 1X TA Buffer (Epicentre, USA) for 4 hours. Complete double digestion was verified by running an aliquot in a 1% agarose gel. The digested plasmid was then dephosphorylated using HK phosphatase enzyme (Epicentre, USA) in the same solution at 30°C for 2 hours at the rate of 1 unit/µg of plasmid. Complete dephosphorylation was checked by setting up three control reactions using small aliquots of the prepared plasmid DNA:

- 1) Prepared DNA in 1X T4 Ligase Buffer (NEB, USA)
- 2) Prepared DNA + T4 DNA Ligase (NEB, USA) in 1X Ligase buffer
- 3) Prepared DNA + T4 Ligase + T4 Polynucleotide Kinase (NEB, USA) in 1X Ligase buffer.

The reactions were incubated at 16°C for 90 min. The reactions were then run on a 1% agarose gel to check for self-ligation, which is an indication of incomplete dephosphorylation. The phosphatase was heat-inactivated at 65°C for 20 min. The vector was stored at -20°C and used as such, without extraction. A working solution of about 20 µl was kept at 4°C.

The YAC vector pYAC4 (Fig. 1-3) - The vector was prepared from *E. coli* HB101 by a standard maxi-prep procedure using the alkaline lysis method (Sambrook et al 1989). The isolated vector was purified twice by continuous gradient equilibrium centrifugation in cesium chloride-ethidium bromide gradients using a Beckman Ti50 rotor (Beckman, Illinois, USA) at 45,000 rpm for 48 hours. About 50 µg vector DNA was subjected to double digestion, first with 30 units of *Bam*HI and then with 30 units of *Eco*RI in 1X *Eco*RI buffer. The digested vector was dephosphorylated using shrimp alkaline phosphatase (USB, Ohio, USA) in the same buffer for 2 hours at 30°C at a rate of 1 unit/µg of plasmid DNA. Once complete dephosphorylation was confirmed as before, the enzyme was heat inactivated at 70°C for 20 minutes. The DNA was purified by phenol-chloroform extraction, precipitated with ethanol, resuspended in 1X TE and stored at -20°C till ready to use for ligation.

Ligation of HMW DNA

BAC ligation - The ligation was carried out in a 100 µl volume in a molar ratio of 10:1 in vector excess. Typically the composition of the ligation mixture was as follows:

Insert HMW DNA	80.0 µl (About 150 ng of size selected agarased DNA)
Plasmid vector DNA	02.5 µl (About 40 ng of prepared pBeloBAC11 DNA)
10X Ligase Buffer	10.0 µl (NEB, USA)
T4 DNA Ligase	01.0 µl (400 units, NEB, USA)
Sterile Water	06.5 µl
Total volume	100.0 µl

The ligation reaction was carried out at 16°C for 14-16 hours. The ligation mixture was then dialyzed against 1x TE in an Ultrafree-MC filter tube (Millipore, Massachusetts,

USA) at 4°C for 24 hours with one change of TE at 12 hours, to remove the excess salts. The mixture was then heated to 65°C for 10 min to inactivate the ligase enzyme and stored at 4°C until transformation.

YAC ligation - The gel slice containing the large DNA fragments was equilibrated with the equilibration buffer (66 mM Tris-HCl pH 7.5, 5 mM MgCl₂, 30 mM NaCl, 0.75 mM spermidine, 0.30 mM spermine) at room temperature (about 22°C) for 30 min. The equilibrated gel slice was cut into 2-3 mm sections. About 200 mg of these slices were weighed out into an Eppendorf tube. Vector DNA (2 µg) was added to this tube (molar volume of 10:1 in vector excess). The agarose slices were melted at 68°C for 20 min. The tube was then cooled to 37°C for 10 min. and 1/10 volume of 10x T4 ligase buffer (prewarmed to 37°C) was added. After mixing the contents gently by rotating the tube and incubating at 37°C for a further 10 min., 1,600 units of T4 DNA ligase were added. Gentle mixing was repeated and the tube left at 37°C for 45 min for the enzyme to diffuse. The ligation was then left to proceed at room temperature overnight. The ligation mixture was melted at 68 °C for 20 min. About 100 µl of the melted mixture was put in a new tube (with a cut pipette tip), transferred to 40°C, and after 10 min 2 µl of β-agarase enzyme was added. The digestion of agarose was allowed to proceed for 3-4 hours. The completely aqueous solution was then dialyzed as before against 1x TE in an Ultrafree-MC filter tube. With each HMW ligation, a control ligation was also set up with total digested DNA of Lotus and soybean to compare ligation and transformation efficiencies. All conditions were the same as the HMW DNA ligation. Chromosomal DNA of *L. japonicus* ecotype "Gifu" and *Glycine max* v. Bragg, from which the insert DNA was prepared for this control, were isolated according to Dellaporta et al (1983).

Transformation

BAC transformation: Transformation of electro-competent *E. coli* DH10B strain (ElectroMax DH10B, Gibco-BRL, USA) was carried out by electroporation using either an *E. coli* PULSER (Bio-Rad, USA) or a Cell Porator-Voltage Booster system (Gibco-BRL, USA). 1.5 µL of the ligation mix was added to 19 µL of competent cells and then electroporated under the following settings:

Voltage 2.5 kV, Capacitance 25 μ F, Resistance 100 ohms in the E. coli PULSER or Voltage 400 V, Capacitance 330 μ F, Voltage Boosted Resistance 4000 ohms, low impedance in the Cell Porator-Voltage Booster system.

After electroporation the cells were transferred to 1 ml SOC solution (2% Bacto tryptone, 0.5% Bacto yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose pH 7.0) and shaken at 37°C, 200 rpm for 45 minutes to allow expression of the Chloramphenicol resistance gene. The cells were then spread on LB plates containing chloramphenicol (12.5 μ g/ml), X-gal (40 μ g/ml), and IPTG (0.072 μ g/ml). The plates were incubated at 37°C for 24 hrs to allow the blue/white color selection to work.

YAC transformation: Transformation of the host *Saccharomyces cerevisiae* strain AB1380 was carried out by spheroplasting (Burgers and Percival, 1987) with some modifications. About 200 ml of culture grown in YPD media was harvested when the cell density reached 3×10^7 cells/ml. The culture was divided into 4 tubes of 50 ml each. They were then spun down and washed first with distilled water and then with 1 M sorbitol. The cells were then resuspended in 20 ml of spheroplasting buffer (1 M sorbitol, 10 mM sodium phosphate, 10 mM EDTA pH 8, 30 mM β -mercaptoethanol). Thirty, 50, 75 and 100 microliters of 500 mg/ml lyticase (ICN, USA) were added to the different tubes. Spheroplasting was carried out at 30°C for 10 min. The extent of spheroplast formation was judged by spectrophotometric reading at 800 nm. For this 200 μ l each of cells were added to 800 μ l of water and 800 μ l of sorbitol. A ratio of the water and sorbitol readings was taken as an indication of the extent (%) of spheroplasting. The batch of cells showing 60-70% spheroplasting was selected for transformation. Cells were spun down at 900 rpm for 5 minutes in a Labofuge-B (American Scientific Products, USA). The pellet was washed in 20 ml SSP (1 M sorbitol, 10 mM sodium phosphate). The pellet was then resuspended in 20 ml YPD buffer (Ausubel et al., 1995) in 1 M sorbitol. The tube was then gently shaken on sides at 30°C for 30 min. Ten ml of STC (1 M Sorbitol, 10 mM Tris-HCl pH 7.5, 10 mM $CaCl_2$) was added to this solution and the cells were then harvested at 900 rpm for 5 min. The cells were then resuspended in 2 ml STC. These cells were used for

transformation. Ligation mix (50 µl) was added to 300 µl of the competent cells for each set of transformation (one 50 ml culture tube). The tubes were left at room temperature for 10 min. Then 3 ml PEG solution (20% PEG 8000-Sigma, 10 mM Tris-HCl pH 7.5, 10 mM CaCl_2) was added to this tube with just enough force to resuspend the cells. The tubes were left at room temperature for a further 10 minutes. Spheroplasts were then pelleted at 900 rpm for 4 min. The pellet was resuspended in 450 µl of SOS (1 M sorbitol, 6.5 mM CaCl_2 , 0.25% yeast extract, 0.5% bactopectone, uracil 20 mg/ml, tryptophan 20 mg/ml, pH 5.8) and incubated at 30° C for 20 min. Then 10 ml top agar (0.8% Difco in YPD; 45-48°C) was added to each tube. Cells were mixed with the top agar by inverting the tube 2-3 times. The mixture was then plated on single selection plates (182 g sorbitol, 20 g agar, 20 g dextrose, 0.8 g -URA dropout powder BIO 101 Inc., 1.7 g yeast nitrogen base (without amino acids and ammonium sulfate), 5 g ammonium sulfate in 1,000 ml total volume) and left at 30°C for 2-3 days.

Insert size characterization

BACs : White colonies containing putative inserts were picked on to a new LB plate for a second color screen. The clones were then transferred to 96 well microtiter plates containing 60 µl of freezing buffer (36 mM K_2HPO_4 , 13.2 mM KH_2PO_4 , 1.7 mM Sodium Citrate, 0.4 mM MgSO_4 , 6.8 mM $(\text{NH}_4)_2\text{SO}_4$, 4.4% v/v glycerol, 12.5 µg/ml chloramphenicol in LB). The plates were then stored at -80°C. About 10 clones from each transformation were checked for insert size. Plasmid DNA was isolated by the standard alkaline lysis plasmid miniprep method (Sambrook et al, 1989). The isolated plasmid DNA was resuspended in 20 µl of TE. 10 µl of this was digested with *NotI* restriction enzyme (*NotI* has restriction sites on either side of and bordering the cloning site). Being a rare cutting enzyme few cutting sites were expected within the insert DNA. The composition of the restriction reaction was as follows:

Plasmid DNA	10.0 µl (around 1 µg)
10X Buffer	2.0 µl (NEB #3)
<i>NotI</i> enzyme	1.0 µl (10 units, NEB)
10X BSA	2.0 µl (Bovine serum albumin, diluted from 100X stock, NEB)

Sterile water	5.0 μ l
Total	20.0 μ l

The reaction was allowed to proceed to completion at 37°C for about 3 hours. The restricted plasmid DNA was then subjected to PFGE in a 1% agarose gel at 150 V with a 10 second switch time for 16 hours at 14°C in a CHEF-DRII or at 5 V/cm, 5-15 seconds pulse, 120° angle, 14°C for 13 hours in a CHEF-MAPPER. Appropriate markers (PULSE MARKER, Sigma, USA) were loaded to judge the size of the inserts. The gel was stained and viewed as before (Fig.4a).

The gel was then blotted on to a Nylon membrane and probed with ³²P labeled total *L. japonicus* "Gifu" genomic DNA for confirmation. Hybridization was carried out exactly as before (Fig. 4b).

For confirmation, Southern blotting and hybridization was resorted to. The gel was blotted on to a Zeta-Probe® GT Nylon membrane (Bio Rad, CA, USA) by using the Vacugene (Pharmacia-LKB, Sweden) vacuum blotting system following the manufacturer's instructions. To depurinate the DNA, a 0.25 M solution of HCl was added to the surface of the gel, while maintaining the vacuum at 20 cm water pressure. The vacuum was increased to 30 cm water pressure and for 10 minutes. After this period the vacuum was brought down to 20 cm and the excess HCl removed. Keeping the vacuum at the same level, 0.4 M NaOH was added to the surface of the gel to denature the DNA and effect the transfer. The vacuum was increased to 30 cm for 10 minutes. The vacuum was again brought down to 20 cm and the NaOH discarded by pipetting. Fresh NaOH was added so as to immerse the gel, vacuum increased to 40 cm and the transfer allowed to continue for about 1 hour. NaOH was removed as above, the gel discarded and the membrane with the imprinted DNA taken out. The membrane was washed in 2X SSC (17.53 g NaCl + 8.82 g of sodium citrate in a one liter solution, pH 7.0) to bring the pH down to 7.0. The membrane was then dried and auto-crosslinked in a UV Crosslinker (Stratagene, USA) at 1200 μ J/cm². The membrane thus prepared was used for Southern hybridization.

Membrane was probed with *Lotus japonicus* ecotype Gifu and *Glycine max* cv Bragg genomic DNA, extracted as mentioned before. The probe was labeled with ^{32}P (DuPont, USA) by random priming (Feinberg and Vogelstein, 1982) using the Megaprime Labeling Kit (Amersham, UK) according to the manufacturer's protocol. Hybridizations were carried out according to the standard protocol (Sambrook et al, 1989) with differences in the composition of the solutions. The prehybridization and hybridization solutions used were the same in composition (1mM EDTA, 0.5 M NaHPO_4 pH 7.2, 7% SDS). The hybridizations were carried out at 60°C in The Belly Dancer® hybridization water bath (Stovall Life Sciences Inc., NC, USA) or in an Incubator-Shaker (Boekel, USA) for about 20 hours. The membranes were washed twice in Wash I (1 mM EDTA, 40 mM NaHPO_4 pH 7.2 and 5% SDS) and twice in Wash II (1 mM EDTA, 40 mM NaHPO_4 pH 7.2 and 1% SDS). Each wash was done at 60°C for 20-30 minutes. The membranes were then dried on blotting paper and autoradiographed using Cronex® medical X-Ray films (DuPont, USA), typically for 3-5 days depending on the strength of the probe. The films were developed in a Kodak GBX developer (Eastman Kodak, USA) for about 5 minutes, briefly rinsed in water and treated with GBX Fixer (Sigma, USA) for another 5 minutes. The film was then washed with water and dried in the air.

YACs: the yeast colonies that appeared on the single selection plates were picked on to double selection plates (same as single selection plates except in that it contained -URA-TRP dropout powder instead of -URA powder). After incubation at 30°C for another 3 days about 20 of those colonies that turned red, were picked up for an insert size check. Each of these colonies was inoculated into 5 ml of 1.5x AHC (2x AHC- 3.4 g YNB, 10 g ammonium sulfate, 20 g casein-hydrolysate acid, 40 mg adenine hemisulfate, 100 µl concentrated HCl, 40 g dextrose) containing 50 mg/ml ampicillin. The cultures were shaken at 30°C/250 rpm till the cultures turned fully red (about 48 hours). Cells were spun and then washed with 5 ml SE (0.9 M sorbitol, 50 mM EDTA). The washed cells were resuspended in 150 µl LMP agarose containing 0.25 mg/ml lyticase and 30 mM β -mercaptoethanol. Each of the cultures was then put into pre-chilled plug molds. Once solidified the plugs were extruded into separate Eppendorf tubes containing 1.5 ml

spheroplasting solution (0.9M sorbitol, 50 mM EDTA, 30 mM β -mercaptoethanol and 0.25 mg/ml lyticase). The tubes were incubated at 37°C overnight. The spheroplasting solution was then replaced with ESP (0.5 M EDTA pH 8.0, 1% Sarkosyl, 1 mg/ml proteinase K). Incubation was done overnight at 50°C. The old solution was replaced with fresh ESP and incubation was continued overnight at 50°C. Small pieces of the treated plugs thus obtained were subjected to PFGE in a 1% agarose gel at 200 V/14° C with a 20-70 seconds switch time for 16 hours in a CHEF-DRIL. The gel was stained and viewed under UV to check the insert sizes of the artificial chromosomes (Fig. 3a).

The gel was then blotted on to Nylon membranes and probed with ^{32} P labeled pBR322 (which has high sequence homology to the YAC vector) as well as *Lotus japonicus* ecotype Gifu genomic DNA. The blotting, hybridization and autoradiography were carried out exactly as mentioned above.

Screening of BAC pools for candidate clones

Two BAC pools (#14 & #40, containing about 350 clones each) from the soybean BAC library (Noble Foundation, Oklahoma, USA) were found to be positive for pUTG-132a in an initial screening by scientists at the Noble foundation (personal communication). These two pools were screened further to locate pUTG132a candidate BAC clones. A PCR based screening method was followed.

The pools were diluted 1:10⁴ with LB media and plated out onto LB plates that were usually used to plate BACs. 768 clones from each pool were picked onto 96 well microtiter plates in the same way BAC clones were picked before. Eight 96 well plates were used for each pool. The plates were put at 37°C for one day to allow the clones to grow.

Corresponding to the rows, columns and numbers of the plates, 28 sub-pools were made from the eight plates for each pool. These consisted of 8 "X" co-ordinate pools (96 clones each, the plate rows), 12 "Y" co-ordinate pools (64 clones each, the plate columns) and 8 "Z" co-ordinate pools (96 clones, the plates themselves). The pools were made by inoculating 5 ml of LB, containing 15 μ g/ml chloramphenicol, with the clones using a

toothpick. The pools were grown at 37°C overnight. 1.5 ml of this culture was used for DNA isolation, while the rest were stored as 30% glycerol stocks in the -80 °C freezer.

Plasmid DNA from the pools were isolated by the standard alkaline lysis method. The DNA was extracted by treating with a 25:24:1 phenol-chloroform-isoamyl alcohol mixture (Ambion, USA) and precipitated with ethanol. After washing once with 70% ethanol, the DNA pellet was resuspended in 20 µl of TE and the concentration measured.

The pUTG-132a sequence was retrieved from the GENBANK database. Sequence comparisons were done with the published primer sequences (Kolchinsky et al, 1991) using the OLIGO-4.0 program. At first *Glycine max* v. Bragg genomic DNA, *Glycine soja* genomic DNA and the cloned pUTG-132a were amplified using different primers to give PCR products of the desired size. Different PCR conditions were tried out. The following left (LI & LII) and right (RI & RIII) primers were used:

LI 5' CAC GGT CAC TCA TGG GCC ATG G 3' (22 mer)

LII 5' GCA GCA GTG TTG GGC ATG TCT CT 3' (23 mer)

RI 5' CTG CAG AAT TGG ATT CCC AAA AGC 3' (24 mer)

RIII 5' CAC TCA CTC GTT CTA AAA GG 3' (20 mer)

The primers were ordered from Integrated DNA Technologies, Inc. (Iowa, USA).

Reaction mixtures were prepared as follows:

30 µl total reaction volume, 1X PCR buffer (diluted from 10X stock - Perkin Elmer, USA), 2.0 mM MgCl₂, 200 µM dNTP, 0.1-0.25 µM each primer, 50-75 ng DNA, 0.5 units Amplitaq Polymerase enzyme (Perkin-Elmer, USA).

The following PCR conditions were used (slightly modified from Cai et al, 1995):

Samples were denatured for 5 min at 94°C and then let through 35 cycles as follows: 94°C denaturation for 30 sec, 58°C annealing for 1 min, 72°C extension for 1 min. The samples were then let to stay for an additional cycle at 72°C for 10 min.

About 10 µL of each reaction was then loaded on to a 1% agarose gel and electrophoresed. The gels were stained and viewed as before to see the PCR products.

For amplifying the pooled BACs, different PCR conditions had to be followed even though the composition of the mixture was essentially the same. Genomic and the cloned pUTG-132a DNA were used as controls. A touchdown PCR run was done as follows: denaturation at 94°C for 30 sec, annealing touchdown from 62°C to 55°C at 1°C per cycle (8 cycles) for 1 min and extension at 72°C for 2 min. Once the 55°C annealing temperature was reached, the run was repeated for 35 cycles. At the end of the run the samples were left at 72°C for 5 min.

The amplified samples were checked as above, by electrophoresis to view the PCR products. The gels were then blotted onto Nylon membranes and hybridized with pUTG-132a fragment. For hybridization, the probe fragment was isolated from a 5 ml overnight culture by the standard alkaline lysis method. The isolated DNA was purified by the Qiagen plasmid purification kit (Qiagen, Germany). About 5 µg of this plasmid DNA was digested to completion with 50 units of *Pst*I in a total digestion volume of 100 µL. Digestion was checked by electrophoresing a small aliquot. Once digestion was verified a preparative electrophoresis was run in a 1% LMP agarose gel. The sample was loaded into a long middle well. The gel was stained to view the 1.7 kb pUTG-132a fragment. A slice of the gel containing the fragment was cut out from the gel. The DNA in this slice was isolated by using the β -agarase enzyme (NEB, USA) following the manufacturer's specifications. The isolated DNA was resuspended in 50 µL of TE buffer. Probe labeling, hybridization, washing and autoradiography were done as before.

Confirmation of the candidate BAC clones

Based on amplification and subsequent hybridization, candidate BAC clones were picked up from pools 14 and 40 by aligning the "X", "Y" and "Z" pools. The candidates thus picked were first characterized for size as before by *Not*I digestion followed by PFGE. The gel was then blotted and hybridized with the pUTG-132a fragment.

For further confirmation, the isolated DNA from the candidate BACs were digested with *Pst*I enzyme (in a total volume of 20 µL) to release the putative 1.7 kb fragment. The samples were electrophoresed in a 1% agarose gel to separate all the restricted fragments.

The gel was stained to view the restriction digestion pattern. The gel was then blotted and hybridized with pUTG-132a. To avoid plasmid DNA contamination of the probe, pUTG-132a was isolated by PCR amplification (Kolchinsky et al, 1991) and gel purified as before and then labeled. The labeled probe was hybridized to the same membrane after stripping off the previous probe (Sambrook et al, 1989). The same procedure was repeated for a *DraI* digest of the candidate BAC DNA.

The candidate BAC DNA was also PCR amplified to check for the putative amplification product. A "nested" PCR approach was followed. At first, the BAC DNA was amplified (Kolchinsky et al, 1996) using primers LII and RI to amplify the 1.7 kb fragment. Then 1 µl of this 30 µl PCR mixture, already enriched for the 1.7 kb fragment, was used as the template for a second round of amplification. For the second round, primers LII and RIII were used to amplify the 1.2 kb fragment. The products were then electrophoresed, blotted, hybridized and autoradiographed.

Isolation of BAC end clones

BAC end clones were isolated according to published protocols (Cai et al, 1995).

The BAC library was constructed in the *Bam*HI cloning site of the vector. Hence, both the left and right end clones could be isolated by plasmid end rescue. This involved restriction digestion of the entire BAC clone, self-ligation and transformation to select for the chloramphenicol resistance gene that is located close to the right end of the BAC. Separate aliquots of the isolated BAC DNA were digested completely with *Eco*RI, *Kpn*I and *Sac*I for the right end and with *Hind*III for the left end clone, each in a total reaction volume of 20 µL. The digested DNA was purified once by phenol-chloroform extraction, precipitated with ethanol and resuspended in 20 µL of TE. These DNA solutions were self-ligated in a 40 µL total volume at 16°C overnight, as specified. The samples were then dialyzed overnight against 1x TE. 1.5 µL of this ligation mixture was used to transform 19 µL of ElectroMAX DH10B competent cells according to the protocol described before. The transformed cells were plated out on chloramphenicol selection plates, as before. DNA was extracted from three positive colonies on each plate, digested with the same enzyme that

was used for subcloning and electrophoresed on a 1% agarose gel. The idea was to see the sizes of the end clones and select those which are in the 10-25 kb range.

The left end could also be rescued by inverse PCR. Even though this was not needed the procedure was tried out just to know whether it is possible to obtain better end clones. Separate aliquots of BAC DNA were digested with the same enzymes used for isolating the right end. The digested DNA was self ligated as before. Two μL of this ligation reaction were used in an inverse PCR reaction as specified. Two Primer pairs - IPCR1 & IPCR2 and BAC1 & BAC2 - were used in the inverse PCR reaction following the specifications. PCR products were electrophoresed to see the sizes of the PCR products.

Sequencing of BAC end clones

The isolated end clones were sequenced using the ABI automated sequencer. Pure plasmid DNA for sequencing was isolated using Qiagen's Plasmid mini-kit following the manufacturer's recommendation. The plasmid was resuspended to a final concentration of 200 ng/ μL and sent to the University's Molecular Biology Resource Facility, for sequencing. About 2.5 μL of this DNA (500 ng) was used per reaction using the M13 reverse primer. Once the sequence was available, primers were designed (ordered from IDT Technologies, IA) to facilitate further screening for construction of contigs. Only right end clones (REs) were taken up for analysis and primers were designed only for BAC5 (5RE).

PCR amplification using end clone primers

Genomic and plasmid DNA were PCR amplified using primers designed for 5RE. Genomic DNA from *G.max* (cultivars Bragg and Williams), *G.soja* and *Lotus japonicus* (ecotypes Gifu and Funakura). Plasmid DNA from BAC3 and BAC5 (positive control) as well as the end clones from both the BACs were also amplified. The PCR protocols followed were the same that was followed while screening the library. In a later experiment, a lower annealing temperature of 55° C was used to amplify a pUTG-132a fragment from the above samples. All other conditions were the same. The PCR products

were electrophoresed to determine the sizes of the amplified products. Hybridization with the end clone sequence was done to confirm the products. The procedure followed was exactly the same as mentioned before for southern hybridization.

CHAPTER III

RESULTS

A *Lotus* BAC library of about 10,000 clones with an average insert size of around 65 kb was produced. The sizes of the clones range from 30 kb to 160 kb. The average insert size is small when compared to other BAC libraries (discussed in the introduction), but the library will be of use since it is the only HMW DNA library available in *Lotus* and also since about 15-20% of the clones are 100 kb or larger in size. Soybean ligations of up to 75 kb in average insert size were produced but were not carried through to complete the library because three libraries of average insert sizes larger than 100 kb are already available (see introduction). A YAC transformation was done with *Lotus* genomic DNA. The clones averaged about 150 kb in size, but were very few in number (about 500). The decision was made not to proceed with YACs because of the technical difficulties and problems associated with YACs.

The utilization of BAC libraries was examined by screening two BAC pools for clones containing pUTG-132a. The screening was carried out by PCR. This work was done in collaboration with Dr. Artem Men (Knoxville, TN) and Dr. Madan Bhattacharya (Ardmore, OK). Three candidate clones of sizes 50, 165 and 70 kb respectively were isolated. End clones for further screening of the library were isolated. The clones are being characterized further by Dr. Men.

Construction of HMW DNA clones

The observations at the different stages of HMW DNA cloning are detailed below.

Isolation of HMW DNA

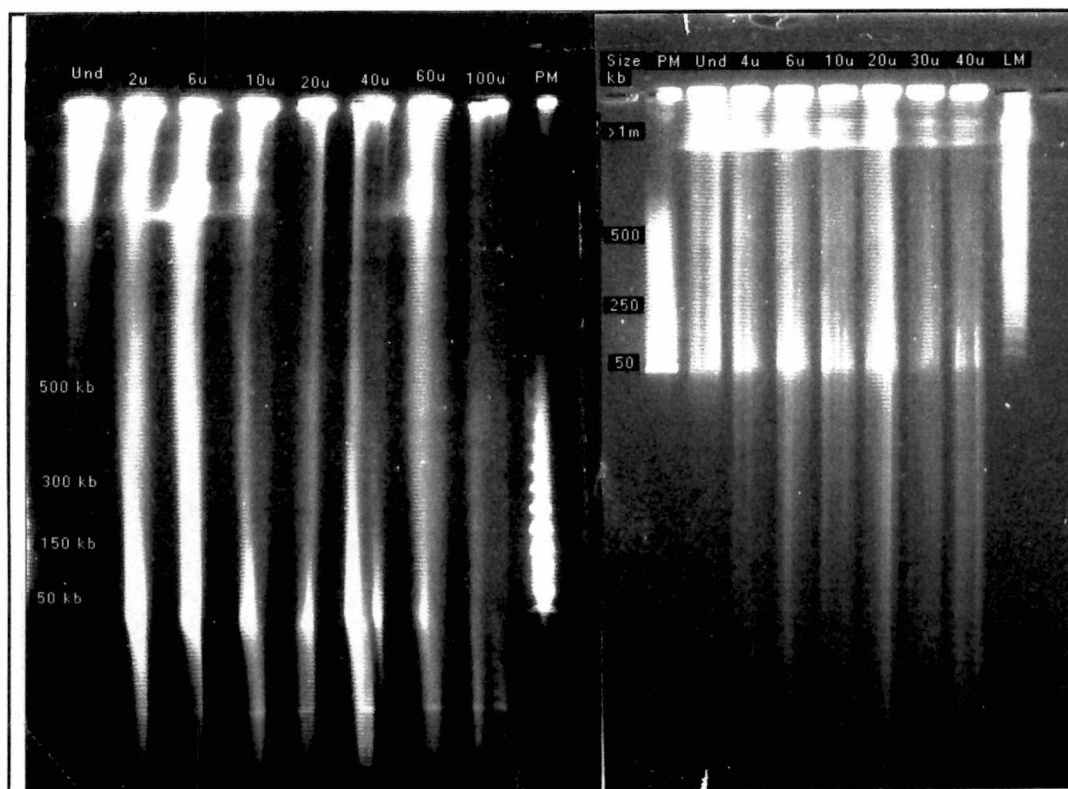
It was possible to isolate relatively undegraded, good quality HMW DNA from soybean and *Lotus*. HMW DNA was isolated from protoplasts as well as nuclei. The protoplast method (Re: previous chap.) was followed for *Lotus*, because of the small size of the leaves. This makes it easier to process the leaves in small size Petri-dishes. The nuclei method (Re: thesis pp 38) was followed for soybean because the size of the leaves

permits easier grinding in liquid nitrogen. The quality of the HMW DNA varied from preparation to preparation. HMW DNA isolated from nuclei was always higher in concentration. This may be because of the higher quantity of material used. However, protoplast plugs were found to be more amenable to restriction digestion. Nuclei plugs contained more contaminating material like starch and that may have reduced the effectiveness of restriction enzymes. Protoplasts were handled much more carefully than nuclei and hence were devoid of most of the contaminants that were found with nuclei. It was also noticed that plants grown in growth chambers gave better quality HMW DNA than plants grown in greenhouses.

Partial digestion and size selection by PFGE

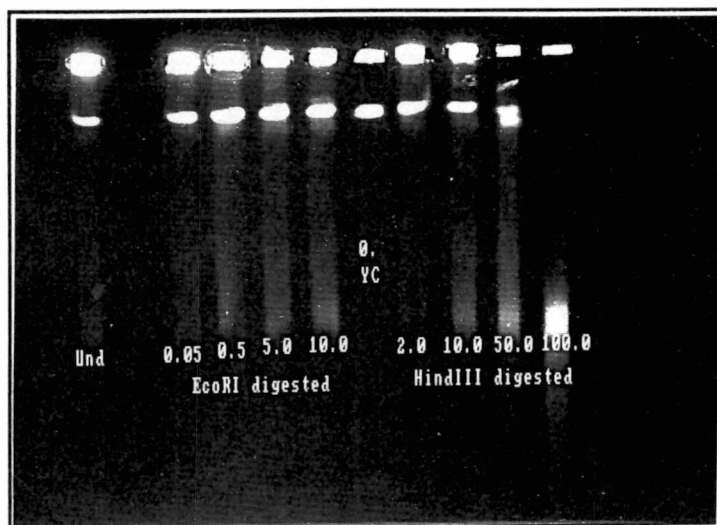
The quality of the isolated HMW DNA was judged by digesting with different amounts of restriction enzyme and running the DNA under PFGE (Fig. 3-1a & 3-1b).

The quality of the HMW DNA varied between preparations. In case of the best preparations, a very small amount of enzyme was enough to give the maximum amount of DNA in the desirable size range (250-500 kb). In such cases it was possible to see a clear difference between the undigested control sample and the digested sample. Most of the DNA in the undigested sample stayed up in the wells or in the compression zone, while there was progressive smearing of DNA with higher concentrations of the enzyme. In case of low quality preparations, it was found that there was a lot of degraded DNA of lower size. This made it difficult to distinguish clearly between the undigested DNA sample and the sample digested with minimal amount of enzyme. Many times, this led to using more than the optimum amount of enzyme for partial digestion resulting in small insert sizes. At times, the plugs could not be digested at all even after incubation with the enzyme for long hours. This may be due to the presence of inhibitory substances that could not be removed by washing with detergents or the inability of the enzymes to get access into the plugs. In such cases, the plugs were melted and made into a viscous solution to allow access to the enzyme. While this process increased the effectiveness of the restriction enzymes, it also resulted in a lot of broken and degraded DNA.



(A) (B)
a: PFGE of partially digested soybean and *Lotus* HMW DNA. Soybean nuclear DNA (A) and *Lotus japonicus* protoplast DNA (B) were digested using different units of *Hind*III, as shown on top of the picture. (A) was ran at 200V, 20-70 sec switch for 16 hours while (B) was ran at 200V, 60-90 sec switch for 16 hours. The size of the DNA at different zones of the gel are given in the picture and so are the amounts of enzyme used.

Figure 3.1: PFGE of partially digested DNA.



b: PFGE of *EcoRI* and *HindIII* partially digested *Lotus japonicus* DNA. These enzymes were used for cloning HMW DNA into YAC and BAC vectors respectively. The PFGE was ran under compression zone conditions: at 150v, 8 sec switch for 16 hours at 14°C. The amount of enzyme needed for optimum partial digestion was judged from the quantity of DNA in the compression zone seen just below the wells. The enzyme units are given on top of the picture. YC denotes yeast chromosomes that were ran as size markers.

Figure 3-1 (continued)

Understandably, the amount of enzyme needed for optimum partial digestion varied between preparations, with the best preparations giving optimum digestion with very small quantities of enzyme. As a standard, about 6-8 units of *HindIII* gave the maximum amount of DNA in the 250-500 kb range. In case of *EcoRI*, 0.02-0.05 units gave the optimum partial digestion. In most of the cases about 50 units of *HindIII* or 10 units of *EcoRI* gave complete digestion of the DNA in the plugs.

As stated before, different PFGE conditions were experimented with for getting large sized clonable DNA. At times, even double size selection was attempted. When the concentration of DNA in the plugs was higher, like in the case of nuclei plugs, better results were obtained when PFGE condition facilitating a wider spread of DNA was performed. This might be because a higher DNA concentration (>5 µg) may lead to entrapment of small molecules in the compression zone. When run under separating conditions (Thesis pp 45), there is a greater chance for the small molecules (<50 kb) to separate out from the large size range (>100 kb). In the case of protoplast plugs, where DNA concentration was low (< 5 µg), compression zone conditions (Thesis pp 45) gave good results. Double size selection, wherein the DNA was first ran under separating conditions, size selected, and then ran under compression zone conditions, resulted in larger insert sizes. However, this could be attempted only when the DNA concentration was sufficiently high and hence done only a couple of times.

Isolation and preparation of the vectors

Isolating sufficient quantities of pure BAC vector was a very difficult process, mainly because the plasmid is a single copy one. Usually, it required growing at least 3 L of bacterial culture. As a standard, one should be able to get around 25-30 µg of pure plasmid DNA from this culture volume. But many times, the yield was under 10 µg! This might be due to the excessive loss of DNA in the various intermediate processes. A single round of purification using the Qiagen's Maxi plasmid isolation kit did not give pure plasmid DNA. A lot of chromosomal DNA contamination was found. Even treating with 50 units of Plasmid-Safe DNase enzyme did not eliminate the chromosomal DNA completely (Fig. 3-2). Only a cesium chloride-ethidium bromide (CsCl-EtBr) equilibrium

centrifugation gave sufficiently pure plasmid DNA. But the yields were found to be too low to do all the preliminary experiments and the final cloning. In one of the trials, direct CsCl-EtBr was attempted without the intermediate steps. The yields were found to be good and the purity was found to be sufficient. This solved the plasmid isolation problem. Isolating the YAC plasmid was easy since it is a multi-copy plasmid and could be isolated by following standard procedures (Thesis pp 48).

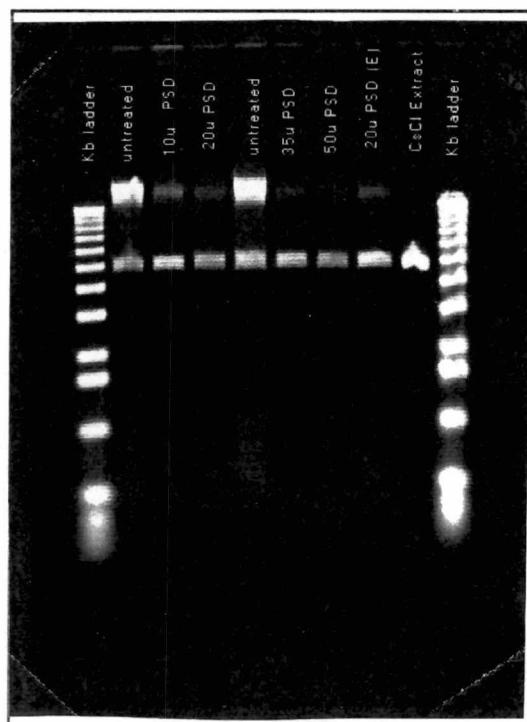


Figure. 3-2: Plasmid-Safe DNase enzyme treatment of the BAC vector. This was done to eliminate chromosomal DNA contamination of the BAC vector DNA. As seen in the picture, even after treating with 50 units of enzyme chromosomal DNA still persisted. On the other hand, CsCl-EtBr extracted vector shows no contamination.

Preparation of the vector (digestion and dephosphorylation) was done as per protocol (Thesis pp 48). It was found that precipitating the plasmid DNA after digestion gave better dephosphorylation in case of the BAC vector, where HK phosphatase was used as the dephosphorylating enzyme. This might be because the presence of EDTA in the solution chelates CaCl_2 , which is essential for the activity of the dephosphorylating enzyme HK phosphatase. However, this resulted in significant loss of valuable plasmid DNA. When checking for dephosphorylation, a single band indicated a perfectly digested dephosphorylated plasmid, while more than one band or a smear indicated incomplete digestion or dephosphorylation. Incubation with kinase and ligase served as a check for an

intact cloning site. If the cloning site is intact, the plasmid should circularize by the activity of ligase. Only after confirming all these facts was the plasmid carried over for further experiments (Fig.3-3).

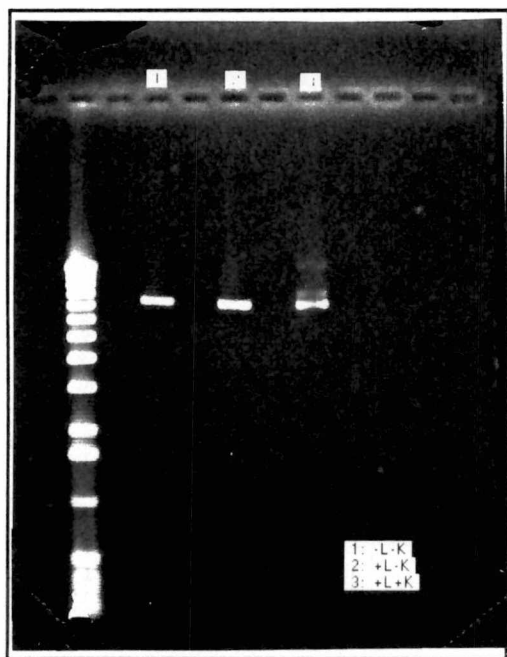


Figure 3-3: Preparation of the 7.4 kb BAC vector. Lane 1 has untreated digested dephosphorylated plasmid. Lane 2 has digested dephosphorylated plasmid treated with ligase. Lane 3 has digested dephosphorylated plasmid treated with ligase and kinase. A single band in lanes 1 and 2 confirm a digested dephosphorylated plasmid. Multiple bands in lane 3, formed by recircularization of the plasmid confirms an intact cloning site.

Preparing and verifying the YAC vector was easier. Shrimp Alkaline Phosphatase, the enzyme used for dephosphorylating the YAC vector was found to be perfectly compatible with the *EcoRI* buffer. Since sufficient amounts of plasmid was available, the loss of DNA from extraction and purification was not significant. Since the plasmid gives three distinct bands after double digestion, verifying digestion and dephosphorylation was easy. The kinase test gave a single band, the size of which is the sum of the sizes of the three separate bands that come after digestion (Fig. 3-4). At times, increased concentration of BSA (about one fold more than normal) was found to be essential for complete dephosphorylation.

Ligation and transformation

In the case of BAC ligation, problems were encountered when at times the agarase digestion was found to be incomplete. This could be because of the presence of

contaminating materials, higher salt concentration or different pH. The manufacturer's (NEB, USA) protocol was never found to work exactly. Either longer incubation times or higher enzyme concentrations were found to be essential. A completely agarased DNA solution was found to be absolutely critical for efficient ligation. Even though a 10:1 vector:insert concentration was used as standard, often higher vector concentrations were found to work better. When dialyzing the ligation mixture to reduce salt concentration, it was found that changing the buffer 2-3 times in 24 hr. greatly reduced salt concentration.

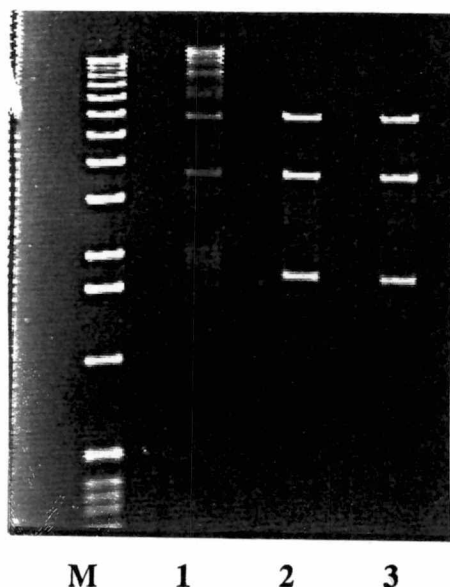


Figure 3-4: Preparation of the YAC vector. M- Kb ladder (DNA size marker). 1- Digested dephosphorylated plasmid treated with ligase and kinase. 2- Digested dephosphorylated plasmid treated with ligase alone. 3- Untreated digested dephosphorylated plasmid. The three distinct bands in lane 2 confirms that the plasmid is fully digested and dephosphorylated.

YAC ligation was found to proceed efficiently in agarose. However, it was important to melt the gel slice very well (for 20 minutes) before adding the vector and the ligase enzyme, and keep the contents molten for sometime for the enzyme to diffuse and make proper "contact". To ensure proper ligation in agarose a large excess of vector was added. In this case, the agarase digestion could be allowed to proceed for longer times (up to 30 minutes) since the HMW DNA molecules present are already ligated to the vector.

Published protocols were followed for BAC transformation by electroporation (Thesis pp 49). Control transformations were done using the same reactions that were done to confirm the vector preparation. Salt concentration of the ligation mixture was very critical. Arcing of the current was observed when the salt concentration was too high. This

led to poorer transformation efficiencies. With increased capacitance and resistance the total applied voltage could be reduced. This was found to be critical in transforming *E.coli* with large insert molecules. Self-made competent cells were found to be of much lower transformation efficiency than commercially available ones (Gibco-BRL, USA). This was revealed when control transformations using different kinds of DNA were tried (Table 3-1).

Despite repeated attempts, it was not possible to produce cells with the same level of electro-transformation efficiency as commercial ones. One strange observation was that the same self-made bacterial cells that transformed well with BACs isolated from other bacterial cells did not transform well with recombinant plasmids made in the test tube. It is possible to speculate that these cells transform well with supercoiled plasmids, however large they are, but not with circular ligated molecules. Because of all these problems, it was later decided to use only commercial cells.

Table (3-1)
Control transformations done with *E. coli* DH10B electrocompetent cells

Transforming plasmid DNA	Transformation efficiency (Self-made cells)	Transformation efficiency (Commercial cells)
Original BAC vector	$>10^{10}$	$>10^{10}$
Prepared vector+L	<100	<100
Prepared vector+K+L	3×10^4	$\sim 10^9$
Prepared vector+TI+L	10^4	3×10^8
Control BAC (95 kb)	2.9×10^6	6×10^6
Control BAC (170 kb)	1.5×10^5	2.7×10^5

(L - DNA ligase; K - Polynucleotide Kinase; TI - Test insert (total digested genomic DNA))

Even then, transformation efficiencies were found to fluctuate widely in different experiments. On an average, about 150 recombinant clones were obtained from each nanogram of ligated DNA. Thus, on an average, the transformation efficiency obtained was

around 1.5×10^5 colonies per microgram of ligated DNA. The transformation efficiency dropped more than 75% with double size-selected DNA, with hardly a few colonies on each plate. As a rule, the control total digested DNA always transformed much better (>10 times) than the BAC ligations. On an average the transformation efficiency was found to be 0.7×10^7 transformants per μg of ligated DNA.

YAC transformation efficiencies were much lower than bacterial transformation efficiencies. Typically an entire compression zone of size-selected DNA had to be used for every transformation. One such transformation, on an average, yielded around 500 clones of *Lotus* genomic DNA. Considering that about 100 ng of ligated DNA was used per transformation, the average transformation efficiency was around 5×10^3 transformants per μg of ligated DNA. Control transformations with self-ligated plasmid yielded in the range of $1-1.5 \times 10^4$ clones per microgram of DNA. Estimating the right amount of lyticase for optimum (60-70%) spheroplasting of the yeast cells was the most important step. Many times it was not possible to get spheroplasting in the desired range and hence transformation had to be carried out with either under or over-spheroplasted yeast cells (Table 3-2).

Table (3-2)
Spheroplasting of yeast cells for transformation.

Measurement	Lyticase 50 μl	Lyticase 75 μl	Lyticase 100 μl	Lyticase 125 μl
Aw	0.161	0.129	0.091	0.049
As	0.202	0.220	0.218	0.231
% spheroplasted	21	42	58.3	79

(Lyticase amounts given are of a 500 mg/ml stock solution. Aw - Absorbance in water. As - Absorbance in sorbitol)

The 100 μl treated cell fraction was used for transformation because it gave close to 60% spheroplasting. The next fraction was about 80% spheroplasted and may not be good enough to be used in transformation involving HMW DNA. It was also found that different

amounts of lyticase gave optimum spheroplasting in different trials, even though the cells were always harvested at more or less the same stage of growth. YACs had to be incubated for 2-3 days in single selection plates and for a further 2 days on double selection plates for proper selection of recombinant amber colored colonies.

Insert size characterization of HMW DNA clones

The utility of BAC and YAC clones are based on the large insert sizes. All the good results obtained in previous steps are negated if the insert sizes of the clones are small (< 50 kb). Since it is unreasonable to characterize each clone for its insert size, a set of 10-20 BAC clones were selected from each transformation for characterizing the inserts. The results obtained were assumed representative for the entire transformation/ligation and the decision was made whether to proceed with further transformations or not. Despite repeated attempts, it was not possible to produce a BAC transformation with an average clone size of 100 kb or above. This was true for both *Lotus* and soybean. The best transformations had an average insert size of around 65 kb for *Lotus* (Fig. 3-5) and around 75 kb for soybean (Fig. 3-6), while most of the transformations had average insert sizes below 50 kb (Fig. 3-7). In many cases it was found that white recombinant clones contained no inserts at all or that they were so small in size that they ran off the gel. The biggest clones obtained were around 160 kb in size but the ligations were discarded because most of the other clones were below the expected average size.

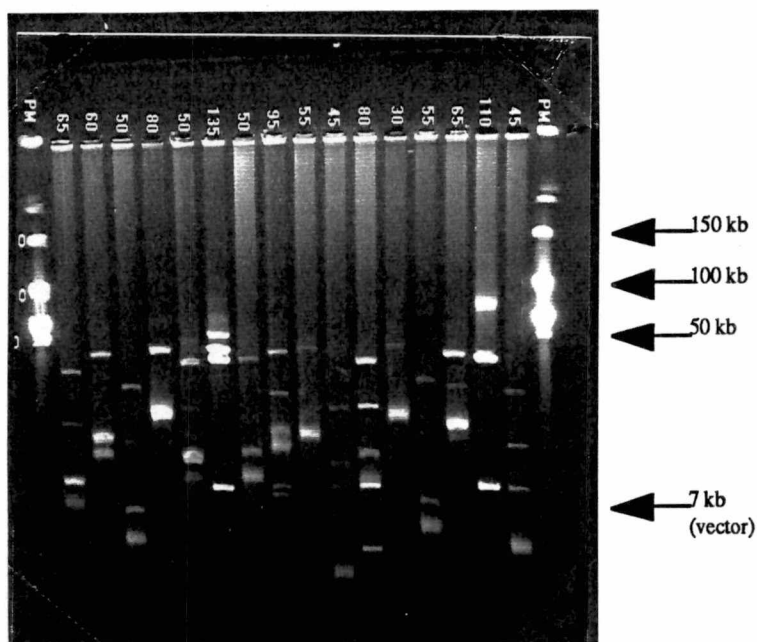


Figure 3-5: PFGE insert size check of *Lotus* BACs. This ligation was chosen to construct the library. Sixteen BACs were chosen at random for this experiment. The BAC DNA was digested with *NotI* to release the insert. The numbers given at the top are the respective insert sizes of each clone. PM (Pulse Marker) refers to the DNA size standard. The sizes are indicated by arrows. The average insert size of this transformation measured at 66.8 kb, on checking 16 clones.

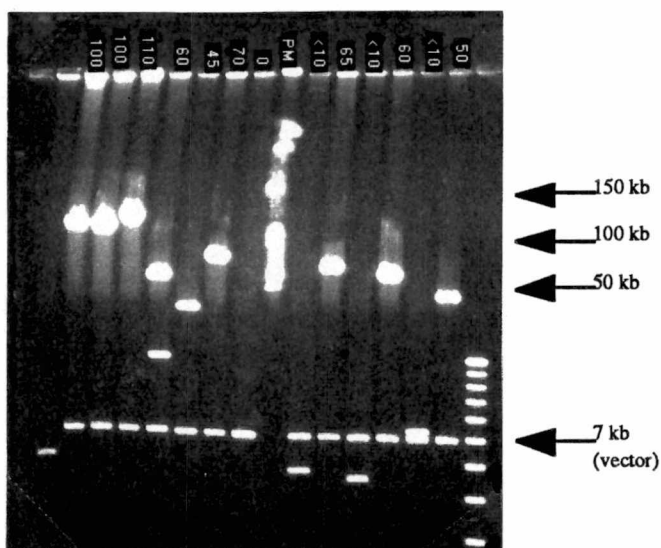


Figure 3-6: PFGE insert size check of a set of soybean BACs. The sizes of the clones are given on top. PM (Pulse Marker) refers to the HMW DNA size standard. The sizes are indicated by arrows.

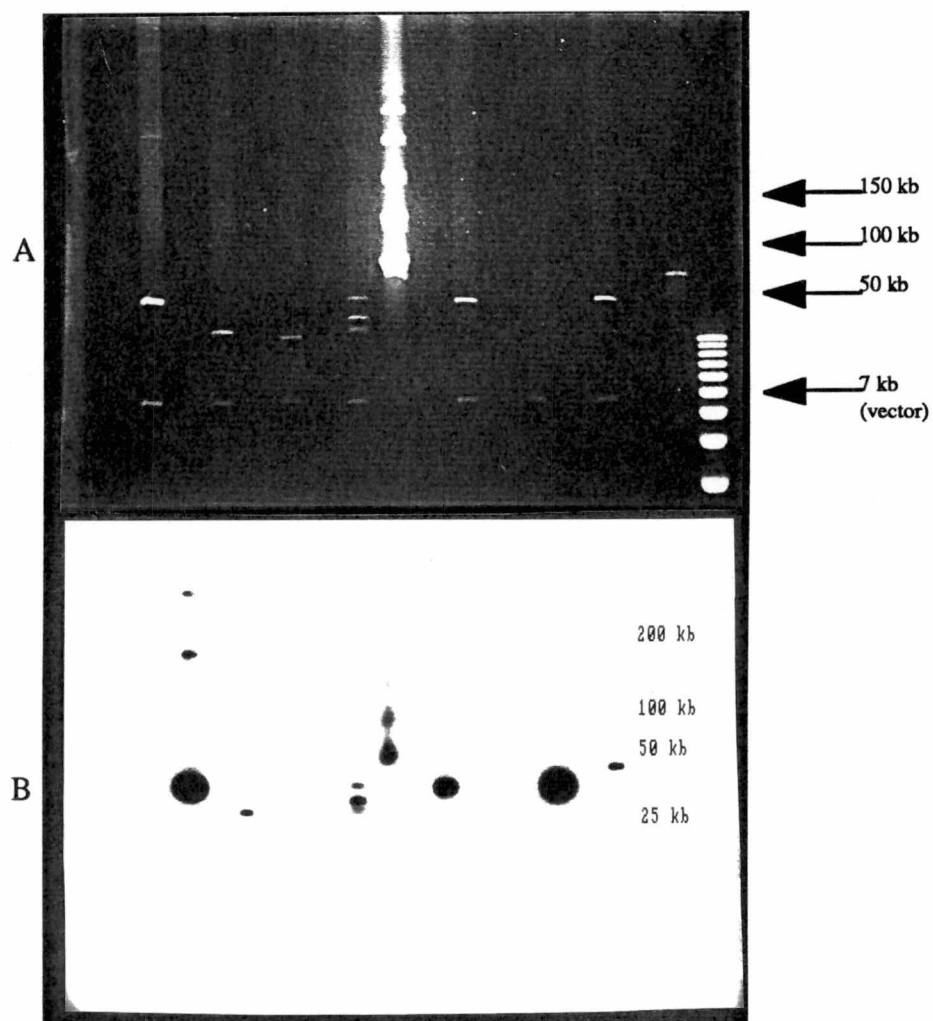


Figure 3-7: An initial set of *Lotus* BACs. The BACs were checked for insert size (A) and hybridized with *Lotus* genomic DNA (B) for confirmation. Note that most of the clones are below 50 kb in size. Sizes are either indicated by arrows (A) or shown in the picture (B).

Double-size selected ligations did yield clones of size 100 kb and above, but the transformation efficiencies were too small to justify proceeding with them. Those ligations could yield no more than a few hundred clones. In almost all the cases, it was noticed that compression zones that clearly showed DNA size-range over 200 kb, gave significant number of transformants below 50 kb in size. Some of them were even smaller than the vector (7.4 kb). The same results were obtained even with the most advanced PFGE system (Bio-Rad's CHEF MAPPER). Despite extensive trouble shooting at the various stages of HMW DNA clone construction, these problems persisted.

YAC insert sizes were better than that of BACs. The only YAC transformation done with *Lotus* DNA showed an average insert size around 150 kb (Fig. 3-8), while most of the soybean transformations had an average insert size below 100 kb. But there were other problems also. Most of the clones showed multiple transformations (more than one YAC in one yeast cell, see Fig. 3-8). In one of the clones, a native yeast chromosome was seen lost while a YAC was found to be present, suggesting integration.

These unreliable results prompted a rethinking as to proceed with YACs or not. As said before, an entire ligation had to be used for every single transformation. Hence, even if the insert sizes were found to be good for one transformation, it was not possible to go ahead and complete the library because another transformation required a new set of size selected DNA. The insert sizes were never found to be constant between transformations even if the same batch of HMW DNA was used following the same procedures.

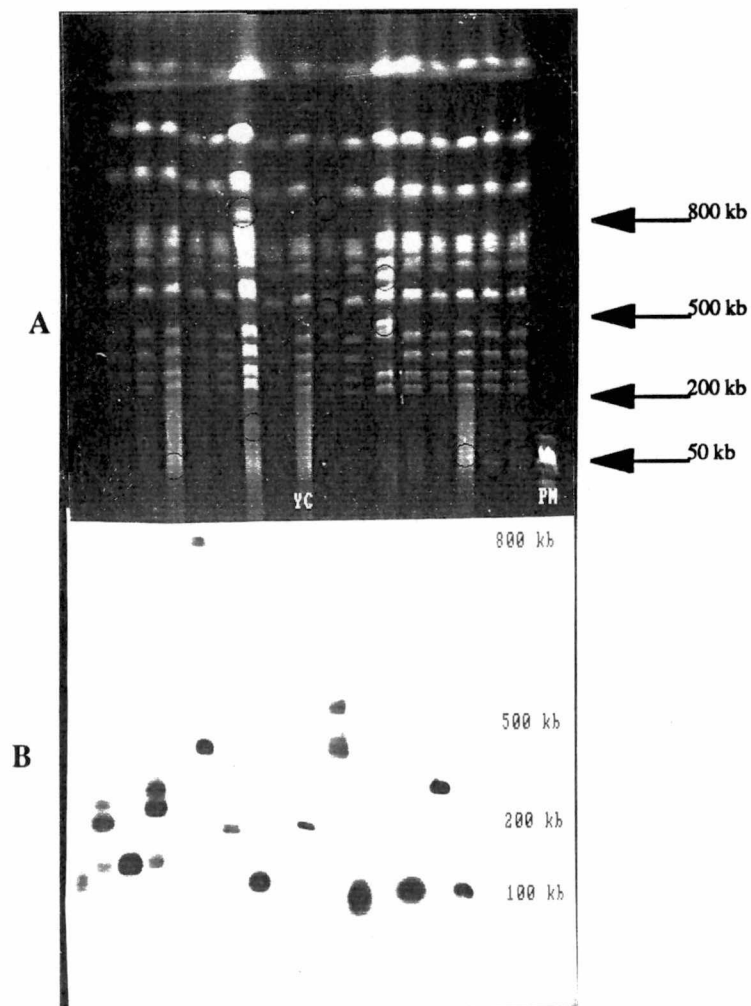


Figure 3-8: PFGE separation of *Lotus* YACs

(A) 15 clones were chosen at random for this experiment. Each clone contains sixteen native yeast chromosomes and a YAC. The YACs (indicated by black circles) are not always visible. The sizes are indicated by arrows. YC refers to control sample of native yeast chromosomes alone. PM refers to a HMW DNA size standard. Note the (B) Southern hybridization of the same set of samples with pBR322 plasmid DNA (which has high homology to the YAC vector pYAC4) to reveal the presence of YACs. Note that some lanes have more than one signal indicating multiple YACs.

Compared to checking BAC insert sizes, YACs were much more difficult and time consuming. It required growing the clones over 2-3 days and later embedding the yeast DNA in plugs. The plugs needed lysis treatments for a further 2 days. Also, many a time it was not possible to detect the YAC until after hybridization with genomic or plasmid DNA. Hybridization results were at times confusing with many signals showing up in a single clone. At times some yeast chromosomes were found to be hybridizing with the probe used. This process was much easier in BACs because the recombinant plasmid is isolated free of contaminating chromosomal DNA and the insert released from the vector by digestion with the rare cutting *NorI* enzyme.

The southern transfer of HMW DNA was not found to be problematic as expected. Even the largest yeast chromosomes (>2 Mb) were transferred efficiently. This procedure did not require any special adjustments except longer depurination and transfer times (Thesis pp 52). In case of YAC hybridization, it was found that the plasmid probe gave much more clearly defined signals than the genomic DNA probe. This might possibly be due to eukaryotic DNA homology, that gives diffused signals with the genomic DNA probe.

Utilization of BACs

This section describes the results obtained from the different experiments that were done to assess the utility of BACs in positional cloning. As stated before, the DNA sequence targeted for screening the BACs was pUTG-132a, an RFLP marker that is closely linked to the soybean supernodulation gene (*nts*; see introduction)

Screening of BAC pools for pUTG-132a candidate clones

A PCR based screening method was followed to isolate BAC clones that might harbor the pUTG-132a marker.

A GENBANK search was done to get the exact sequence of pUTG-132a and to decide which primers to use for amplification. The results were as follows, with the

primers in italics underlined. The primers shown are LII, LI, RIII and RI respectively from 5' to 3'.

5'

CTGCAGCAGT GTTGGGCATG TCTCTCTTTG CTGAAACAAG AATCGCATT A CTGTCCTCT
TGCACATTAA AAAATAAAGC TACCGACTTG TGAGTGAAAA GAAGCAGAGA GGTAAGGTAG
AGCCTGAGCC CTTGTGTCTT TGTTCGATTT CATTTAATAT AGAATTGTCT ATCTAATGCT
TAATGAACAT AAGGTCTGAA CTTGAGCCT AGCATCTATT AAATTAAATA TTTGACTTAG
AAACAAAGAA GACGTTTCGAT AATTTTTTAAC ACTTGCGTGG TTTTCTTTA TTGCTTGAAA
CATACAGACA TGCTCTTTTT TGGAATTGTT TTTTCTTCAA ATCTAAAGAG ATATCAAGCA
AGGAACAACA GCTATAGCAA CATACCAACA ATCCGTTAAA TTAGTAGAAA ATTTAGTATT
TTATGCTAGA ACGAGGCATA TTTAGAGGTG CATTATTCTA GATTTTTTTT TTTTGCGAGT
GACACGAAGT ATATAAAAAA ATGAATGCCA AATTACACAT GGATATACAA ATAACGAT
ATATTATGAT CCTGAAGCAT TCAGAGAAAA ATGTAATAAT TATTTTCTAC GCAATTTCAC
GGTCACTCAT GGGCCATGGC AAGTAGGAAA ATTGCTGCCA CAGGCTATGT TACATAGTCG
AGTTGAACAG CTGGACTAAT ACTTGTCGTA GTTAATATAA TTATGTGTGA ATAGTAAAGA
TAAATACTTT TGA CTGGACC ACATATAAAT TATAAGATTG ATTTAATGGT AAAAAAGAA
AAAAAGTGAA AAAAATTGTG TGTTTAATTT TTCCGTCAAC AAAAATTAAT AATTAACAAC
TAAACTAATA TTTACTGATA TTAGGGTAAA AACTTTGCC TCTCATTTTT TAATTATTCT
ACCGAATTAT TGATTATTTT AACACTAATT ATTTGTCATT CTAAAAAGTG TACTAATTAT
TTCTTTTCTA GTAATCTCTC TATTTTTTTC ATATTTTCATC TAATTAAAAAT TAAATTACAA
CTATATATTT TAACTTATTA TTTTGTGTTT AAATATTCAT TTTTATCAAC ATAACCTATA
TAATTAATAG AAAAAAATTA ATATTAAATG AAAAGCTAAG ATGATATTTA TTAGGAGTAT
TAATAGAAAC AAATAATATT GCATTTCAAA ATTAACATAA CTCTTATTTT GAATTATTTT
TTCTTCAAAT GTGATACCTT TTAGAACGAG TGAGTGATTA AGGGAGACTT ATTTTAAGAA
ATTTATAAAA GCTGAAAATT AACTTCAAG TACGTTTTTA ATTTATTTTG GATGGTTATA
AAATCCCTTT TTGATTGCAA AAATATTATT TTAATCCTCC CGCCCTCACT TTGTCTGTGC
CTTCGTTTAG GTGCCCTTTG CTCTTTTCTG TTCCCTCCGC GCGAGGTAGG TTTTTTACGT
CTGCCCCAGT AGTTCTGTTC TGTTTCTTCT GTTTCACCTT CACTGTTGAT CATGCCTATC
AGTGGTTTAT GCTTCATTTT CATAAACTG GGGTGAAAAT TTGTTCAAGC ATTTAACGTT
GTAGCCATAA CATAGTACAA TCATCTCCAC GATTCACCTG GAAATGGGG CTGGATGTCC

TGAGTAAGGA CTTTTACTGA GTTTAAATAC TGTAACATTC TCATCATTAT TTTGCTCCCT
TATTAATTAA AAGGGGCTTTT GGAATCAA TTCTGCAG 3'

To standardize PCR conditions, genomic and plasmid DNA were amplified as controls, using the same primers. The PCR products showed up at the expected sizes, at 1.7 kb for Bragg genomic DNA and the cloned pUTG-132a DNA and at around 850 bp for *G. soja* DNA. The amplification was very specific and the products were sharp. It was found that higher primer concentration somehow inhibits amplification of genomic DNA. A primer concentration of 0.1 μ M gave a strong and specific product of the expected size. At a primer concentration of 0.25 μ M, the product appeared much fainter. At 0.5 μ M primers no amplification could be detected. The cloned plasmid DNA amplified at all concentrations of the primers, but there was a progressive decrease in the amount of the product.

However, the same conditions did not give any amplification product when used for amplifying the BAC pools. This could be because of the inhibitory effects of some bacterial components found along with BAC DNA (see discussion). Hence, different PCR conditions had to be worked out (Thesis pp 56). The conditions had to be worked out in such a way that not only the BAC pooled DNA showed some amplification products, but also the control plasmid and genomic DNA. *G. soja* DNA was omitted at this point, since it was not of relevance. When the 28 sub-pools from each master pool were screened a lot of non-specific signals showed up. It was difficult to pick up candidate clones from those results. Hence, for further confirmation, the gels of PCR product electrophoresis were blotted and hybridized with purified pUTG-132a fragment DNA. Even after hybridization, the results were confusing with many pools showing the desired size product (Fig. 3-9A & B and Fig. 3-10). The same experiment was repeated for pool 40. While testing pool 40 the PCR products from all 28 pools were run in the same gel using two rows of wells. This allowed uniform hybridization conditions and made the work easier (Refer to Fig. 3-10 for details)

A

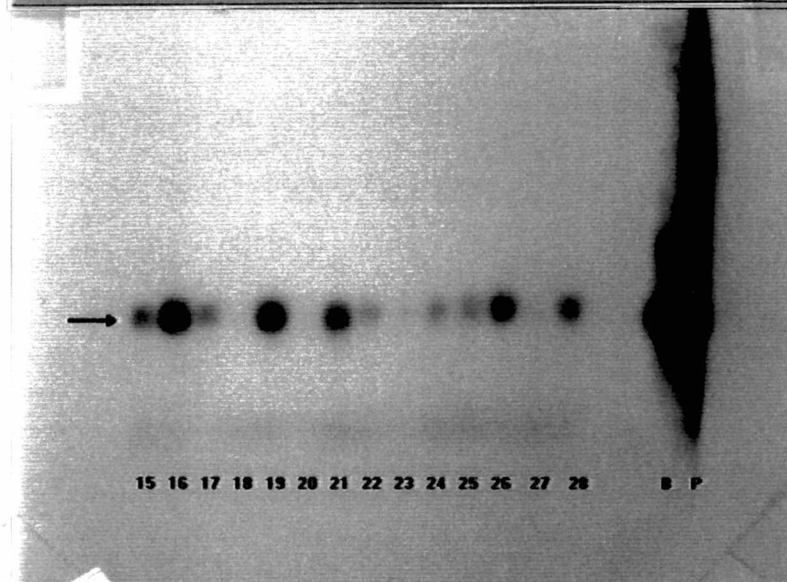
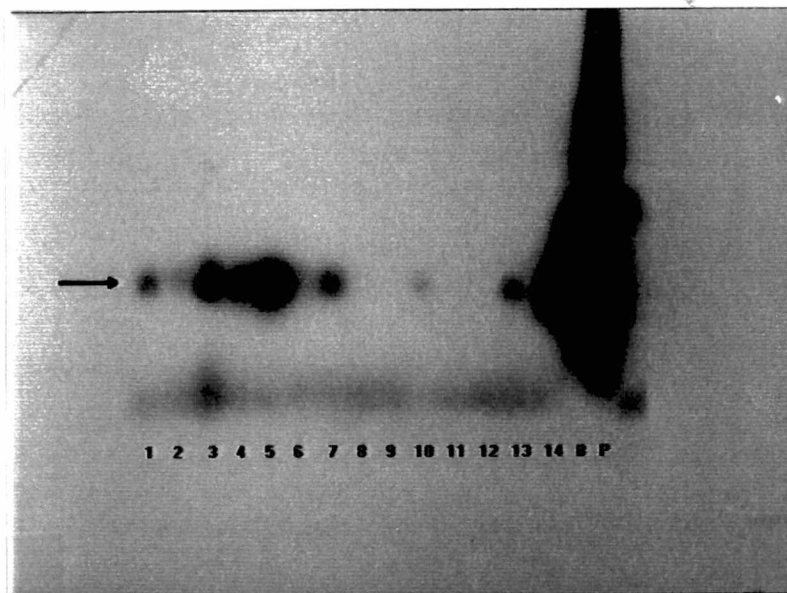


Figure. 3-9
Hybridization of the
PCR products
obtained on
amplifying pool 14

(A) Sub-pools 1-14 of pool 14 amplified with primers LI and RIII and hybridized with pUTG-132a. The arrow indicates the PCR products of 614 bp. Lanes 1-14 indicate the sub-pools. B - control *G. max* "Bragg" DNA amplified by the same primer pair. P - control pUTG-132a plasmid DNA amplified by the same primer pair. (B) Hybridization of the PCR products, obtained on amplifying sub-pools 15-28 of pool 14 with the same primers mentioned above.

B

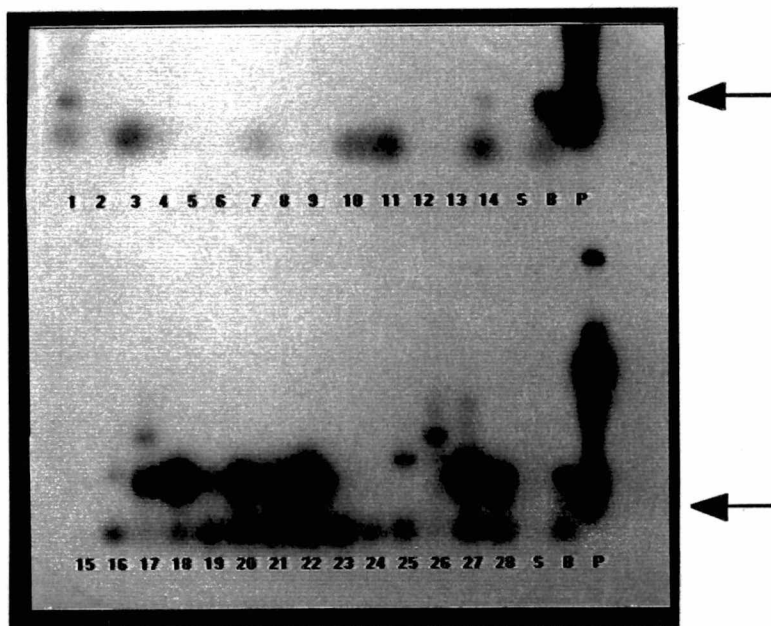


Figure 3-10: Hybridization of the PCR products obtained from amplifying pool 40. Sub-pools 1-28 of pool 40 were amplified using primers LI and RIII. The expected PCR products of 614 bp are indicated by the arrows. Lanes 1-28 indicate the 28 sub-pools. B- Control *G. max* "Bragg" DNA; P- Control pUTG-132a plasmid DNA, both amplified by the same pair of primers.

It was later decided to use the intensity of the product signal as a criterion for picking up the candidate clones. The pools showing the strongest intensity of hybridization were selected for further analysis. The pools were then aligned with each other as "X", "Y" and "Z" axes. The intersection points were used to pick up the individual clones. The following clones were picked up as candidates:

Pool 40 - Plate 1, Row H, Columns 2, 3, 5 & 10.

[BACs 1, 2, 3 and 4]

Pool 14 - Plate 3, Row H, Columns 3, 5 & 10.

[BACs 5, 6 and 7]

Plate 5, Row H, Columns 3, 5 & 10.

[BACs 8, 9 and 10]

The similarity in the numbers is purely co-incidental. The clones were numbered BACs 1-10 in the above order.

Confirmation of the candidate BAC clones

The candidate BACs were digested with *NotI* and electrophoresed to view the size of the inserts (Fig. 3-11A). This would eliminate some of the candidates if they are found to be too small for building contigs.

The sizes of the BACs from 1-10 were as follows:

BAC1-50 kb, BAC2-60 kb, BAC3-165 kb, BAC4-90 kb, BAC5-70 kb, BAC6-55kb, BAC7-65 kb, BAC8-60 kb, BAC9-155 kb, BAC10-50 kb.

Hybridization with pUTG-132a revealed homology in clones 1, 3, 4, 5, 7 & 8 (Fig. 3-11B). However, except in clone 3, the signals were very faint. It was concluded that hybridization with such large sizes of DNA might not be very efficient. Clone 3 was found to be a definite candidate. It was possible to eliminate clones 2, 6, 9 and 10 from further studies.

To enhance the strength of the signals, the individual BACs were digested with *PstI* to release the 1.7 kb fragment. However, electrophoresis did not clearly reveal a 1.7 kb fragment in any of the clones. Hybridization with pUTG-132a revealed a very confusing picture. The probe was found to clearly hybridize with many of the digested fragments in each clone. Thus this experimental approach did not help in narrowing the search.

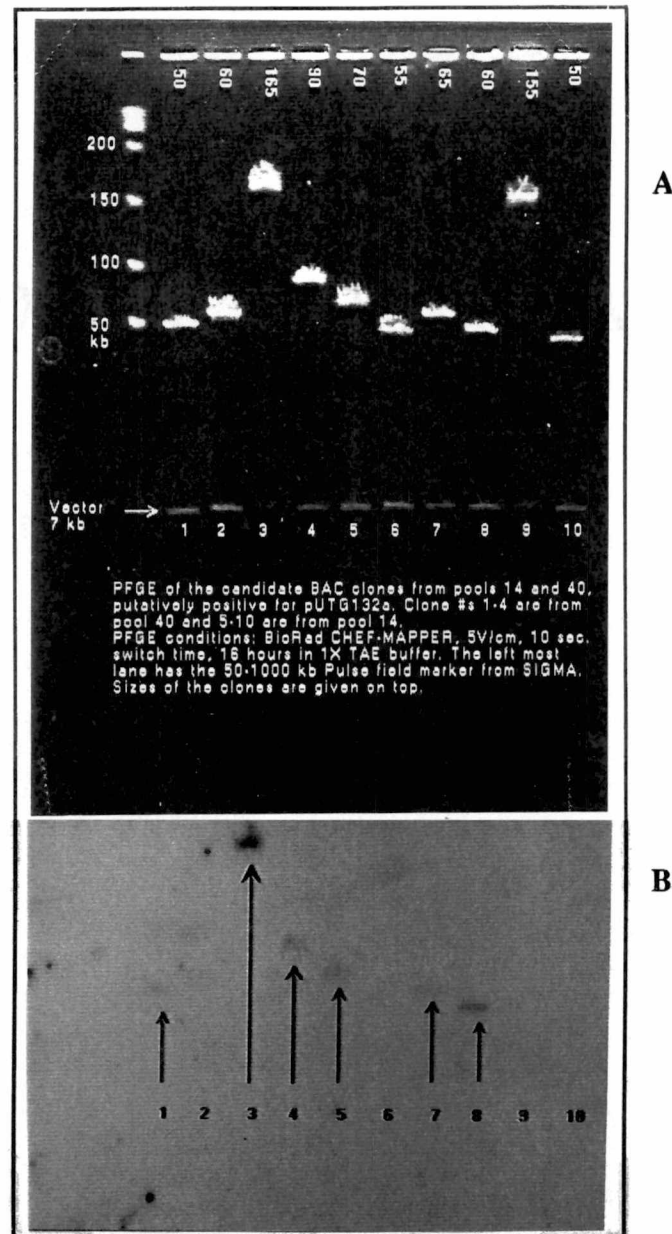


Figure 3-11: PFGE of the candidate BACs (A) PFGE of the candidate BACs after digestion with *NotI* to excise the insert from the vector. Details are given on the picture. (B) The same gel boltted and hybridized with pUTG-132a to reveal homology. Except in the case of BAC3 the signals (indicated by arrows) are weak.

To narrow down the search further, each of the candidates were amplified by a nested PCR. This had to be done because regular PCR failed to amplify any of the clones. BACs 6 and 8 were omitted at this stage because they failed to give any signals on whole-BAC hybridization. The first round of PCR amplification was done with primers LII and RI. After this, a second round of amplification was done with primers LI and RI (primer LI has an annealing site within the first PCR product). This approach helped to isolate three of the ten clones as definite candidates. Nested PCR amplified clones 1, 3 and 5 (sizes 50, 165 and 70 kb, respectively) and gave the right size product of 1.7 kb. This was revealed after hybridization of the PCR products with pUTG-132a (Fig. 3-12).

After this experiment it was decided to proceed only with clones 1, 3 and 5.

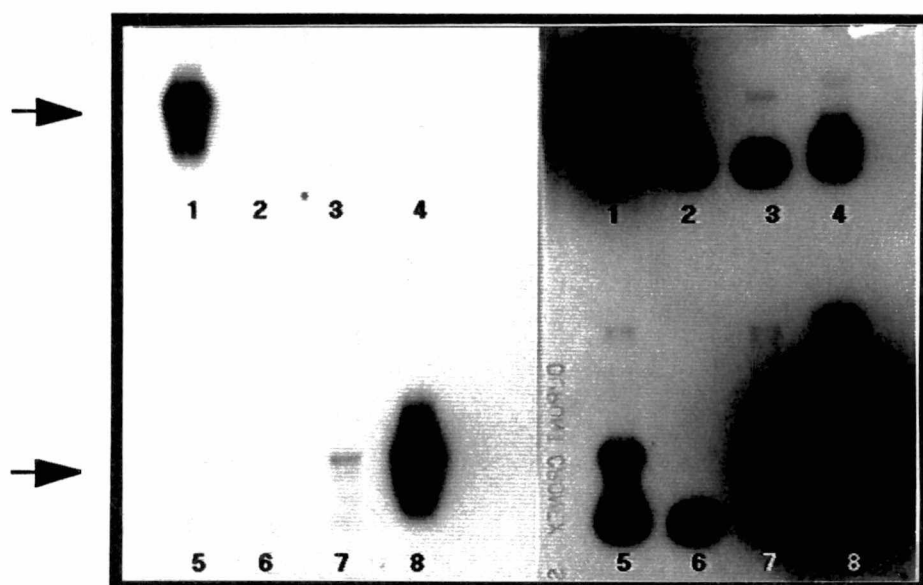


Figure 3-12 : Nested PCR amplification of the 8 candidate BACs. The expected PCR product (1.1 kb) is shown by arrows. The lanes 1-8 contain BACs 1, 2, 3, 4, 5, 7, 9 and *G. max* genomic DNA (positive control) respectively. The blot was exposed for 5 min. (left panel) and 120 min. (right panel).

Isolation of endclones

After isolating the three candidate BACs, the next step was to isolate the end clones to facilitate contig building in the direction of the *nts-1* locus. Both the right and left end clones were isolated by plasmid end rescue. Three enzymes (*Eco*RI, *Kpn*I and *Sac*I) were used for isolating the right end clone, while *Hind*III was used for isolating the left end clone. Three clones were checked for size from each restriction digestion. In all the cases, the sizes of all three clones from each set were found to be the same. The sizes of the end clones varied from 2 -13 kb (Fig. 3-13 and Fig. 3-14). As part of this work only the right end clone was characterized further.

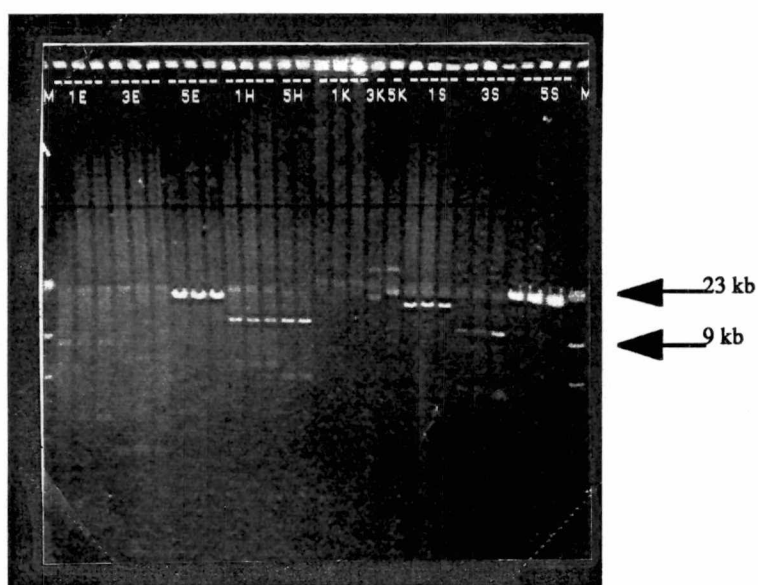


Figure 3-13 End clone analysis: Endclones isolated from the three candidate BACs and digested with the respective enzyme. The sizes are shown by arrows. Lane assignments are as follows: M - DNA markers; 1E, 3E & 5E - *Eco*RI isolated endclones from BACs 1 (3 clones), 3 (3 clones) and 5 (3 clones) respectively; 1H & 5H - *Hind*III isolated endclones from BACs 1 (3 clones) and 5 (2 clones) respectively; 3K & 5K - *Kpn*I isolated endclones from BACs 3 and 5 respectively (1 clone each). 1S, 3S & 5S - *Sac*I isolated endclones from BACs 1 (3 clones) 3 (3 clones) and 5 (3 clones)

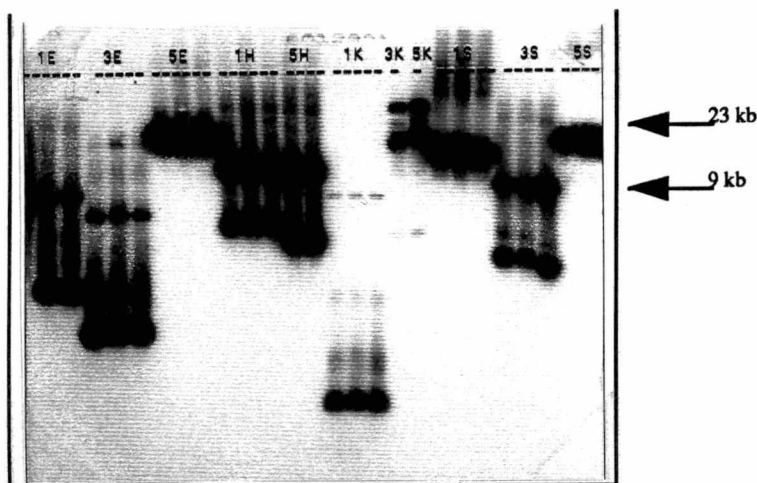


Figure 3-14:
Hybridization of
endclones. BAC end
 clones hybridized with
 soybean genomic DNA for
 confirmation. The legend
 for Fig. 3-13 holds goods
 here also.

The *Sac*I-generated endclones were taken up for further analysis, for all the three BACs. Approximately 500 bp of the right end clones (RE) of BACs 1,3 and 5 were sequenced and PCR primers designed for further screening of the library. The samples were sent to the University's Molecular Biology Resource Facility. Sequencing was done using the M13 reverse primer at a DNA concentration of 2.5 μ L per reaction. The sequencing results were as follows, with the primers underlined. (Primers were designed only for BAC5).

Sequence of BAC1 right end clone

TGACCATGAT	TACGCCAAGC	TATTTAGGTG	ACACTATAGA	ATACTCAAGC	TTGCATGCCT
GCAGGTCGAC	TCTAGAGGAT	TCTTCTTTA	TTTATCTGCT	AGTAGACCTG	ATATTATGTT
TAGTGTATGC	ATGTGTGCTA	GGTTTCAATC	CAACCCCAA	CAATCACATT	TAAGTGCAGT
AAAGAGAATC	ATGAGATATC	TATTAGGAAC	AATCAATTTA	GGATTATGGT	ATCCTAAGAA
TTCAACATGT	AACTTAATAG	GATATTCTGT	TTCTGATTTT	GCCGGATCTA	AAACTGATAG
AAAAAGTACA	AGTGGAACCT	GTCAATTTAT	TGGATCGGCT	CTTGTCTCAT	GGCATAGTAA
GAAACAAAAC	AGTGTTGCTT	TATCTACTGC	TGAAGCGGAG	TATATCTCTG	CTGGTAGTTG
TTGTGCACAG	ATTTTATGGG	TTGAAAGCAA	CAATTATCTG	ACTATGGCAT	CATTCTTGAT
CGCATACCTA	TTAAGTTGTT	GATATTACTA	GTGCCATAAA	TCTATCCAAA	AACCCAGTTC
CAACACTCNA	GAACTAAACA	TATAGAAGAA	TTAGACACCA	CTTCTTTAGA	AAATCATGTT

CCTAAAGGGA	NAATGTGTAT	TANAATTTGT	TTTGAATACT	AAGAAATCAA	CNTGCTGAAT
ATTTTCCCTA	AACCTCTCCC	AGGGAANTTT	TTCTTCCCTA	TAAAAAAAAA	AATNAGGTCT
CTTAAANTTN	NNAAATTANN	AAAATGGAGA	NTGGNNNNNN	NNNNNNNNNG	NNNNNNNNNN
NNNNNNNNNN					

Sequence of BAC3 right end clone

AACANCTGCC	GACTTNAATT	ACGCCAAGCT	ATNTAGGTGA	CACTATAGAA	ATACTCAAGC
TTGCATGCCT	GCAGGTCGAC	TCTAGAGGAT	CATTCCAAGT	TAACTCTTTA	ACTTAATCAA
CCTTTCTTCA	GTTAATGATT	TTATAAAAAT	ATCAGCAACT	TGTTTCATCAA	TGGGTACAAA
CTATAAATCC	ACTGTACCAT	TTTGAACATG	ATCTCATATA	AAATGATGTT	TAATTTTTAT
ATGCTTTGAT	TGGGAATGCA	ATGCTAGATT	TTTTGAAAGA	CTTATAGTAT	CTCTATTATC
ACAATAGATA	GGAATTTTAC	TCTCATAGAT	GCTATAGTCT	TCAATATCCA	TAGAAGTTGA
GTGCAACAAC	TTGTTGTTTG	ACATGTATTT	TGCTTCAATA	GTAGATAATG	CACTTGATCC
TTGCTTCTTG	CATATCCATG	TTGACTAAAG	TTTGCATCTT	ATCAAGTGAA	AAACTTCCAC
TTGTACCTTT	TCTTTCAATT	TTATCACCAT	TATAGTTTTC	ATTGCAATAG	CTTGTGATCC
CTGAAAACCT	TCTCTTCTTT	TGGAACAAGA	GACAGGTTAG	AAAANTTCCT	AATTAAATGT
CTGGATNATG	TTCCNANNNC	CNNNNAAATG	AACNCCCNT	GAACCTTTNN	GNANTCNANA
ACNNCNNCGN	NCNTNNNCTC	NTGTNNNGCN	NGANGCCTNN	GGGNNAANCT	NCTTNNNCNN
NNNNTNGNNC	GGNNNNNGGN	CNNGCNCGCT	CCNNNNNNNC	CNCNNNNNNC	NNNNNCNNGN
NNNNNNNNCG					

Sequence of BAC5 right end clone (primers underlined)

GACCATGATT	ACGCCAAGCT	ATTTAGGTGA	CACTATAGAA	TACTCAAGCT	TGCATGCCTG
CAGGTCGACT	CTAGAGGATC	<u>TTTGTCTTAA</u>	<u>CACATTGGAG</u>	<u>GATACATCCT</u>	TTGTGGTATA
AGTAGAGGGT	ACATCTACTT	GGGTTTGACT	GAGAACAAGA	GAGGATACAT	CTCTTGTTGA
TCAGTTCTAG	TGGAGGGTAC	ATCCACTAGG	TTCAAAGAGA	ACAAGGGAGG	GTACATCCCT
TGTGGATCTT	TGCTTGTAAG	AGGATTTTTA	CAAGGTTGAA	ATAAATTTCA	AGGACCGCAG
GTCGCTTGGG	GACTGGATGT	AGGCACGGGT	TGTTGCCGAA	CCAGTATAAA	AACTCTTGTTG
CGTTTGTTTC	CTTCTTCCCT	ACTCTTTTAC	TTTCCGCTGT	GCATTTAATT	TTGCTTTTAA
CTTTCTGTTA	AGTTTCTCTT	CTACTCCTCA	TTCTCTTAAC	AATTTANTAA	ATGCCTTANA

AANAATAAAT	TTTTTAATTA	GTAAAGGTTT	AAGGAATAAT	<u>TAATTCAACC</u>	<u>CCCCCTTCTT</u>
<u>AATTATNCTG</u>	AAGNCACCCA	TCCAATNTGA	TNCNAAAATC	CANTGCTANC	NCANCTAATG
TTNNNNCCCC	CCTNNCTATA	CNNNAAGGGT	NCNANAAAAT	GGCTNCCNCN	ANNTTTCCCN
NNCTNNNNNN	NNGAANNCNN	AAANNAANTN	TGGGGGGAAN	NAAAAANNNN	NNNNCNNNNN
NCCNNNNNNN					

On searching the database, it was found that the right endclone of BAC1 (RE1) contains a 80 bp motif that has 85% homology to a maize retrotransposon, *Opie-2* (SanMiguel et al, 1996). A BLAST search was done to reveal this (see search results below).

BLAST search results for RE1

BAC#1 endclone

Zea mays retrotransposon *Opie-2*

```

TCTTCTTTATTTATCTGCTAGTAGACCTGATATTATGTTTAGTGTATGCATGTGTGCTAGG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TTTACTTTATTTATGTGCTAGTAGACCGGATATTATGCTTAGCGTATGCATGTGTGCTAGA

TTTCAATCCAACCCCAAACAATCACATTTAAGTGCAGTAAAGAGAATCATGAGATATCT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
TTTCAATCCGATCCTAAGGAGTGTCACTTAGTGCGGTGAAGCGGATTCTTAGATATTT

ATTAGGAACAATCAATTTAGGATTATGGTATCCTAAGAATTCAACATGTAACCTAATAGG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
GGTTGCTACGCCTTGCTTCGGGCTCTGGTATCAAAGGGGTCTACCTTTGACTTGTTGG

ATATTCTGTTTCTGATTTTGCCGGATCTAAACTGATAGAAAAAGTACAAGTGGAACCTG
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ATACTCAGACTCCGACTATGCTGGATGTAAGGTCGATAGGAAGAGTACATCGGGGACGTG

TCAATTTATTGGATCGGCTCTTGTCTCATGGCATAGTAAGAAACAAAACAGTGTTGCTTT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
CCAATTCTTAGGAAGGTCCCTGGTGTCGTGGAACCTCTAAGAAACAACTTCTGTTGCCCT

ATCTACTGCTGAAGCGGAGTATATCTCTGCTGGTAGTTGTTGTGCACAGATTTTATGGGT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
ATCCACCGCTGAGGCCGAGTATGTTGCCGAGGACAGTGTTGCGCGCAACTACTTTTGAT

```

Retrotransposons are known to be highly repeated throughout the genome. However, for endclone mapping there is a need for unique sequences and hence this is not desirable. So it was decided to proceed only with BACs 3 and 5. The right endclones of BACs 3 and 5 were termed 3RE and 5RE respectively

The 5RE primers (a 27mer left primer and a 28mer right primer) were designed in such a way that the product would be 471 bp in size. At first, the end clone primers were used to amplify the sequences from the BACs and from the end clones themselves. It was found that the primers amplified not only BAC5 but also BAC3 (Fig. 3-15). The primers, however, did not amplify BAC1. The same was true for the end clones as well. Since 5RE-specific primers amplified BAC3 total DNA and a PCR product of the same size was obtained, it is possible to speculate that BAC3 and BAC5 are closely related sharing at least two STSs - pUTG-132a and 5RE.

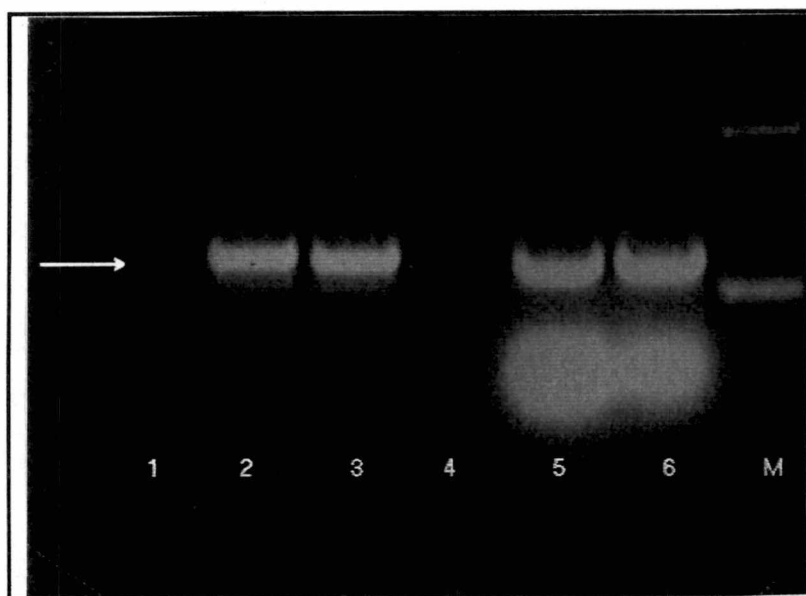


Figure 3-15: PCR amplification of BACs using 5RE primers. Lane assignments are as follows: 1 - whole BAC1; 2 - whole BAC3; 3 - whole BAC5; 4 - BAC1 right endclone; 5 - BAC3 right endclone; 6 - BAC5 right endclone (positive control). The expected PCR product (471 bp) is shown by the arrow.

Later, the right end clone of BAC5 (5RE) was PCR amplified from the parental genotypes *G. max* "Bragg" and *G. soja*. On sequencing, the PCR products were found to be identical (see sequencing results below). Hence, there was no polymorphism between the parents. This made it impossible, at present, to map 5RE using the F2 segregating population.

Sequences of the PCR products amplified from the parental genotypes *G. max* "Bragg" and *G. soja* using 5RE primers. From comparing sequences, the products were found to be nearly identical.

Glycine max "Bragg"

CGGTAANNNN	TGTGTTAAGN	AANAGAGGGT	ACATCTACTT	GGGTTGTACT
AANAAAAAAA	GAGGGTGGGT	CTCTCNTGNA	TCAATCCTNN	TGNANGGTGN
NTCCNCNNAG	TTNNNAAAAA	AAAAAGGAGG	GGGNANNNCC	CTTCNATNTT
GGNTCNTGGN	NGGAGATGGT	NANAGNGNAG	NAAAAAANNN	CNGACACCGN
GTGNCGGNGG	GAGAGGAANT	GNNTNNGGGT	NNTGTTNNAA	NCNAAACAAA
AANNCTNTTN	TNTTTTTTTC	CCCCNCCCCC	CCNCTCTTNT	TTTTNCCCNT
GTGTNTANTT	TCCCCCCTN	NTTNTNTNTN	GGTTTCTCTC	CNNCNCNCCN
TNNTCTTTTC	ANNANNNTNT	AAANNANNCN	AAAANAANTT	TTTNTTTNNN
NNANNGGGGG	GGNNNANCAN	CNCCCCCCCC	CNNNNNNNNN	NNNNNNNNCC
CCCCCTTTTN				

Glycine soja

GNTNTNAANN	TGTGTTAAGN	AANAGAGGGT	ACATCTACTT	GGGTTGTACT
AANAAAAANA	GAGGGTGGNT	CTCTCTTNNA	TCAATCCTNN	TGGANGGTGG
NTCCNCNNNG	TTNNTAAAAA	AAAAAGGAGG	GGGNAGGNCC	CTTCCNTNTT
GGNTNTTGGN	NGGNGATGGT	TNNNGNGNNG	NAAAAAANNN	CNNACACCGC
GNGNNGGTGG	GAGNGGAAAT	GTNTNTGGGT	NTTNTTNNA	NCNAAANNA
NANNCTNTTN	TNTTTTTTTC	CCCCNCCCCC	CCNCTNTTNT	TTTTTNCNNT
GTNTTTNNTT	TNNCCCCCTN	TTTNTNTNTN	GGTTTNTNTN	CNNCNCNCCCT
TNNTCNCTTN	ANNANCNNNT	NAANNANACC	NAAATAATTT	TTTTNTTTNN
NANNGGGGGG	GGNTTANNNN	NNCCCCCCCC	NNNNNNNNNN	NNNNNNCCCC
CCCTTTNTTN				

A significant result came about when *Lotus japonicus* DNA was PCR amplified using 5RE primers (Fig. 3-16). It was possible to obtain a putative polymorphism between the two *Lotus* ecotypes Gifu and Funakura. However, this result is preliminary needs to be confirmed.

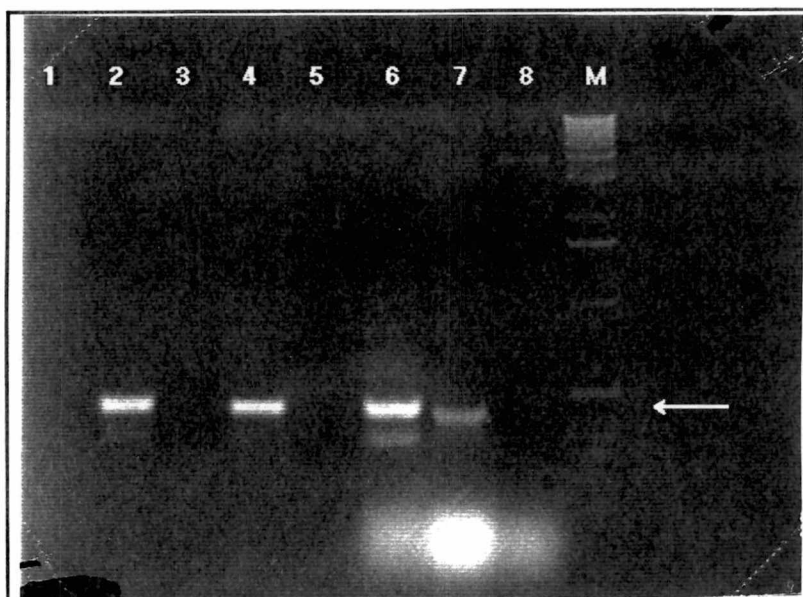


Figure 3-16: PCR of soybean and *Lotus* genomic DNA using 5RE primers. Lane assignments are as follows : 1 - BAC1; 2 - BAC3; 3 - BAC4; 4 - BAC5; 5 - BAC9; 6 - *G. max* "Williams" genomic DNA; 7 - *L. japonicus* "Gifu"; 8 - *L. japonicus* "Funakura"; M - Marker. The expected PCR product (471 bp) is shown by the arrow. BAC9 was included because of its large size.

In one of the confirmatory experiments it was possible to directly PCR amplify the 1.1 kb LI-RI primer pair product from BACs 3 and 5 as well as from *Lotus japonicus* "Gifu" (Fig. 3-17). This was made possible by using a lower annealing temperature (55° C instead of 60° C). It should be noted that previously, only nested PCR could amplify this product from the BACs.

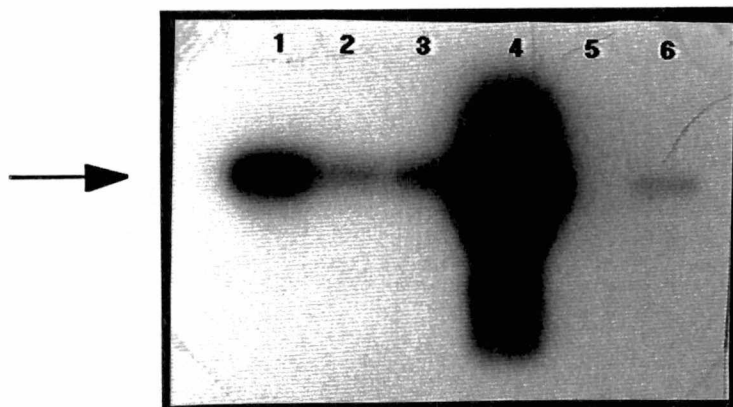


Figure 3-17: PCR amplification of a 1.1 kb pUTG-132a fragment from the candidate BACs and *Lotus japonicus* genomic DNA. Lane assignments are as follows: 1 - BAC3; 2 - BAC5; 3 - BAC9; 4 - G. max genomic DNA (positive control); 5 - *L. japonicus* Funakura; 6 - *L. japonicus* Gifu. The gel was blotted and hybridized with the pUTG-132a fragment. The 1.1 kb PCR product is shown by the arrow.

There is a possibility that a sequence similar to pUTG-132a is present in *Lotus*. If so, this makes it possible to put 5ER into the *Lotus* genetic map and opens up a new avenue for isolation of the *nts* gene, through synteny between *Lotus* and soybean. The advantages of using *Lotus japonicus* to isolate the *nts* gene are discussed in detail in the discussion section of this thesis.

CHAPTER IV

DISCUSSION AND CONCLUSIONS

Positional cloning has necessitated the construction of plant HMW DNA libraries. Plant HMW DNA libraries have been slow to come when compared to other systems (see introduction) . This might have been due to the lack of marker-enriched genetic maps or technical difficulties associated with isolating and manipulating HMW plant DNA libraries. However, with the advent of the model plant *Arabidopsis thaliana* and sharing of resources and technical expertise among the various research centers, a lot of progress has been made in positional cloning in plants (see introduction). Most of the major cultivated crops either have well-represented HMW DNA libraries or are in the process of constructing them. This prompted us to go in for an attempt to construct and characterize HMW DNA clones in the legumes soybean and *Lotus*.

In this work many fundamental technical questions regarding the construction and characterization of HMW DNA libraries were addressed. Progress was also made in identifying BAC clones that might harbor the pUTG-132a marker. Technical problems hampered the process a lot more than expected. These problems were encountered at various stages of HMW DNA cloning and took a lot of time to solve or address.

Technical aspects of HMW DNA library construction

At the stage of HMW DNA isolation itself technical problems were encountered and observations could be made. The growth conditions of the plant were found to influence the quality of HMW DNA isolated. Plants that were grown in growth chambers under controlled conditions and leaf material harvested at the right stage gave much better starting DNA than plants that were grown in green houses without proper environmental controls. When the plants are grown under not very precisely controlled conditions, the fluctuations might induce stress in them. Depending upon how close to these fluctuations were the leaf material harvested, it could harbor different degrees of DNA degrading substances, which in turn influence the stability of isolated HMW DNA. Some of the batches of HMW DNA isolated were not easily amenable to restriction enzyme digestion. The stress that the plants

were exposed to could trigger the induction of substances that give the plant the ability to withstand the stress. These kind of substances might have proteasing activities that might disrupt the activity of restriction enzymes. The difficulty in digesting the DNA could also be because of physical barriers like excess starch or other carbohydrates that limit the access of the restriction enzymes to the DNA. At times, even overnight incubation with the enzyme did not yield sufficiently digested DNA. This was especially true in case of nuclei plugs where a lot of contaminating material could be present because of the larger amount of leaf material used. Even repeated filtering and cleaning of the nuclei solution could not prevent the occurrence of contaminating material. This implies the importance of starting with the best possible material for HMW DNA isolation.

One of the most significant differences between normal cloning and HMW DNA cloning is that the DNA used for HMW DNA cloning is partially digested and size-selected for large DNA molecules. Partial digestion is achieved by digesting the DNA with minimal amounts of enzyme for short times and size selection is achieved by PFGE. The titration conditions were such that the undigested control sample would stay way up in the gel while the digested samples would run down through the desired size range. But that degree of assessment could be achieved only a few times. Many times, the undigested and the digested samples looked nearly the same in the HMW region, while in the lower molecular weight region (<50 kb) they looked significantly different. This possibly leads to a conclusion that the enzymes digest the DNA in one portion of the plug completely, leaving the other, more inaccessible part, intact. It is certainly better to leave the plugs with the enzyme (in buffer devoid of $MgCl_2$) overnight on ice to allow better diffusion of the enzyme. If even that does not work, it is worth attempting melting the plugs before adding the enzyme. This process (even though it degrades the DNA to a certain extent) allows better digestion of the DNA when resilient contaminants are present. Ideally, although, one will not put the DNA in a molten state, even for a short while.

PFGE is expected to separate out the HMW DNA from smaller sized undesirable DNA. However, it is intriguing to notice that a lot of small molecules are still present in the HMW DNA zone. The difficulty is that one can detect this only after going through a cycle

of size-selection, ligation and transformation. One would expect to find a few smaller sized DNAs in the HMW DNA zone, but a lot of the transformations showed that most of the DNA were small sized. Different PFGE conditions, using different equipment, allowing varying degrees of separation were tried out to get rid of the smaller DNA, but the problem still persisted. It is not clear whether the quality of the starting DNA has any bearing on this problem, even though logically one would assume that it does not. However, it is logical to assume that very high concentrations of DNA tend to trap smaller DNA molecules in the compression zone. It is also known that high concentrations of EDTA in the digested DNA plug traps small DNA molecules in the compression zone (Funke et al., 1994). Hence care was taken to keep the EDTA concentration low (~50 mM). Additionally, the gel containing the partially digested DNA was equilibrated in 1X TAE before starting the electrophoresis. Despite all these precautions, the problem of small molecule trapping persisted. This problem was the most difficult and insurmountable one in the construction of a HMW DNA library.

A few technical problems were also experienced with regard to the vector and the host competent cells. Not all dephosphorylating enzymes dephosphorylated the BAC vector efficiently. Shrimp Alkaline Phosphatase (SAP), which was used for YAC vector dephosphorylation did not give good results with the BAC vector. This might be due to buffer incompatibility or slight exonuclease activity. Many other labs have also had problems with this enzyme (B. Birren, MIT; personal communication). It was HK phosphatase that gave the best results, with the added advantage that the vector need not be extracted organically for further use in cloning. It was also found that the dephosphorylated BAC vector was not very efficient in cloning after extended periods of storage (>2 months), even if stored at -20°C. It is possible that the cloning site was destroyed. In YAC vector dephosphorylation, it was found that more than prescribed amounts of BSA was needed to bring about complete dephosphorylation.

Minor technical problems cost a lot of time since the reasons were not obvious enough to be noted at the first sight. For example, it took quite a while to figure out that self-made bacterial electro-competent cells were not efficient in taking up large sized ligated

DNA molecules. When the cells transformed well with control BACs as large as 170 kb, it was concluded that the cells were good enough for uptake of large sized molecules. It was not until a control experiment, in which self-ligated vector was transformed into the cells, that it was clear that the cells were not efficient in taking up circular ligated DNA molecules. The BAC controls used earlier were supercoiled DNA molecules. It was only then that it was realized that the cells were capable of taking in large sized supercoiled molecules, but not so in taking up even small sized circular molecules. The technical skills involved in making competent cells were probably not good enough. Since a library would need thousands of clones, it was decided not to proceed with standardizing procedures for making electro-competent cells and instead to purchase commercially available ones. The commercially available cells (Gibco-BRL, USA) worked much better for DNA of all different sizes and structures.

The ligation and transformation procedures were standardized without much difficulty. Experiments showed that a larger excess of vector can give better transformations. It was also noticed that inactivating the ligase enzyme prior to transformation enhanced transformation efficiencies. Since the attempt was to construct large sized DNA clones and not merely to get good transformations, utmost care had to be taken in all the individual transformation steps. For example, the electroporated *E. coli* cells were to be shaken at 37°C for about 45 minutes before plating them out, to allow the expression of the chloramphenicol-resistance gene. It was noted that when the temperature was slightly below 37°C it caused a significant reduction (about two fold) in colony numbers. In YAC transformation, it was found to be better to choose a range of spheroplasting close to the estimated optimum, than to proceed with only one set of spheroplasted cells. At times, extents of spheroplasting, slightly greater or smaller than the estimated optimum, gave better transformations. This might stem from variations in spectrophotometric readings that result from handling errors.

Even though isolating BAC DNA for characterization was much easier than doing the same with YACs, there were some technical difficulties experienced there also. When the temperature in the incubator was slightly higher than normal (by 2-3°C), it was noticed

that many of the BACs tend to lose their plasmids. This might be because the small stress induced by the change in temperature forces the bacteria to abandon the replication of the large BAC plasmid, after the antibiotic is used up. The other major problem experienced with BAC DNA was its relatively short storage life. It was not possible to store intact BAC DNA for more than a couple of weeks. BAC DNA comes with a lot of contamination, not only of chromosomal DNA, but also of nucleases. These nucleases tend to degrade the large sized BAC DNA relatively fast. Even treatment with phenol-chloroform did not significantly reduce the degradation. Hence it was better to extract DNA fresh, especially for digestion and PCR.

The *Lotus* BAC library constructed as part of this work has an average insert size of 65 kb. This insert size does not compare favorably with other plant BAC libraries available, most of which have average insert sizes higher than 100 kb (Table 4-1). But this BAC library can still be used as a HMW DNA library in genome analysis. About 15-20% of clones in this library are larger than 100 kb. While screening the library for candidate clones, it is quite possible that large (>100 kb) clone will come out positive. The smaller clones can be used to fill gaps between the large clones identified. The rate at which progress is being made in research involving this model legume certainly necessitates the availability of genomic libraries. This library will be a good starting resource for genome analysis since there are no other HMW DNA libraries available for *Lotus japonicus*.

Table 4-1
BAC libraries in plant genome analysis

Plant	Institute	Av. insert size (kb)
Rice (IRBB21)	UC Davis	125
Rice (Lemont)	Texas A&M	150
Rice (Teqing)	Texas A&M	130
Rice	Japan	175
<i>Arabidopsis</i>	UC Davis	65
Soybean	Iowa State	150
Sorghum	Texas A&M	135
Wheat	Texas A&M	In progress
Common bean	Mexico	100
<i>Medicago</i>	Texas A&M	100
Apple	Cornell	110
Citrus	U Florida	125
Grapes	UC Davis	In progress
Lettuce	UC Davis	In progress
Cotton	Texas A&M	100

BACs vs YACs - a comparison based on technical feasibility

Yeast and bacterial artificial chromosomes have gained prominence as the vectors of choice for HMW DNA cloning. A recent trend in the field of plant genome analysis shows that BACs are more popular with researchers than YACs. The various advantages and disadvantages with both systems have been discussed in the introduction. By working with both the systems, it was possible to reasonably compare the technical aspects of BACs and YACs.

The initial steps leading to the isolation of digested HMW DNA and ligation to the respective is more or less the same for both the systems. However, the later stages are vastly different in ease, time and effort. Once transformed, it is possible to see recombinant colonies of BACs within a matter of 24 hours. In case of YACs it could take nearly 3 days for the colonies to appear, before they can be transferred onto double selection plates to look for recombinant red colonies. The latter process takes a couple of days more. In total, it could take 5-6 days after transformation to view the recombinant clones.

The next comparison can be made about the easiness in isolating pure BAC or YAC DNA from the clones. BAC DNA can be isolated from an overnight culture of the clone by following a simple miniprep procedure. Since it is a single copy plasmid, the yield is relatively low, but BAC DNA from a 1.5 mL overnight culture should be more than enough to do a variety of tests including insert size check and PCR. Moreover, improvised procedures and kits are available for extracting larger quantities of purer BAC DNA. In case of YACs the isolation of pure recombinant DNA is much more time-consuming and difficult. Since the yeast has 16 native chromosomes and the YAC is just an additional chromosome, it is impossible to extract pure YAC DNA straight from the yeast cell. The entire yeast DNA has to be extracted and put into plugs to prevent shearing. The process of lysing away the other cell components takes 2 days. After PFGE separation the YAC DNA is isolated by cutting out from the gel. This comes with a lot of contaminants since the YAC might run very close to native yeast chromosomes.

Checking the insert size is pretty simple in case of BACs. One only needs to digest the isolated BAC DNA with *NotI* to release the insert from the vector. The insert size can be directly observed after running a PFGE. Hybridization with genomic DNA need to be done only for confirmation. As opposed to this YAC insert size cannot be clearly viewed on a PFGE gel since the YAC runs along with 16 other chromosomes. Only hybridization can reveal the exact size of the YAC.

Another avenue for comparison is the ease with which the required number of clones can be obtained to complete a library. In the case of BACs one good ligation mixture

of 100 μ L is enough to complete the entire library of an organism of 1000 Mb genome size. From a single transformation (at least 350 clones) it is possible to ascertain whether the ligation is of good enough quality to construct the library. Once found to be ideal all that needs to be done is repeated electroporation. It is possible to do four electroporations at a time. In less than a month it should be possible to complete the library. In the case of YACs it is a different story. Even if a single transformation is found to be good, a complete library is still a long way to go. An entire ligation is used for a single transformation. This means that to do another transformation, the entire process starting from the partial digestion HMW DNA has to be repeated many times over. Given the stability of the source HMW DNA and the unpredictability of partial digestion, this process could take well over a couple of years.

Viewing purely from the technical side and easiness of operations the balance is tilted heavily in favor of BACs. However, as stated in the introduction, for sheer cloning capacity YACs are better cloning vehicles. Moreover, some types of eukaryotic DNA cannot be cloned in prokaryotic bacteria, but can be cloned in eukaryotic yeast. A look at the number of plant genes of commercial value that have been cloned using YACs bears testimony to their utility (see introduction). BACs are a pretty recent choice for cloning and hence it could take some time before we can see a number genes cloned using this system.

The aspects mentioned above are in addition to the scientific facts given in the introduction comparing YACs and BACs. Given enough time and resources, it is still worth to go in for YAC libraries. But since these resources are pretty limited in today's tight situation it is likely that many plant genomics researchers will go for BAC libraries.

Identification of candidate BAC clones

As part of this work, attempts were made to isolate BAC clones harboring the pUTG-132a molecular marker mapped close to the supernodulating *nts-1* locus in soybean. The objective was to obtain a large sized clone containing pUTG-132a, which could further be used for building contigs towards the *nts-1* locus. Results obtained are preliminary and needs further confirmation. The work was done on two BAC pools previously screened by

PCR and identified to contain pUTG-132a. At first the decision had to be made whether to do further PCR screening or to do colony hybridization. With the exciting possibility of zeroing in on one clone from each pool, the decision was made to continue PCR screening.

In reality, the process was not as well defined as expected and a lot of technical difficulties were encountered. Since the pools contained nearly 350 clones each, it was decided to be sufficient to pick up around 750 clones. However, it is quite possible that there are certain regions of the genome that are drastically underrepresented in a BAC library. Personal communications with researchers at the source of the library gave information that there were cases where nearly ten-fold number of clones from a pool had to be screened to locate specific sequences. Hence picking up twice the number of clones might not have been enough and hence that decision was questionable. Anyway, it was decided to proceed with the clones picked into pools hoping that the target sequence is properly represented. Another question arises in the method of growing the pools. Pools were made by inoculating each clone into a common dish containing LB media and grown overnight. It is possible that there will be a size-based bias in the growth of the BACs, with the smaller sized BACs growing more efficiently than the bigger sized ones. It was decided to deal with this question only if the isolated candidates turned out to be much smaller than the average size of the library (105 kb).

As in case of any PCR experiment, standardizing conditions for the amplification of the specific sequence was an essential step. The original pUTG-132a sequence was isolated from the soybean cultivar "Bragg" (Landau-Ellis et al, 1991), and hence the PCR conditions were standardized for the same. This also involved preparation of a special PCR buffer (Kolchinsky et al, 1997). However, the library from which the pools were isolated was made in the cultivar "Williams". These cultivar differences could adversely affect PCR amplification. Hence, the same PCR conditions that were followed for the isolation of the pools was followed for amplifying control genomic and plasmid DNA. It was also decided to work with the primer pair that gave the smallest product (614 bp) for the ease of amplification. Even though these conditions worked well in amplifying genomic and plasmid DNA, it was not possible to amplify the pools. Hence, a touch down PCR

approach was adopted that was more likely to give amplification products. A whole range of products including the right sized ones were observed. The confusion remained even after hybridization to the pUTG-132a probe since a lot of non-specific signals showed up. Signals were visible even at spots where nothing was seen on the gel. It was later decided to pick up and align those pools that showed the highest intensity of the signal.

The difficulty in amplifying the BACs leads to some speculation about the nature of pUTG-132a and also the sequences surrounding it. It is possible to speculate that there is a region of the genome quite close to pUTG-132a that is in some way homologous to at least one of the primers or both, and gets amplified non-specifically. In one of the control experiments it was seen that the addition of an equal amount of BAC DNA suppressed the amplification of genomic DNA. This could also be because of possible contaminants in the isolated plasmid DNA that could be inhibiting amplification. In conclusion, it can be asked whether the BAC DNA is resilient to amplification or whether there is a sequence structure in the vicinity of pUTG-132a that adversely affects amplification.

After careful consideration, ten candidate BACs were selected for further analysis. The insert sizes ranged from 50-165 kb. Only two clones were larger than 100 kb. However, since two of them were really large and the smallest was still around 50 kb, it was decided not to pursue the original speculation that smaller clones grow preferentially in pool cultures. Since pUTG-132a is unique in the genome, it was hard to accept the fact that ten different clones harbor it. Here again the question of cultivar differences surfaced.

When probed with pUTG-132a it was possible to omit four clones from further consideration. Hybridization results led to another dilemma. The signals were very weak and needed a couple of weeks exposure. At this time the question came about whether it is advisable to hybridize whole BACs with the target sequence. It might be difficult for the probe to hybridize efficiently to a homologous sequence amongst DNA fragments as large as 165 kb. Hence it was decided to digest the BAC with *Pst*I that would cut out the 1.7 kb pUTG-132a fragment. A whole size-range of fragments resulted from the digestion, but it was not possible to locate the 1.7 kb fragment. *Pst*I is not known for efficient restriction

digestion and hence it might be that the fragment is somewhere among the larger fragments. But hybridization gave the most surprising and confusing results. A whole range of fragments in almost every clone was found to hybridize strongly to the pUTG-132a probe. This was really surprising considering the fact that the whole BAC hybridization with the same probe gave such weak signals. Even stranger, how can an unique sequence hybridize to that many fragments in a single clone? It was not possible to derive any valid conclusion from this experiment and hence it was not followed up.

A totally contrasting result was obtained when these individual BACs were amplified using pUTG-132a primers. None of them amplified the desired size product. The question at this stage was quite intriguing - how can the BAC DNA molecules that show such strong signals on hybridization with the probe, not amplify the same sequence? It is possible to explain this by the above stated speculation about the nature of the sequences surrounding pUTG-132a. Genomic DNA amplifies because the sequence is very much diluted in there when compared to BACs wherein the template mixture is many fold enriched for the sequence. This experiment also gave no conclusive result.

Taking cultivar differences into consideration, nested PCR was tried to amplify pUTG-132a from the BACs. This process consisted of two rounds of amplification, first one using one pair of primers that would enrich the mixture for the entire pUTG-132a sequence (1.7 kb) and a second round using a set of primers that would amplify a smaller portion of the sequence (1.1 kb). This process was expected to offset any cultivar related sequence differences. To narrow down the search, only those BACs that showed signals on whole BAC hybridization were used. The results were more conclusive but definitely not absolute. The results became more apparent only after hybridization. Clones 1,3, 5 and 10 showed signals along with the control genomic DNA. Since clones 1 and 10 were both 50 kb and showed similarity on digestion it was decided to omit clone 10 from further investigation. Thus three candidate clones of sizes 50, 165 and 70 kb were selected. The hope was that the two smaller clones would in some way be inside the big clone and thus the 165 kb BAC could be used for further experiments.

Isolation of end clones was the next logical step, even though it can always be argued that there was a need to do more affirmative tests on the isolated BACs. The library was constructed using the *Bam*HI cloning site of the vector. This facilitated the isolation of both left and right end clones by plasmid end rescue. The *Hind*III cloning site would have forced a more difficult IPCR process for isolation of the left end clone. Any end clone smaller than around 25 kb in size could be used, but an end clone around 10-15 kb would be ideal and that was what was obtained. As part of this work only the right end clones was characterized further. Experiments are underway in our laboratory (by Dr. Artem Men) to characterize the left end clones.

Sequence information of the right end clone of BAC1 (1RE) revealed 85% homology to a maize retrotransposon element *Opie-2* (SanMiguel et al., 1996). Retrotransposons are found to occur in many parts of the genome (Grandbastien, 1992). Hence it was concluded that a sequence bearing close homology to a retrotransposon is definitely not good for mapping an end clone. It was decided to omit 1RE from further investigation.

The end clone that was characterized most was 5RE. After sequence information was available, primers were designed for this end clone. Some encouraging results have come out of PCR amplifications using these primers. The fact that it was possible to amplify the 450 bp fragment from BAC3 argues that the smaller BAC5 is located within the 165 kb BAC3. In yet another experiment, it was possible to amplify the same PCR product from 3RE. This means that BAC5 and BAC3 could share the same right end, which in turn means that BAC3 could be located at one end of BAC3. This lightens up the possibility of using the large BAC as the sole candidate for pUTG-132a. To identify the orientation of the BAC (to determine which end of the BAC is closer to *nts*) it is necessary to map both the left and right end clones of the candidates. Since BAC5 is only 70 kb in size it could be difficult to map the ends. On the other hand a 165 kb BAC will make it easier to map the end clones. Given the fact that *nts-1* could be anywhere between 120-600 kb away from pUTG-132a, getting a 165 kb BAC harboring pUTG-132a could help a lot more in even isolating the gene.

To map any sequence it is necessary to find molecular level polymorphism between two parents. Unfortunately it was not possible to find a polymorphism between *G. max* and *G. soja* for 5RE. Both parents amplified the same product and sequencing showed that they are identical. Thus it became obvious that it is not possible to map 5RE using our F2 segregating population. However, it was possible to find a putative polymorphism between ecotypes Gifu and Funakura of the model legume *Lotus japonicus*. The results are very preliminary and need to be confirmed. With molecular synteny between soybean and *Lotus japonicus* being a distinct possibility, it could be possible to use the model legume system to isolate a gene homologous to *nts-1* from *Lotus*, provided an equivalent supernodulating mutant is available.

Future perspectives

The ultimate aim of this work was to standardize conditions to construct complete HMW DNA libraries of the legumes soybean and *Lotus japonicus*. Because of technical difficulties only a part of this aim could be achieved. The first question to be resolved is whether to attempt to construct a YAC library or a BAC library. Most of the findings of this work suggest the construction of a BAC library. Most of the methodologies have been standardized and the only major problem that was left untackled was the presence of undesirably small molecules in the DNA that was being cloned. If this difficulty can be overcome, constructing a BAC library should not be very difficult. The key to this lies in preparing really high quality HMW DNA. High quality in this case would mean good concentrations of readily digestible intact HMW DNA. It would be wise to allot an extended period for isolating good quality HMW DNA and proceed further only after getting several batches of good quality source DNA. Once this has been achieved the problem of small molecule trapping can be overcome by double size selection or varying pulsed field gel running conditions. A running technical collaboration, with a laboratory that has already made HMW DNA libraries would be immensely helpful.

The next aspect to be discussed is about fixing priorities. Soybean already has three BAC libraries available. While it would be worthwhile to have a BAC library of the cultivar Bragg, it should be possible to get most of the needed information and resources

from the existing libraries. The first priority should be given for the construction of a large insert size (>100 kb) BAC library of *Lotus japonicus*. With this plant emerging fast as the model legume of choice, a BAC library of the same is imperative. Since the technology has been standardized in this laboratory and no other lab is at present capable of attempting such a venture, this is quite a reasonable goal to aim for. It is also reasonable to foresee a future for *Lotus* in the same lines as *Arabidopsis* and hence the demand for such resources would be very high. Also, once the technology is established it will be easy to transfer the same into constructing a soybean BAC library.

As regards cloning the *nts* gene, which is one of the ultimate goals of this laboratory, positional cloning is still the best way. Agreed, that the results obtained were not definitely conclusive, but an avenue is clearly open for isolating this gene. If the results cannot be confirmed, it is still possible to re-screen the pools by colony hybridization to pick up candidate colonies directly. It is also not unreasonable to go in for a PCR re-screening of the entire library to obtain new, more definite pools.

One of the ultimate tests for an isolated gene is to functionally complement the mutant phenotype with the gene. Here, the drawback with soybean is that transformation is very difficult. It could be quite frustrating to isolate a gene and not be able to confirm it by transformation. It is here that the promise of *Lotus japonicus* holds good. The availability of a putative polymorphism between Gifu and Funakura is an added advantage, provided the BACs can be confirmed. As we have seen in the introduction, this plant has a lot of desirable characters including high frequency of transformation and regeneration and high seed set. If it is possible to isolate a *Lotus* mutant with a similar supernodulating phenotype, there exists an exciting possibility of using this model system for getting the *nts* gene. The use of other model organisms bear testimony to this fact. The logic is that if a nematode worm can be used to study human aging genes, it should be possible to isolate a soybean nodulation gene from another legume.

Positional cloning is getting more powerful as a genome analysis tool. The difficulty in traveling extended stretches of DNA to isolate genes of interest will prompt

scientists to look for high-volume marker technologies to enrich the areas surrounding the genes of interest. But there could be a lot of genes where these technologies cannot be applied. For isolating such genes it is imperative that positional cloning strategies be used. The future will see the handling and manipulation of even larger pieces of DNA, not only for genome analysis, but also for such applications as gene therapy, genetic engineering and bioinformatics. Hence it is important that scientists be equipped technically to work with them. If so, we can all look forward to an exciting future for HMW DNA technology and its applications in positional cloning.

REFERENCES

Adam, G., Mullen, J.A., and Kindle, K.L. (1997) Retrofitting YACs for direct transfer into plant cells. *Plant J.* **11**, 1349-1358.

Albertsen, H.M., Abderrahim, H., Cann, H.M., Dausset, J., Le, P.D., and Cohen, D. (1990) Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl. Acad. Sci. USA* **87**, 4256-4260.

Albertsen, H.M., Abderrahim, H., Cann, H.M., Dausset, J., and Le Paslier, D. (1992). In *Plant Genomes: Methods for Genetic and Physical Mapping*, J.S. Beckman and T.C. Osborn, eds (Netherlands: Kluwer Academic Publishers), pp 141-165.

Anand, R., Villasante, A., and Tyler-Smith, C. (1989) Construction of yeast artificial chromosome libraries with large inserts using fractionation by pulsed-field gel electrophoresis. *Nucleic Acids Res.* **17**, 3425-3432.

Anand, R., Ogilvie, D.J., Butler, R., Riley, J.H., Finniear, R.S., Powell, S.J., Smith, J.C., and Markham, A.F. (1991) A yeast artificial chromosome contig encompassing the cystic fibrosis locus. *Genomics* **9**, 124-130.

Anderson, C. (1993) Genome shortcut leads to problems. *Science* **259**, 1684-1687.

Antonio, B.A., Emoto, M., Wu, J., Ashikawa, I., Umehara, Y., Kurata, N., and Sasaki, T. (1996) Physical mapping of rice chromosomes 8 and 9 with YAC clones. *DNA Res.* **3**, 393-400.

Appleby, C.A. (1984) Leghemoglobin and rhizobial respiration. *Annu. Rev. Plant Physiol.* **35**, 443-478.

Arondel, V., Lemieux, B., Hwang, I., Gibson, S., Goodman, H., and Somerville, C.R. (1992) Map-based cloning of a gene controlling omega-3 fatty acid desaturation in *Arabidopsis*. *Science* **258**, 1353-1355.

Asakawa, S., Abe, I., Kudoh, Y., Kishi, N., Wang, Y., Kubota, R., Kudoh, J., Kawasaki, K., Minoshima, S., and Shimizu, N. (1997) Human BAC library : construction and rapid screening. *Gene* **191**, 69-79.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., Struhl, K. (1995) *Current Protocols in Molecular Biology*, Publ. John Wiley & sons.

Baker, A., and Cotton, M. (1997) Delivery of bacterial artificial chromosomes into mammalian cells with psoralen-inactivated adenovirus carrier. *Nucleic Acids Res.* **25**, 1950-1956.

Barker, D.G., Bianchi, S., Blondon, et al. (1990) *Medicago truncatula*, a model plant for studying the molecular genetics of the *Rhizobium*-legume symbiosis. *Plant Mol. Biol. Rep.* **8**, 40-49.

Bent, A.F., Kunkel, B.N., Dahlbeck, D., Brown, K.L., Schmidt, R., Giraudat, J., Leung, J. and Staskawicz, B.J. (1994) RPS2 of *Arabidopsis thaliana*: a leucine-rich repeat class of plant disease resistance genes. *Science* **265**, 1856-1860.

Birren, B., and Lai, E. (1993) In Pulsed Field Gel Electrophoresis : A Practical Guide, (Academic Press Inc., San Diego, CA).

Bonnema, G., Hontelez, J., Verkerk, R., Zhang, Y.Q., vanDaelen, R., van Kammen, A., and Zabel, P. (1996) An improved method of partially digesting plant

megabase DNA suitable for YAC cloning : application to the construction of a 5.5 genome equivalent YAC library of tomato. *Plant J* 9, 125-133.

Bouffard, G.G., Idol, J.R., Braden, V.V., Iyer, L.M., Cunningham, A.F., Weintraub, L.A., Touchman, J.W., Mohr-Tidwell, R.M., Peluso, D.C., Fulton, R.S., Ueltzen, M.S., Weissenbach, J., Magness, C.L., and Green, E.D. ((1997) A physical map of chromosome 7 : an integrated YAC contig map with average STS spacing of 79 kb. *Genome Research* 7, 673-692.

Boysen, C., Simon, M.I., and Hood, L. (1997) Analysis of the 1.1-Mb human alpha/delta T-cell receptor locus with bacterial artificial chromosome clones. *Genome Res.* 7, 330-338.

Burke, D.T., Carle, G.F., and Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236, 806-812.

Burke, D.T. (1991) The role of yeast artificial chromosome clones in generating genome maps. *Curr Opin Genet Dev* 1, 69-74.

Caetano-Anolles, G., and Gresshoff, P.M. (1991) Plant genetic control of nodulation. *Annu. Rev. Microbiol.* 45, 345-382.

Cai, L., Taylor, J.F., Wing, R.A., Gallagher, D.S., Woo, S-S., and Davis, S.K. (1995) Construction and Characterization of a Bovine Bacterial Artificial Chromosome Library. *Genomics* 29, 413-425.

Cai, D., Kleine, M., Kifle, S., Harloff H.J., Sandal N.N., Marcker, K.A., Klein-Lankhorst, R.M., Salentijn, E.M.J., Lange, W., Stiekema, W.J., Wyss, U., Grundler F.M.W., and Jung, C. (1997) Positional cloning of a gene for nematode resistance in sugar beet. *Science* 275, 832-834.

Cai, L., Schalkwyk, L.C., Schoeberlein-Stehli, A., Zee, R.Y., Smith, A., Haaf, T., Georges, M., Lehrach, H., and Lindpaintner, K. (1997) Construction and characterization of a 10-genome equivalent yeast artificial chromosome library for the laboratory rat, *Rattus norvegicus*. *Genomics* 39, 385-392.

Camargo, A.A., Fischer, K., and Lanzer, M. (1997) Construction and rapid screening of a representative yeast artificial chromosome library from *Plasmodium falciparum* strain Dd2. *Parasitol Res* 83, 87-89.

Carroll, B.J., McNeil, D.L., and Gresshoff, P.M. (1985a) Isolation and properties of soybean (*Glycine max*) mutants that nodulate in the presence of high nitrate concentrations. *Proc. Natl. Acad. Sci.* 82, 4162-4166.

Carroll, B.J., McNeil, D.L., and Gresshoff, P.M. (1985b) A supernodulation and nitrate tolerant symbiotic (*nts*) soybean mutant. *Plant Physiol.* 78, 34-40.

Chang, C., Bowman, J.L., DeJohn, A.W., Lander, E.S., and Meyerowitz, E.M. (1988) Restriction Fragment Length Polymorphism map for *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* 85, 6856-6860.

Chang, C., Kwok, S-F., Bleecker, A.B., and Meyerowitz, E.M. (1993) *Arabidopsis* ethylene-response gene *ETR1*: similarity of product to two-component regulators. *Science* 262, 539-544.

Chen, M., SanMiguel, P., de Oliveira, A.C., Woo, S-S., Zhang, H., Wing, R.A., Bennetzen, J.L. (1997) Microcolinearity in sh2-homologous regions of the maize, rice and sorghum genomes. *Proc. Natl. Acad. Sci. USA* 94, 3431-3435.

Chu, G., Vollrath, D., and Davis, R.W. (1986) Separation of large DNA molecules by contour clamped homogeneous electric fields. *Science* 234, 1582-1585.

Clark, R.F., Cruts, M., Korenblat, K.M., He, C., Talbot, C., Van Broeckhaven, C., and Goate, A.M. (1995) A yeast artificial chromosome contig from human chromosome 14q24 spanning the Alzheimer disease locus AD3. *Hum Mol Genet* 4, 1347-1354.

Connelly, C., McCormick, M.K., Shero, J., and Heiter, P. (1991) Polyamines eliminate an extreme size bias against transformation of large yeast artificial chromosome DNA. *Genomics* 10, 10-16.

Cooper, J.B., and Long, S. (1994) Morphogenetic rescue of *Rhizobium meliloti* nodulation mutants by trans-zeatin secretion. *Plant Cell* 6, 215-225.

Coulson, A., Waterston, R., Kiff, J., Sulston, J., and Kohara, Y. (1988) Genome linking with yeast artificial chromosomes. *Nature* 335, 184-186.

Creusot, F., Fouilloux, E., Dron, M., Lafleuriel, J., Picard, G., Billaut, A., Le Paslier, D., Cohen, D., Chaboute, M-D., Durr, A., Fleck, J., Gigot, C., Camilleri, C., Bellini, C., Caboche, M., and Bouchez (1995) The CIC YAC library : a large insert YAC library for genome mapping in *Arabidopsis thaliana*. *Plant Journal* 8(5), 763-770.

Dausset, J., Ougen, P., Abderrahim, H., Billaut, A., Sambucy, J.L., Cohen, D., and Le Paslier, D. (1992) The CEPH YAC library. *Behring Inst Mitt* 91, 13-20.

Deckert, J., Taranenko, N., and Gresshoff, P.M. (1994) Cell cycle genes and their plant homologues. In : *Plant Genome Analysis*, Gresshoff, P.M, ED. CRC Press, Boca Raton, FL, 169-194.

- Del-Favero, J., Vauterin, M., Weyens, G., Edwards, K.E. and Jacobs, M.** (1994) Construction and characterization of a yeast artificial chromosome library containing five haploid sugarbeet (*Beta vulgaris* L.) genome equivalents. *Theoretical and Applied Genetics*, **88** .449-453.
- Delsney, M., Raynal, M., Laudie, M., Varoquaux, F., Comella, P., Wu, H.J., Cooke, R., and Grellet, F.** (1996) Sequencing and mapping the *Arabidopsis* genome : a weed model for real crops. *Symp Soc Exp Biol* **50**, 5-9.
- Denton, P.H., Cullen, J.B., Loeb, D., Lucas, A., and Nunes, K.** (1994) Partitioned PFGE-PCR : a new method for pulsed-field mapping of SRS and microsatellites. *Nucleic Acids Res.* **22**, 1776-1777.
- Dietrich, W.F., Copeland, N.G., Gilbert, D.J., Miller, J.C., Jenkins, N.A., and Lander, E.S.** (1995) Mapping the mouse genome : current and future prospects. *Proc Natl Acad Sci USA* **92**, 10849-10853.
- Dixon, M.S., Jones, D.A., Keddie, J.S., Thomas, C.M., Harrison, K., and Jones, J.D.** (1996) The tomato Cf-2 disease resistance locus comprises two functional genes encoding leucine-rich repeat proteins. *Cell* **84**(3), 451-459.
- Dower, W.J., Miller, J.F., and Rgasdale, C.W.** (1988) High efficiency transformation of *E. coli* by high voltage electroporation. *Nucleic Acids Res.* **16**, 6127.
- Dunford, R., Vilageliu, L., and Moore, G.** (1993) Stabilisation of a yeast artificial chromosome containing plant DNA using a recombination-defecient host. *Plant Mol Biol* **21**, 1187-1189.
- Edwards, K.J., Thompson, H., Edwards, D., De-Saizieu, A., Sparks, C., Thompson, J.A., Greenland, A.J., Evers, M. and Schuch, W.** (1992)

Constuction and characterization of a yeast artificial chromosome library containing three haploid maize genome equivalents. *Plant. Mol. Bio.* **19**, 299-308.

Ehrhardt, D.W., Atkinson, E.M., and Long, S.R. (1992) Depolarization of alfalfa root hair membrane potential by *Rhizobium meliloti* Nod factors. *Science* **256**, 998-1000.

Eyers, M., Edwards, K., and Schuch, W. (1992) Construction and characterization of a yeast artificial chromosome library containing two haploid *Beta vulgaris* L. genome equivalents. *Gene* **121**, 195-201.

Fuchs, J., Kloos, D.U., Ganai, M.W., and Schubert, I. (1996) *In situ* localization of yeast artificial chromosomes on tomato and potato metaphase chromosomes. *Chromosome Res.* **4**, 277-281.

Funke, R.P., Kolchinsky, A. and Gresshoff, P.M. (1993) Physical mapping in a region of the soybean genome containing duplicated sequences. *Plant Mol. Biol.* **22**, 437-446.

Funke, R.P., Kolchinsky, A. and Gresshoff, P.M. (1994) High EDTA concentrations cause entrapment of small DNA molecules in the compression zone of pulsed field gels, resulting in smaller than expected insert sizes in YACs prepared from size selected DNA. *Nucleic Acids Research* **22**, 2708-2709.

Garza, D., Ajioka, J.W., Burke, D.T., and Hartl, D.L. (1989) Mapping the *Drosophila* genome with yeast artificial chromosomes. *Science* **246**, 641-646.

George, S., Clark, A.J. and Archibald, A.L. (1997) Physical mapping of the murine casein locus reveals the gene order as alpha-beta-gamma-epsilon-kappa. *DNA Cell Biol.* **16**, 477-484.

Gibson, F., Lehrach, H., Buckler, A.J., Brown, S.D.M., and North, M.A. (1994) Isolation of Conserved Sequences from Yeast Artificial Chromosomes by Exon Amplification. *BioTechniques* 16, 453-458.

Gianfrancesco, F., Esposito, T., Ruini, L., Houlgatte, R., Nagaraja, R., D'Esposito, M., Rocchi, M., Auffray, C., Schlessinger, D., D'Urso, M., and Forabosco, A. (1997) Mapping of 59 EST gene markers in 31 intervals spanning the human X chromosome. *Gene* 187, 179-184.

Grandbastien M.A. (1992) Retroelements in higher plants. *Trends Genet* 8, 103-108.

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

Green, E.D. and Green, P. (1991) Sequence-tagged site (STS) content mapping of human chromosomes: Theoretical considerations and early experiences. *PCR Methods Applic.* 1, 77-90.

Green, E.D., Reithman, H.C., Dutchik, J.E. and Olson, M.V. (1991) Detection and characterization of chimaeric yeast artificial chromosome clones. *Genomics* 11, 658-669.

Green, E.D., Braden, V.V., Fulton, R.S., Lim, R., Ueltzen, M.S., Peluso, D.C., Mohr-Tidwell, R.M., Idol, J.R., Smith, L.M., and Chumakov I. (1995) A human chromosome 7 yeast artificial chromosome (YAC) resource : construction, characterization and screening. *Genomics* 25, 170-183.

Gresshoff, P.M. (1993) Molecular genetic analysis of nodulation genes in soybean. *Plant Breeding Rev.* 11, 275-318.

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

Gurley, W.B., Hepburn, A.G., and Key, J.L. (1979) Sequence organization of the soybean genome. *Biochim.Biophys.Acta* 561, 167-183.

Guzman, P. and Ecker, J.H. (1988) Development of large DNA methods for plants: molecular cloning of large segments of *Arabidopsis* and carrot DNA into yeast. *Nucleic Acids Research* 16, 11091-11105.

Haldi, M., Perrot, V., Saumier, M., Desai, T., Cohen, D., Cherif, D., Ward, D., and Lander, E.S. (1994) Large human YACs constructed in a *rad52* strain show a reduced rate of chimerism. *Genomics* 24, 478-484.

Hamilton, C.M., Frary, A., Lewis, C., and Tanksley, S.D. (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes *Proc. Natl. Acad. Sci. USA* 93, 9975-9979.

Handberg, K. and Stougaard, J. (1992) *Lotus japonicus*, an autogamous, diploid legume species for classical and molecular genetics. *Plant Journal* 2, 487-496.

Handberg, K., Stiller, J., Thykjaer, T. and Stougaard, J. (1994) Cell Biology: A Laboratory Handbook, Academic Press Inc. pp 119-127.

Hanson, R.E., Zwick, M.S., Choi, S., Islam-Faridi, M.N., McKnight, T.D., Wing, R.A., Price, H.J., Stelly, D.M. (1995) Fluorescent in situ hybridization of a bacterial artificial chromosome. *Genome* 38, 646-651.

Harris, A. and Argent, B. E. (1993) The cystic fibrosis gene and its product CFTR. *Semin. Cell Biol.* 4: 37-44.

Havukkala, I.J. (1996) Cereal genome analysis using rice as a model. *Curr. Opin. Genet. Dev.* 6(6), 711-714.

Heale, S.M., Stateva, L.I., and Oliver, S.G. (1994) Introduction of YACs into intact yeast cells by a procedure which shows low levels of recombinagenicity and co-transformation. *Nucleic Acids Res.* 22, 5011-5015.

Hieter, P. and Boguski, M. (1997) Functional Genomics: It's All How You Read It. *Science* 278, 601-602.

Hieter, P., Connelly, C., Shero, J., McCormick, M.K., Antonarakis, S., Pavan, W. and Reeves, R. (1990) Yeast artificial chromosomes: promises kept and pending. In *Genome Analysis Vol I: Genetic and Physical Mapping* (Cold Spring Laboratory Press), pp 83-119.

Hirsch, A.M. (1992) Developmental biology of legume nodulation. *New Phytol.* 122, 211-237.

Hubert, R.S., Mitchell, S., Chen, X.N., Ekmekji, K., Gadomski, C., Sun, Z., Noya, D., Kim, U.J., Chen, C., Shizuya, H., Simon, M., de Jong, P.J., and Korenberg, J.R. (1997) BAC and PAC contigs covering 3.5 Mb of the Down syndrome congenital heart disease region between D21S55 and MX1 on chromosome 21. *Genomics* 41, 218-226.

Hwang, I., Kohchi, T., Hauge, B.M., Goodman, H.M., Schmidt, R., Cnops, G., Dean, C., Gibson, S., Iba, K., Lemieux, B. et al. (1991) Identification and map position of YAC clones comprising one-third of the *Arabidopsis* genome. *Plant J.* 1, 367-374.

Hyser, C. L. (1989) Unraveling the mysteries of Duchenne and Becker muscular dystrophy. *Mol. Chem. Neuropathol.* **10**, 15-20.

Infante, I., Lo, S., Hall, J.L. (1995) A Chlamydomonas Genomic Library in Yeast Artificial Chromosomes. *Genes* **141**, 87-93.

Jacob, A.N., Kandpal, G., and Kandpal, R.P. (1996) Isolation of expressed sequences that include a gene for familial breast cancer (*BRCA2*) and other novel transcripts from a five megabase region on chromosome 13q12. *Oncogene* **13**, 213-221.

Jiang, J., Gill, B.S., Wang, G.L., Ronald, P.C. and Ward, D.C. (1995) Metaphase and interphase fluorescence *in situ* hybridization mapping of the rice genome with bacterial artificial chromosomes. *Proc. Natl. Acad. Sci. USA* **92**, 4487-4491.

Jiang, Q., and Gresshoff, P.M. (1993) *Lotus japonicus*: A Model Legume System. In Plant Responses to the Environment, Ed. P.M.Gresshoff, CRC Press, Boca Raton, Florida. pp 97-110.

Jiang, Q., and Gresshoff, P.M. (1997) Classical and Molecular Genetics of the Model Legume *Lotus japonicus* *Mol. Plant Microbe Interact.* **10**, 59-68.

Kim, U.J., Shizuya, H., Chen, X.N., Deaven, L., Speicher, S., Solomon, J., Korenberg, J., and Simon, M.I. (1995) Characterization of a human chromosome 22 enriched bacterial artificial chromosome sublibrary. *Genet. Anal.* **12**, 73-79.

Kim, U.J., Birren, B.W., Slepak, T., Mancino, V., Boysen, C., Kang, H.L., Simon, M.I., and Shizuya, H. (1996) Construction and characterization of a human bacterial artificial chromosome library. *Genomics* **34**, 213-218.

Kim, U.J., Shizuya, H., Kang, H.L., Choi, S-S., Garrett, C.L., Smink, L.J., Birren, B.W., Korenberg, J.R., Dunham, I., and Simon, M.I. (1996) A bacterial artificial chromosome-based framework contig map of human chromosome 22q. *Proc. Natl. Acad. Sci. USA* **93**, 6297-6301.

Klein, L.R.M., Salentijn, E.M.J., Dirkse, W.G., Arens-de-Reveur, M. and Stiekema, W.J. (1994) Construction of a YAC library from a *Beta vulgaris* fragment addition and isolation of a major satellite DNA cluster linked to the beet cyst nematode resistance locus Hs1pat-1. *Theoretical and Applied Genetics* **89**, 426-434.

Kleine, M., Michalek, W., Graner, A., Herrmann, R.G., and Jung, C. (1993) Construction of a barley (*Hordeum vulgare* L.) YAC library and isolation of a *Horl*-specific clone. *Mol Gen Genet* **240**, 255-272.

Kohno, K., Oshiro, T., Kishine, H., Wada, M., Takeda, H., Ihara, N., Imamoto, F., Kano, Y., and Schlessinger, D. (1997) Construction and characterization of a *rad51rad52* double mutant as a host for YAC libraries. *Gene* **188**, 175-181.

Koike, K., Yoshino, K., Sue, N., Umehara, Y., Ashikawa, I., Kurata, N., and Sasaki, T. (1997) Physical mapping of rice chromosomes 4 and 7 using YAC clones. *DNA Res.* **4**, 27-33.

Kolchinsky, A., and Gresshoff, P.M. (1994) Plant Telomeres as Molecular Markers. In *Current Topics in Plant Molecular Biology*. P. M. Gresshoff, ed (Boca Raton, Florida : CRC Press), pp. 97-112.

Kolchinsky, A., Landau-Ellis, D., and Gresshoff, P.M. (1997) Map order and linkage distances of molecular markers close to the supernodulation (*nts-1*) locus of soybean. *Mol. Gen. Genet.* **254**, 29-36.

Koncz, C., Schell, J., and Redei, G.P. (1992) T-DNA transformation and insertion mutagenesis. In *Methods in Arabidopsis Research*, C. Koncz, N.H. Chua and J. Schell, eds (Singapore : World Scientific) pp. 119-143.

Korenberg, J.R., Chen, X.N., Mitchell, S., Fannin, S., Gerwehr, S., Cohen, D., and Chumakov, I. (1995) A high-fidelity physical map of human chromosome 21q in yeast artificial chromosomes. *Genome Res* 5, 427-443.

Kouchi, H., and Hata, S. (1993) Isolation and characterization of novel cDNAs representing genes expressed at early stages of nodule development. *Mol. Gen. Genet.* 238, 106-119.

Landau-Ellis, D., Angermuller, S., Shoemaker, R., and Gresshoff, P.M. (1991) The genetic locus controlling supernodulation in soybean (*Glycine max* L.) co-segregates tightly with a cloned molecular marker. *Mol. Gen. Genet.* 228, 221-226.

Landau-Ellis, D., and Gresshoff, P.M. (1992) Supernodulating soybean mutant alleles, nts382 and nts1007, show no recombination with the same RFLP marker. *Mol. Plant Microbe Interact.* 5, 428-429.

Landau-Ellis, D., and Gresshoff, P.M. (1994) The RFLP molecular marker closely linked to the supernodulation locus of soybean contains three inserts. *Mol. Plant Microbe Interact.* 7, 432-433.

Lapitan, N.L.V (1992) Organization and evolution of higher plant nuclear genomes. *Genome* 35, 171-181.

Larin, Z., Monaco, A.P. and Lehrach, H. (1991) Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci USA* 88, 4123-4127.

- Larin, Z., Monaco, A.P., Meier-Ewert, S., and Lehrach, H. (1993) Construction and characterization of yeast artificial chromosome libraries from the mouse genome. *Methods Enzymol* **225**, 623-637.
- Leeb, T., Rettenberger, G., Hameister, H., Brem, G., and Brenig, B. (1995) Construction of a porcine YAC library and mapping of the cardiac muscle ryanodine gene to chromosome 14q22-q23. *Mammalian Genome* **6**, 37-41.
- Leung, J., Bouvier-Durand, M., Morris, P-C., Guerrier, D., Cheddor, F., and Giraudat, J. (1994) *Arabidopsis* ABA response gene *ABI1*: features of a calcium-modulated protein phosphatase. *Science* **264**, 1448-1451.
- Libert, F., Lefort, A., Okimoto, R., Womack, J. and Georges, M. (1993) Construction of a bovine genomic library of large yeast artificial chromosome clones. *Genomics* **18**, 270-276.
- MacMurray, J.A., Weaver, A., Shin, H-P., and Lander, E.S. (1991) An automated method for DNA preparation from thousands of YAC clones. *Nucleic Acids Res.* **19**, 385-390.
- Marek, L.F., and Shoemaker, R.C. (1996) Construction and Size Characterization of a Bacterial Artificial Chromosome (BAC) library from Soybean. *Soybean Genetics Newsletter* **23**, 126-129.
- Martin, G.B., Williams, J., and Tanksley, S.D. (1991) Rapid identification of markers linked to a *Pseudomonas* resistane gene by using random primers and near isogenic lines. *Proc. Natl. Acad. Sci. USA* **88**, 2336-2340.
- Martin, G.B., Ganai, M.W. and Tanksley, S.D. (1992) Construction of a yeast artificial chromosome library of tomato and identification of cloned segments linked to two disease resistance loci. *Mol. Gen. Genet* **233**, 25-32.

Martin, G.B., Brommonschenkel, S.H.S.H., Chungwongse, J., Frary, A., Ganai, M.W., Spivey, R., Wu, T., Earle, E.D., and Tanksley, S.D. (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262**, 1432-1436.

Mathews, A., Carroll, B.J., and Gresshoff, P.M. (1990) The genetic interaction between non-nodulation and supernodulation in soybean : an example of developmental epistasis. *Theor. Appl. Genet.* **79**, 125-130.

Mejia, J.E., and Monaco, A.P. (1997) Retrofitting vectors for Escherichia coli-based artificial chromosomes (PACs and BACs) with markers for transfection studies. *Genome Res.* **7**, 179-186.

Meyer, K., Leube, M.P., and Grill, E. (1994) A protein phosphatase 2C involved in ABA signal transduction in *Arabidopsis thaliana*. *Science* **264**, 1452-1454.

Meyerowitz, E.M. (1987) *Arabidopsis thaliana*. *Annu Rev Genet* **21**, 93-111.

Nagaraja, R., Macmillan, S., Kere, J., Jones, C., Griffin, S., Scmatz, M., Terrell, J., Shomaker, M., Jermak, C., Hott, C., Masisi, M., Mumm, S., Srivastava, A., Pilia, G., Featherstone, T., Mazzarella, R., Kesterson, S., McCauley, B., Railey, B., Burough, F., Nowotny, V., D'Urso, M., States, D., Brownstein, B., and Schlessinger, D. (1997) X Chromosome Map at 75 kb STS Resolution, Revealing Extremes of Recombination and GC Content. *Genome Research* **7**, 210-222.

Nakamura, S., Asakawa, S., Ohmido, N., Fukui, K., Shimizu, N., and Kawasaki, S. (1997) Construction of an 800-kb contig in the near-centromeric region of the rice blast resistance gene *Pi-ta2* using a highly representative BAC library. *Mol. Gen. Genet.* **254**, 611-620.

Nam, H.G., Giraudat, J., Boer, D.B., Moonan, F., Loos, W.D.B., Hauge, B.M., and Goodman, H.M. (1989) Restriction fragment length polymorphism map of *Arabidopsis thaliana*. *Plant Cell* **1**, 699-705.

Nelson, D.L., Ledbetter, S.A., Corbo, L., Victoria, M.F., Ramirez-Solis, R., Webster, T.D., Ledbetter, D.H., and Caskey, C.T. (1989) *Alu* polymerase chain reaction : a method for rapid isolation of human specific sequences from complex DNA sources. *Proc. Natl. Acad. Sci. USA* **82**, 6686-6690.

Olson, M.V. (1993) The human genome project. *Proc. Natl. Acad. Sci. USA* **90**, 4338-4344.

Parimoo, S., Kolluri, R., and Weismann, S.M. (1993) cDNA selection from total yeast DNA containing YACs. *Nucleic Acids Res* **21**, 4422-4423.

Peterson, R.K., Clegg, C.H., Huxley, C., Josephson, B.M., Haugen, H.S., Furukawa, T., and Stamatoyannopoulos, G. (1993) Transgenic mice containing a 248 kb yeast artificial chromosome carrying the human β -globin locus display proper developmental control of human globin genes. *Proc. Natl. Acad. Sci. USA* **90**, 7593-7597.

Phillips, D.A. (1992) Flavonoids : plant signals to soil microbes. In: *Recent Advances in Phytochemistry*. Vol 26. Stafford, H.A., and Ibrahim, R.K., Eds., Plenum Press, New York, 201-231.

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

Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand. R., Smith, J.C., and Markham, A.F. (1990) A novel rapid method for isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18**, 2887-2890.

Rolfe, B.G., and Shine, J. (1984) Rhizobium-Leguminosae symbiosis : The bacterial point of view. In : *Genes Involved in Microbe-Plant Interactions*, Verma, D.P.S., and Hohn, T.H., Eds. Springer-Verlag, New York, pp. 95-128.

Rouquier, S., Batzer, M.A., and Giorgi, D. (1994) Application of bacterial artificial chromosomes to the generation of contiguous physical maps : a pilot study of human ryanodine receptor gene (*RYR1*) region. *Anal. Biochem.* **217**, 205-209.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning - A Laboratory Manual*, Second Edition (1989). Cold Spring Harbor Laboratory Press.

SanMiguel P., Tikhonov A., Jin Y.K., Motchoulskaia N., Zakharov D., Melake-Berhan A., Springer P.S., Edwards K.J., Lee M, Avramova Z, and Bennetzen J.L. (1996) Nested retrotransposons in the intergenic regions of the maize genome. *Science* **274**, 765-768.

Schmidt, R., Cnops, G., Bancroft, I. and Dean, C. (1992) Construction of an overlapping YAC library of the *Arabidopsis thaliana* genome. *Australian Journal of Plant Physiology* **19**, 341-351.

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

Schmidt, R., Love, K., West, J., Lenehan, Z., and Dean, C. (1997) Description of 31 YAC contigs spanning the majority of *Arabidopsis thaliana* chromosome 5. *Plant J.*, **11**, 563-572.

Schwartz, D.C. and Cantor, C.R. (1984) Separation of yeast chromosome sized DNAs by pulsed field gradient gel electrophoresis *Cell* **37**, 67-75.

Sheng, Y., Mancino, V., and Birren, B. (1995) Transformation of *Escherichia coli* with large DNA molecules by electroporation. *Nucleic Acids Res.* **23**, 1990-1996.

Shizuya, H., Birren, B., Kim, U.J., Mancino, 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. Natl. Acad. Sci. USA* **89**, 8794-8797.

Short, N. (1995) On first looking into the human genome. *Supp Nature* **377** (The Genome Directory), 1.

Silverman, G.A., Ye, R.D., Pollock, K.M., Sadler, J.E., and Korsmeyer, S.J. (1989) Use of yeast artificial chromosome clones for mapping and walking within human chromosome segment 18q21.3. *Proc. Natl. Acad. Sci. USA* **86**, 7485-7489.

Snell, R.G., Doucette-Stamm, L.A., Gillespie, K.M., Taylor, S.A., Riba, L., Bates, G.P., Altherr, M.R., MacDonald, M.E., Gusella, J.F., and Wasmuth, J.J. (1993) The isolation of cDNAs within the Huntington disease region by hybridization of yeast artificial chromosomes to a cDNA library. *Hum Mol Genet* **2**, 305-309.

Song, W-Y., Wang, G-L., Chen, L-L., Kim, H-S., Pi, L-Y., Holsten, T., Gardner, J., Wang, B., Zhai, W-X., Zhu, L-H., Fauquet, C., and Ronald,

P.C. (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. *Science* 270, 1804-1806.

Sorokin, A., Azevedo, V., Zumstein, E., Galleron, N., Ehrlich, S.D., and Serrero, P. (1996) Sequence analysis of the *Bacillus subtilis* chromosome region between the *serA* and *kdg* loci cloned in a yeast artificial chromosome. *Microbiology* 142, 2005-2016.

Spaink, H.P., Bloemberg, G.V., van Brussel, A.A.N., Lugtenberg, B.J.J., van der Drift, K.M.G.M., Haverkamp, J., and Thomas-Oates, J.E. (1995) Host specificity of *Rhizobium leguminosarum* is determined by the hydrophobicity of highly unsaturated fatty acyl moieties of the nodulation factors. *Mol. Plant-Microbe Interact.* 8, 155-164.

Stone, N.E., Fan, J.B., Willour, V., Pennacchio, L.A., Warrington, J.A., Hu, A., de la Chapelle, A., Lehesjoki, A.E., Cox, D.R., and Myers, R.M. (1996) Construction of a 750-kb bacterial clone contig and restriction map in the region of human chromosome 21 containing the progressive myoclonus epilepsy gene. *Genome Res.* 6, 218-225.

Szczyglowski, K., Hamburger, D., Kapranov, P., Shaw, R.S., and de Bruijn, F.J. (1995) Symbiotic mutants and nodule-specific ESTs of the model legume *Lotus japonicus*. In: *Nitrogen Fixation : Fundamentals and Achievements*. Tikhanovich, I., Proporov, N., Romanov, V., and Newton, W.E. Eds. Kluwer Academic Publishers, Dordrecht. pp 484.

Strauss, W.M., Jaenisch, E., and Jaenisch, R. (1992) A strategy for rapid production and screening of yeast artificial chromosome libraries. *Mamm. Genome* 2, 150-157.

Taketo, A. (1988) DNA transfection of *Escherichia coli* by electroporation. *Biochim. Biophys. Acta* **949**, 318.

Tanksley, S.D., Ganai, M.W., Prince, J.P., De Vincente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovanonni, J.J., Grandillo, S., Martin, G.B., Messeguer, R., Miller, J.C., Miller, L., Paterson, A.H., Pineda, O., Roder, M.S., Wing, R.A., Wu, W., and Young, N.D. (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics* **132**, 1141-1160.

Tanksley, S.D., Ganai, M.W., and Martin, G.B. (1995) Chromosome landing : a paradigm for map-based cloning in plants with large genomes. *Trends Genet* **11**(2), 63-68.

Umehara, Y., Inagaki, A., Tanoue, H., Yasukochi, Y., Nagamura, Y., Saji, S., Otsuki, Y., Fujimura, T., Kurata, N., and Minoba, Y. (1995) Construction and characterization of a rice YAC library for physical mapping. *Mol Breed* **1**, 79-89.

Umehara, Y., Kurata, N., Ashikawa, I., and Sasaki, T. (1997) Yeast artificial chromosome clones of rice chromosome 2 ordered using DNA markers. *DNA Res.* **4**, 127-131.

Van-Eck, J.M., Blowers, A.D. and Earle, E.D. (1995) Stable transformation of tomato cell cultures after bombardment with plasmid and YAC DNA. *Plant Cell Reports* **14** , 299-304.

Vasee, J., de Billy, F., Camut, S., and Truchet, G. (1990) Correlation between ultrastructural differentiation of bacteroids and nitrogen fixation in alfalfa nodules. *J. Bacteriol.* **172**, 4295-4306.

Verdoes, J.C., Calil, M.R., Punt, P.J., Debets, F., Swart, K., Stouthamer, A.H., and Van Den Hondel, C.A.M.J.J. (1994) The complete karyotype of *Aspergillus niger* : the use of induced electrophoretic mobility variation of chromosomes for gene assignment studies. *Mol Gen Genet* 244, 75-80.

Verma, D.P.S. and Delauney, A.J. (1992) Signals in root nodule organogenesis and endocytosis of *Rhizobium*. *Plant Cell* 4, 373-382.

Vijayaraghavan, U., Siddiqi, I., and Meyerowitz, E. (1995) Isolation of an 800 kb fragment encompassing a 3.5-cM region of chromosome 1 in *Arabidopsis* using YAC clones. *Genome* 38, 817-823.

Walter, M.A., Surti, U., Hofker, M.H., and Cox, D.W. (1990) The physical organization of the human immunoglobulin heavy chain gene locus. *EMBO J.* 9, 3303-3313.

Wang M., Chen, X.N., Shouse, S., Manson, J., Wu, Q., Li, R., Wrestler, J., Noya, D., Sun, Z.G., Korenberg, J et al (1994) Construction and characterization of a human chromosome 2-specific BAC library. *Genomics* 24, 527-534.

Wang, X., Qian, J., and Ko, M.S. (1994) Simple and robust screening of pooled yeast artificial chromosome libraries by the restriction enzyme digestion of polymerase chain reaction products. *Genet. Anal. Tech. Appl.* 11, 63-68.

Wang, G-L., Holsten, T.E., Song, W., Wang, H-P. and Ronald, P.C. (1995) Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa-21* disease resistance locus. *The Plant J.* 7, 525-533.

Wang, G-L., Warren, R., Innes, R., Osborne, B., Baker, B., and Ronald, P.C. (1996) Construction of an *Arabidopsis* BAC Library and Isoaltion of Clones

Hybridizing with Disease-Resistance Gene-like Sequences. *Plant Mol. Biol. Rep* **14**(2) 107-114.

Warburton, P.E., and Willard, H.F. (1990) Genomic analysis of sequence variation in tandemly repeated DNA in the human genome. *J. Mol. Biol.* **216**, 3-16.

Ward, E.R. and Jen, G.C. (1990) Isolation of single-copy-sequence clones from a yeast artificial chromosome library of randomly-sheared *Arabidopsis thaliana* DNA. *Plant Molecular Biology* **14**, 561-568.

Weaver, C.D., Shower, N.H., Louis, C.F., and Roberts, D.M. (1994) Nodulin-26, a nodule specific symbiosome membrane protein from soybean, is an ion channel. *J. Biol. Chem.* **269**, 17858-17862.

Wicking, C. and Williamson, B. (1991) From linked marker to a gene. *Trends Genet.* **7**, 288-293.

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.

Woo S-S, Rastogi, V.K., Zhang, H-B., Paterson, A.H, Schertz, K.F., and Wing, R.A. (1995) Isolation of megabase-sized DNA from sorghum and applications for physical mapping and bacterial and yeast artificial chromosome library construction. *Plant Mol Biol Rep* **13**, 38-50.

Zhang, H-B., Marin, G.B., Tanksley, S.D., and Wing, R.A. (1994) Map-based cloning in crop plants : tomato as a model systme. II. Isolation and characterization of a set of overlapping YACs encompassing the jointless locus. *Mol Gen Genet* **224**, 513-521.

Zhang, H-B., Martin, G.B., Tanksley, S.D. and Wing, R.A. (1995) Preparation of megabase sized DNA from plant nuclei. *Plant J.* 7, 175-184.

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.

Zhu, T., Shi, I., Funke, R.P., Gresshoff, P.M., and Keim, P. (1996) Characterization and application of soybean YACs to molecular cytogenetics. *Mol. Gen. Genet.* 252, 483-488.

Zimmer, R., and Verrinder-Gibbins, A.M. (1997) Construction and characterization of a large fragment chicken library. *Genomics* 42, 217-226.

VITA

Sanjeev Vasu Pillai was born in Trivandrum, India on February 28, 1969. A citizen of India, he attended elementary schools in Trivandrum and the neighboring district of Quilon, India. After completing a two year pre-degree course at the University of Kerala, Trivandrum, India he joined the Kerala Agricultural University, Trivandrum, India in from where he graduated with a Bachelor of Science (B Sc. Agricultural Sciences) in January, 1992.

After working for a couple of years in Trivandrum, India he came to the University of Tennessee, Knoxville in August 1994 to pursue a masters degree in Biotechnology. He worked in the Plant Molecular Genetics (PMG) lab for his thesis project under the guidance of Dr. Peter Gresshoff.

Having earned his M.S. working on high molecular weight DNA cloning, he will be furthering his career in the industry, working for a biotechnology company in the San Francisco bay area.