1994

Investigations of the organization of the genome of chestnut /

Jiansu Zhang
University of Massachusetts Amherst

INVESTIGATIONS OF THE ORGANIZATION OF THE GENOME OF CHESTNUT

A Thesis Presented

by

JIANSU ZHANG

Submitted to the Graduate School of the

University of Massachusetts Amherst in partial fulfillment

of the requirements for the degree of

MASTER OF SCIENCE

May 1994

Department of Plant and Soil Sciences

INVESTIGATIONS OF THE ORGANIZATION OF THE GENOME OF CHESTNUT

A Thesis Presented

by

JIANSU ZHANG

Approved as to style and content by:

Robert Bernatzky, Chair

Thomas Boyle, Member

F. Abel Ponce de Leon, Member

Lyle E. Craker, Department Head
Department of Plant and Soil Sciences

ACKNOWLEDGEMENTS

Thanks to my adviser Dr. Robert Bernatzky for his support, guidance, and kindness; Dr. Thomas Boyle and Dr. F. Abel Ponce de Leon for serving in the committee and giving valuable comments.

I also thank my coworkers Bruce Rivers, Keith Woeste, Wenqing Liang, and Bindu Chawla, for their friendship and help in the laboratory.

A special thanks to my husband for his immeasurable support.

ABSTRACT

INVESTIGATION OF THE ORGANIZATION OF THE GENOME OF CHESTNUT

MAY 1994

JIANSU ZHANG

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Robert Bernatzky

In order to facilitate the introgression of blight resistance from Chinese chestnut (Castanea mollissima) to the American chestnut (C. dentata), a total of 17 RFLP markers were developed for the chestnut genome. Fifteen markers were derived from a cDNA library of C. dentata and 2 markers from a PstI genomic library of C. dentata and C. mollissima. The PstI genomic library of chestnut was not a good source of single or low copy sequences that would serve as useful markers. Fourteen of the markers comprised five linkage groups.

A ribulose 1,5-bisphosphate carboxylase small subunit (rbcS) clone of chestnut was obtained by screening the cDNA library of C. dentata using rbcS-2 from tomato as a probe. The clone was sequenced. The 5′ missing sequence of the gene was obtained by PCR. Unlike the gene in other plant species, rbcS gene in chestnut does not belong to a multigene family.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The American chestnut (Castanea dentata) was once a economically important tree. It comprised more than 25% of the native eastern hardwood forest in the United States. It was a large tree and grew rapidly -- remarkable for a high-quality hardwood. It also set good crops of nuts every year.

In 1904, the American chestnut became infected with a blight-disease which was caused by the fungus Cryphonectria parasitica. The disease was introduced from Asia. By 1950 the blight had almost eliminated the tree. The tree now exist in the form of shrubs or small trees sprouting from the roots of previous trees. However, these remnants contain the genetic variation of the former population (Burnham, 1988).

Breeders have made an effort to restore the American chestnut. They tried to introduce blight resistance from Chinese chestnut (Castanea mollissima) which is highly resistant to the blight (Clapper, 1954). Along with other approaches, the backcross method has been employed, i.e. hybrids of American and Chinese chestnut have been backcrossed to the American chestnut while selecting for resistance in

progenies of each generation (Burnham et al., 1986). The difficulty breeders have met is that it is not possible to distinguish resistant individuals in the progenies of each backcross generation until sensitive individuals show signs of infection. However, if genetic markers become available that are linked to the resistance, they could be used for early detection of resistant individuals. Therefore It would be useful, in developing molecular genetic approaches to obtaining resistance, to know something about the organization of the chestnut genome.

In recent years, information about plant genome organization has increased rapidly. Genome organization in some species has been studied in detail, e.g. tomato (Zamir and Tanksley, 1988, Ganal et al., 1988), Arabidopsis (Pruitt and Meyerowitz, 1986), Rice (McCouch et al., 1988, Zhao et al., 1989).

Studying plant genome organization and development of chromosome markers has been shown to have many important applications. For example, species-specific repeat DNA sequences have been identified in a number of plant species. Since repetitive DNA sequences are relatively homogeneous within a species, but differ among species, highly polymorphic repetitive DNA sequences can provide abundant markers for

phylogenetic studies (Hoang-Tang et al., 1990), population studies (Condit and Hubbell, 1991), and identification of hybrids from somatic protoplast fusion (Schweizer et al., 1988). The repeated sequences that differ greatly in abundance among closely related taxa can be used for identification of diploid ancestors of polyploids (Dvorák et al., 1988), and as probes to detect introgression of foreign chromatin into crops (Zhang and Dvorák, 1990).

There is no information about the chestnut genome to date, and very little information about forest tree genomes in general. We only know that all Castanea species have the chromosome number of 2n=24 (Jaynes, 1962). The purpose of the present study is to provide some information on chestnut nuclear genome and to develop genetic markers for the chromosomes to facilitate restoration of American chestnut.

CHAPTER 2

LITERATURE REVIEW

2.1 Genome size variation in higher plants

Among diploid flowering plants, nuclear genomic DNA content varies more than several hundred fold, ranging from 0.14 pg in Arabidopsis thaliana to 127.4 pg in Fritillaria assyriaca. There is also variation in DNA content within families and even within species (Ohri and Khoshoo, 1986; Arumuganathan and Earle, 1991).

Some studies have shown intraspecific differences in DNA content. Hammatt et al. (1991) employed the technique of flow cytometry to investigate the DNA content of Glycine canescens. The DNA content in diploid G. canescens ranged from 3.80 to 6.59 pg without change in chromosome number. Intraspecific variation in the DNA content obviously exists in this species. Variation in intraspecific DNA content has also been detected in Zea mays (Rayburn et al., 1989), Poa annua, and Collinsia verna (Price, 1988).

Considerable interspecific and intergeneric variation in DNA content exists even among closely related taxa at the diploid level. It has been reported that there is 40-fold

difference in the family Droseraceae (Ohri and Khoshoo, 1986). In 21 diploid (2n=14) species of Lathyrus, a four-fold difference in DNA content was observed. The difference is less in inbreeding annual species and more in outbreeding perennial species. However, in 13 species of Ficus, a tropical hardwood, there is a significant uniformity in DNA content (Ohri and Khoshoo, 1986).

It has been shown that there can be a correlation between genome size and some ecological features of plants such as cold tolerance or environments such as altitude. McMurphy and Rayburn (1991) determined the DNA content of 8 maize populations and found that populations that were selected for tolerance to cold or freezing have significantly larger genomes than unselected populations. Rayburn and Auger (1990) reported that variation in genome size in maize is not random but seems to follow certain geographical distributions. A significant positive correlation between genome size and altitude was observed in southwestern U.S. Indian maize originating from Arizona. In contrast, a significant negative correlation between altitude and genome size has been reported in southwestern maize originating from New Mexico (Rayburn, 1990). Therefore, the nature of the relationship appears

dependent on the populations observed. This has led to speculation that genome size may have adaptive significance.

2.2 Genome organization in higher plants

Genome organization in higher plants has been studied using four methods: cytology, Mendelian genetics, DNA reassociation kinetics, and recombinant DNA technology. Using the first two methods, it is possible to localize certain genes to particular chromosomes. The third method has allowed the study of the distribution, percentage and average size of repetitive and unique DNA sequences. Recombinant DNA technology has enabled researchers to characterize plant genomes in even greater detail.

2.2.1 Repeated DNA sequences

A distinguishing characteristics of most higher plant genomes is the presence of large amounts of repetitive DNA sequences that can range from 10% to over 90% of the nuclear genome. The species with larger genomes usually contain a larger proportion of repeated DNA versus single copy DNA. There are several types of repeated DNA sequences in plant genomes that can be broadly classified as interspersed and

tandemly arranged sequences (Murray et al., 1981, Tanksley and Pichersky, 1988).

A great deal of repeated DNA exists in the form of tandem arrays, in which a basic sequence is repeated many times to form a large block. A tandem array can contain a very short (5-9 bp) repeated sequence and copy number in these arrays may be very high. Condit and Hubbell (1991) have studied dinucleotide repeat regions in several genera of tropical trees. They demonstrated that poly(AC) and poly(AG) sequences were clustered in tandem arrays and repeated $5x10^3$ to $3x10^5$ times, respectively, per genome. The poly(AC) sites were 16-22 bp in length and adjacent to either poly(AG) or poly(AT) sites. Some of the complex tandem repeat units are several hundred nucleotides long and may be repeated 10^4 to 10^5 times (Mitra and Bhatia, 1986).

Numerous repeated sequences are interspersed in many regions of the genome either with single copy DNA or with other repeats. In tomato, it has been found that a dispersed repeat sequence was scattered throughout nearly all chromosomes with an average of 133 kb between the repeated sequence. The total copy number of the repeat in the genome is about 4200 (Ganal et al., 1988). Another kind of dispersed repeat sequence has about 30 copies in the genome. One of the

copies was found within a chlorophyll a/b binding gene cluster (Bernatzky et al., 1988).

It has been demonstrated that repeated non-coding DNA sequences evolve faster than low-copy coding sequences. In tomato, divergence rates have been compared among different kinds of sequences by using random genomic clones and cDNA (complementary DNA derived from mRNA) clones as probes to hybridize with genomic DNA from different species. Under moderately stringent conditions, most of the single copy and middle repetitive sequence clones from the random genomic library cannot hybridize to DNA from distantly related species. The most highly repeated sequences can only hybridize to DNA from the most closely related species. However, 90% of sequences from cDNA clones can hybridize to all species that were tested (Zamir and Tanksley, 1988).

2.2.2 Ribosomal RNA genes

The 45S ribosomal RNA (rRNA) genes are relatively well studied. They occur in tandem arrays in the nucleolar organizing regions of chromosomes. The copy number is highly variable both between and within species. Each repeating unit consists of a transcribed region and a non-transcribed spacer (NTS). The average length of a repeating unit is around 10 kb,

in which the transcribed region is around 6 kb. The transcribed sequences are highly conserved but the NTS sequences evolve rapidly (Kalm et al., 1986).

In Cucurbita maxima there are 5,000 rDNA repeat units, constituting 7% of the haploid genome and forming many rDNA tandem arrays. A comparison of 29 rDNA clones from Cucurbita maxima showed that 21 are identical in length and restriction site pattern in the NTS sequence, while 8 differ in length (Kelly et al., 1990).

2.2.3 Multigene families

Some protein-coding genes are present in more than 1-2 copies, and there can be several to dozens of copies per genome. These copies are present either in tandem arrays or dispersed at several loci. Some of the well-studied multigene families are listed below.

Storage protein genes

The genes for zeins in maize exist as 54 and 24 copies and code for two kinds of polypeptides, respectively. They are dispersed at different loci (Willson and Larkins, 1984). The hordeins of barley are encoded by two clusters of genes, Hor-1 and Hor-2, each of which consists of 10-25 members. These two

clusters are tightly linked. There is variation in DNA sequence between the two clusters but not within each cluster (Forde et al., 1981).

Ribulose-1,5-bisphosphate carboxylase/oxygenase small subunit genes (rbcS)

It has been found that in all plant species studied so far rbcS genes belong to multigene families, and the rbcS gene families vary in size and organization among different plant species. In Arabidopsis thaliana, the rbcS gene family consists of 4 kinds of genes. Three of the genes are arranged in tandem and linked on 8 kb of chromosome. They share more than 95% similarity in DNA sequence. The fourth gene is not linked to the other genes, and is less similar to the other genes. All four genes are expressed (Krebbers et al., 1988). In tomato, five rbcS genes fall into three classes. Two unlinked loci each contains one copy, another unlinked locus contains 3 copies arranged in tandem (Pichersky et al., 1986).

Comparisons of the rbcS gene family members showed that there is intraspecific conservation at the nucleotide and amino acid sequence level within the coding region of these genes and interspecific homologies at the protein sequence level (Dean et al., 1989).

Chlorophyll a/b-binding protein genes (Cab)

Cab gene families are very complex. In tomato, so far, it has been shown that three loci (Cab-1, Cab-2,and Cab-3) contain at least 8 genes encoding the PSII TypeI Cab polypeptides, two loci (Cab-4, Cab-5) encode the PSII TypeII Cab polypeptides, two loci (Cab-6A, Cab-6B) encode the PSI TypeI Cab polypeptides, one locus (Cab-7) encode the PSI TypeII Cab polypeptides, and one locus (Cab-8) encode the PSI TypeIII Cab polypeptides. They are all independent genetic loci and have some degree of sequence divergence from each other (Pichersky et al., 1987, 1989).

In tomato, Cab-1 contains four genes arranged in tandem, Cab-3 contains at least three genes. The intergenic spacer regions in Cab-1 and Cab-3 have been examined. Most of the spacer regions are single copy. A middle-copy interspersed repeat was found between two of the genes in the Cab-1 gene cluster. Regions flanking the clusters have been shown to contain highly repeated sequences (Bernatzky et al., 1988).

Actin genes

Plant actin genes comprise a large family and exhibit a high level of divergence within species. In Petunia hybrida, the actin gene family contains more than 100 gene members,

which have been divided into 5 highly divergent subfamilies. Each subfamily contains 10 to 34 genes. These five subfamilies are located at four independent loci and two of them are tightly linked (McLean et al., 1988). In tomato, there are ten actin loci dispersed throughout the genome, three of them potentially contain a single gene, others may contain as many as 10 copies (Bernatzky and Tanksley, 1986a).

2.2.4 Single copy sequences

Single or low copy DNA sequences may constitute about 10-90% of a plant's genome. The total amount of single copy DNA varies among plant species even in closely related species. Therefore, it has been suggested that some of the single copy DNA does not have coding function, such as pseudo-genes (Tanksley and Pichersky, 1988), or fast evolving low copy non-coding sequences (Zamir and Tanksley, 1988). Bernatzky and Tanksley (1986b) have studied the copy number and number of loci corresponding to cDNA clones derived from tomato leaves. They investigated the segregation pattern of 34 unique, random cDNA clones in progeny of hybrids from two inbred lines. It was revealed that the majority of the genes (53%) are single locus while 32% of the clones are homologous to 2 loci. The remaining clones belong to gene families which

have 3-5 loci. Many of the single locus markers also contain single copies of the genes.

2.2.5 Genome organization in some best-studied species

Arabidopsis thaliana

Arabidopsis thaliana has the smallest nuclear genome (haploid DNA content C=0.07 pg) in higher plants that has been characterized to date. Leutwiler et al. (1984) employed a DNA reassociation technique to analyze total DNA. They found that the genome contains 10-14% highly repetitive sequences, 23-27% middle repetitive sequences and 50-55% single-copy sequences. The majority of this highly repetitive DNA is organellar DNA. Pruitt and Meyerowitz (1986) examined 50 random genomic DNA clones containing inserts with an average length of 12.8 kb. Of the 50 clones, 4 were chloroplast DNA clones, 32 were single-copy sequence clones, 3 contain both single-copy and repetitive sequences, and 2 contain sequences that are at different loci in the genome. Only 1 clone contains sequences that are present at more than ten copies per haploid genome and are highly conserved. Eight clones contained rDNA sequences. The rDNA exists as about 570 tandem copies of a 9.9 kb repeat unit per cell.

Lycopersicon esculentum

Tomato (L. esculentum) has a relatively small genome size (C=0.72 pg) (Galbraith et al., 1983). In tomato some of the repeated sequences and low-copy sequences in nuclear genome have been studied at the DNA level. Zamir and Tanksley (1988) studied 50 random clones derived from a sheared genomic DNA library. The size range of inserts in these clones was 350-2300 bp. Among these clones, 70% behaved as single copy when used as probes on Southern blots at high stringency, 4-10% contained repeated sequences (4% at high, 10% at moderate stringency), and the remaining clones belonged to multicopy families containing 2-20 copies. All of the repeated sequences contained in these random clones are dispersed repeats. Ganal et al. (1988) focused on the analysis of major families of repeated DNA sequences isolated from the sheared tomato DNA library. They screened 1,000 clones for repeated DNA sequences. Among these clones, 36 contained repeated sequences which belong to at least 15 different classes of repeated DNA sequences. Only 4 classes have more than 1,000 copies per genome. In these 4 major repeat classes, one has about 77,000 copies and consists of a 162 bp tandemly repeated satellite DNA. This indicates that the satellite DNA represents 1.75% of the genome. Another class of tandem

repeats is 45S rDNA genes which is a 9.1 kb repeating unit and exists as about 2,300 copies. The rDNA constitutes 3% of the genome. The third repeat class is interspersed throughout most of the chromosomes, has a size of 133 kb and about 4,200 copies per genome. The fourth class also represents an interspersed repeat, but does exhibit clustering on some of the chromosomes. This repeat has 2,100 copies and has homology with the 45S rDNA.

Zea mays

Zea mays has a haploid DNA content 7.70 pg. Repeated DNA sequences comprise 60-80% of the genome. The majority of repeated sequences are dispersed throughout the genome in a short interspersion pattern with unique or other repetitive sequences (Hake and Walbot, 1980). Gupta et al. (1984) analyzed 14 genomic clones which have inserts with an average size of 5 kb. They found that within the same clone all three classes of sequence (unique, middle and highly repetitive sequence) may be present. This indicates that the short interspersion pattern consists of unique, middle and highly repetitive sequences. The length of the dispersed repetitive elements are 300-1,300 bp. Some small dispersed repeats may also exist as part of a larger repeating unit.

CHAPTER 3

DEVELOPING GENETIC MARKERS FOR THE GENOME OF CHESTNUT

3.1 Introduction

The American chestnut (Castanea dentata) was once one of the most abundant and valuable trees in the eastern hardwood forest. In 1904, the chestnut blight Cryphonectria parasitica was first found in the United States. Within 40 years, the American chestnut had been eliminated as a major canopy tree.

In order to restore the American chestnut, breeders have tried to introduce resistance from Chinese chestnut (Castanea mollissima) to the American chestnut. Backcross breeding is one of the approaches being tried. Backcross breeding for disease resistance can be facilitated by the use of selectable molecular markers (Tanksley and Rick, 1980, Melchinger, 1990). The markers that are linked to disease resistance genes will permit early identification of individuals of backcross progeny which have resistance genes. Having markers scattered throughout the genome will allow selection against the genetic background of the donor parent, and this can reduce the number of backcross generations required to obtain individuals with

the phenotype of the American chestnut but with the blight resistance of the Chinese chestnut.

Herein is the results of an attempt to isolate low copy DNA markers for the chromosomes of chestnut. The sources of cloned sequences were genomic fragments derived from the restriction enzyme PstI, random genomic fragments generated with S_1 nuclease, or fragments derived from expressed sequences (complementary DNA or cDNA). Linkage relationships among markers was investigated in an F_2 population [(C. mollissima x C. dentata) x (C. mollissima x C. dentata)].

3.2 Materials and Methods

3.2.1 Plant materials

Leaves of American chestnut were collected from stump sprouts growing in a forest in New Salem, Massachusetts. The F_2 population was derived from a cross between two hybrids of American and Chinese chestnut growing at the Connecticut Agricultural Experimental Station orchard at the Sleeping Giant State Forest. A different American chestnut parent was used for each hybrid but the same Chinese chestnut parent ('Mahogany') was used as female (F.V. Hebbard, personal communication). Leaves of the Chinese chestnut parent and F_1s

were used as the source of DNA for screening of useful polymorphisms. The F_2 population used for linkage analysis is growing at the American chestnut Foundation Research Farm, Meadowview, VA and leaves were kindly provided by Dr. Fred V. Hebbard.

3.2.2 Isolating total genomic DNA

Total genomic DNA (primarily nuclear but containing a small amount of organellar DNA) was isolated from leaves according to Bernatzky and Tanksley (1986b) with modifications. Two grams of frozen leaves were ground with 5 ml of ice-cold extraction buffer (0.35 M sorbitol, 100 mM Tris-HCl pH 8.0, 50 mM EDTA, and 20 mM sodium metabisulfite). Forty ml of cold extraction buffer was added, and the homogenate was filtered through two layers of miracloth (Calbiochem). The homogenate was centrifuged at 1,500 g for 15 min at 4°C. The pellet was washed with 20 ml extraction buffer and resuspended in 5 ml extraction buffer. Five ml nuclei lysis buffer (200 mM Tris-HCl pH 8.0, 50 mM EDTA, 2 M NaCl, and 2% hexadecyltrimethylammonium bromide) and 5 ml of 5% sarcosyl were added, mixed, and incubated for 20 minutes at 60°C. The lysate was extracted with chloroform/isoamyl alcohol (24:1). DNA from the aqueous phase was precipitated with

isopropanol, washed twice with 70% ethanol, and dissolved in 0.2 ml of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). DNA concentrations were estimated by gel electrophoresis and utilizing uncut lambda DNA standards.

3.2.3 Cloning of genomic PstI DNA fragments from C. dentata, C. mollissima, and Brassica rapa.

Five μg of plant DNA was digested with 50 units of PstI (all restriction endonucleases and modifying enzymes used in the thesis were from Promega). The digested DNA was fractionated on 0.9% agarose gel. Approximately 0.5 μg of 0.5-2 kb DNA fragments were collected through electroelution using Model-UEA unidirectional electroelutor (IBI) following manufacturer's instructions.

All cloning procedures were done according to Sambrook et al. (1989). Ten μg of the plasmid pUC18 was completely digested with 50 units of PstI, and dephosphorylated with calf intestinal phosphatase. Approximately 0.2 μg of vector DNA was ligated with 0.1 μg of 0.5-2 kb plant DNA fragments. E. coli DH5α competent cells were made according to Sambrook et al. (1989) and transformed with these chimeric plasmids. Recombinant clones were selected on Luria broth plates containing ampicillin, X-gal and IPTG. Recombinant white

colonies were replated on the same media for a second screening.

3.2.4 Cloning of random, S_1-generated DNA fragments from C. dentata

Ten μg of total DNA was cut with 200 units of S_1 nuclease to the extent that most DNA fragments were 0.5-2 kb. Two μg of 0.5-2 kb DNA fragments were isolated through electroelution. 0.5 μg of DNA fragments were tailed with poly(dC) under the conditions of 0.1 mM dCTP, 15 units terminal deoxynucleotidyl transferase (TdT), at 37°C for 5 min. The reaction was stopped by addition 0.1 volume of 0.5 M EDTA. After phenol/chloroform (1:1) extraction, the DNA fragments were passed through Sephadex G50-80 column equilibrated with annealing buffer. Approximately 0.5 μg of PstI digested pUC18 were tailed with poly(dG) under the conditions of 0.01 mM dGTP, 15 units of TdT, at 37°C for 30 min, following by organic extraction and column chromatography as for poly(dC) tailed DNA fragments. An equal volume of dG-tailed pUC18 and dC-tailed plant DNA fragments were mixed, and annealled at 65°C for 5 min, 55°C for 1 hour, followed by 25°C for 2 hours. Transformation was performed as in section 3.2.3.

3.2.5 Screening genomic DNA libraries for low-copy DNA clones

Plasmids DNA were isolated from recombinant clones using the LiCl-boiling method (Willimzig, 1985). Plasmid DNA was digested with PstI, electrophoresed on 1% agarose gel and blotted onto Zetaprobe nylon membrane (Bio-Rad). Approximately 0.1 μg of total genomic DNA purified by a Geneclean kit (Bio 101 Inc.,) was labelled with ^{32}P using a random primer method (Feinberg and Vogelstein, 1983), and hybridized to their corresponding recombinant clones. All hybridization and wash procedures were according to Bernatzky and Schilling (1992). Clones were classified as low-copy or repetitive sequences based on intensity of autoradiographic signal. Homology to chloroplast DNA was determined for the PstI library of C. dentata and C. mollissima by hybridization of plasmid blots to ^{32}P-labelled radish (Raphanus sativa) DNA.

3.2.6 Construction of a cDNA library from C. dentata

Isolation of total RNA from young leaves was according to Harris and Dure (1978) with modifications. One gram of leaves were ground in 10 ml of 0.1 Tris-HCl (pH 7.6), 0.1 M NaCl, 0.001 M EDTA, and 1% SDS at room temperature (RT). An equal volume of phenol and chloroform (1:1) was added, vortexed for 20 min, and centrifuged at RT for 20 min at 3,000 rpm in an

IEC HN-SII centrifuge. One tenth volume of 3 M NaAc (pH 7.0) was added to the aqueous phase and the nucleic acids were precipitated with 2.5 volumes of ethanol at-20 °C for 2 hours. The precipitate was resuspended in 2 ml of 0.1 M Tris-HCl (pH 7.6), 0.1 M NaCl, 1 mM EDTA, and insoluble substances removed by centrifugation. The nucleic acids were reprecipitated with ethanol. The precipitate was washed once with 80% ethanol and dissolved in 1 ml of H_2O. From 1 g young leaves, approximately 5 mg of total RNA was obtained.

Poly-A RNA was prepared from total RNA utilizing the PolyATtract system (Promega Corp.) following manufacturer's instructions. Approximately 10 μg of poly-A RNA was obtained from 1 mg of total RNA. cDNA was synthesized from 1 μg of poly-A RNA using a cDNA synthesis kit (Pharmacia). After addition of EcoRI/NotI adaptors, the cDNA was passed through a Chroma spin-400 column (Clonetech Laboratories, Inc.) to remove excess adaptors and very small cDNA sequences. The cDNA was ligated to the vector pBluescript SK (+) (Stratagene) and used to transform competent DH5α cells. Colonies were screened on ampicillin/X-gal/IPTG plates.

3.2.7 Probe preparation

Recombinant plasmids DNA were isolated according to Willimzig (1985). The DNA inserts were amplified via the polymerase chain reaction (PCR) (Saiki et al., 1988). Amplification was carried out in 0.05 ml reaction volume containing 10 mM Tris-HCl (pH 8.8), 50 mM KCl, 20 mM MgCl2, 0.1% Triton X-100, 0.15 mM of each dNTP, 500 nM of each vector primer (KS, SK), 2.5 units of Taq DNA polymerase (Promega Corp.), and 1 ng of plasmid DNA denatured by boiling for 5 min. PCR reaction conditions were: 95°C, 30 sec; 55°C, 30 sec; 72°C, 2 min; 30 cycles. The PCR products were isolated by passing the sample over a Sephadex G50-80 column equilibrated with modified TE (10 mM Tris, pH 7.5, 0.5 mM EDTA). DNA concentration was estimated by running 1/10 of the reaction on 1.5% agarose gel along with lambda DNA standards.

3.2.8 Testing inheritance of selected low-copy clones with an F_2 population

Genomic Southern analysis was performed on the C. mollissima parent, an F_1 hybrid between the C. mollissima parent and C. dentata, and on F_2 progeny from a cross between two hybrids. DNA from the C. mollissima parent and F_1 hybrid were digested with restriction endonucleases, EcoRI, EcoRV,

and PstI. The digested DNA (1 μg/lane) were separated on 0.9% agarose gel and blotted onto Hybond N^+ nylon membranes (Amersham Corp.). The DNA inserts from cDNA clones and low-copy genomic DNA clones were used as probes to hybridize to those membranes. The clones that show polymorphism between the F_1 hybrid and C. mollissima parent were used as probes against 60 F_2 DNA samples digested with the restriction enzyme that gave rise to the polymorphism. Hybridization was carried out in a solution containing 5 X SSC, 50 mM sodium phosphate, pH 7.2, 0.6% SDS, 5 X Denhardt's solution, 2.5 mM EDTA, 100 ug/ml salmon sperm DNA, at 68°C for 16 hours as described in Bernatzky and Schilling (1992). The filters were washed at 68°C in 2 X SSC, 0.1% SDS for 10 min, and in 1 X SSC, 0.1% SDS for 20 min. then in 0.5 X SSC, 0.1% SDS for 20 min. After autoradiography, the probe was stripped from the filters by two washes in 0.1 X SSC, 0.1% SDS at 80°C for 10 min for reuse of the filters. The linkage relationships of segregating markers and goodness-of-fit tests to detect distorted segregation ratios were analyzed with the multipoint linkage analysis software LINKAGE-1 (Suiter et al., 1983).

3.3 Results

3.3.1 PstI genomic clones

One hundred and sixty-seven plasmids were isolated from the PstI genomic library of C. dentata and screened against ^{32}P-labelled total genomic DNA from C. dentata. Compared with the photographs of each corresponding gel, only 4% of clones did not show any bands with five days exposure to film and were classified as low or single copy sequence clones (Figure 3.1). One clone was identified as chloroplast (cpDNA) clone after probing with cpDNA from radish.

Ninety-seven plasmids were isolated from PstI genomic library of C. mollissima and 7% of clones contained single or low copy sequences (Figure 3.2). No cpDNA clones were found.

The low proportion of low-copy sequences in the PstI library is in contrast to what was found in tomato (Miller and Tanksley, 1990), rice (McCouch et al., 1988), and Brassica rapa (Figdore et al., 1988). In order to determine that the lack of low copy PstI sequences was not due to artifacts generated by the present methodology, a control PstI library was made from Brassica rapa. Sixty-three clones were tested from this library and 40% of clones were low-copy sequence clones (Figure 3.3).

Figure 3.1 Screening sample of PstI fragment clones of Castanea dentata for low-copy DNA sequences.

Recombinant genomic clones were digested with PstI, separated on 1% agarose gel, and probed with total genomic DNA of C. dentata. The clones that showed no hybridization signal were classified as low or single copy sequence clones, and the clones that showed strong or weak signal were classified as repetitive sequence clones. The arrow indicates an example of a low copy sequence clone. 1, 2 indicates different blots. Molecular weights are given based on HindIII fragments of lambda DNA.

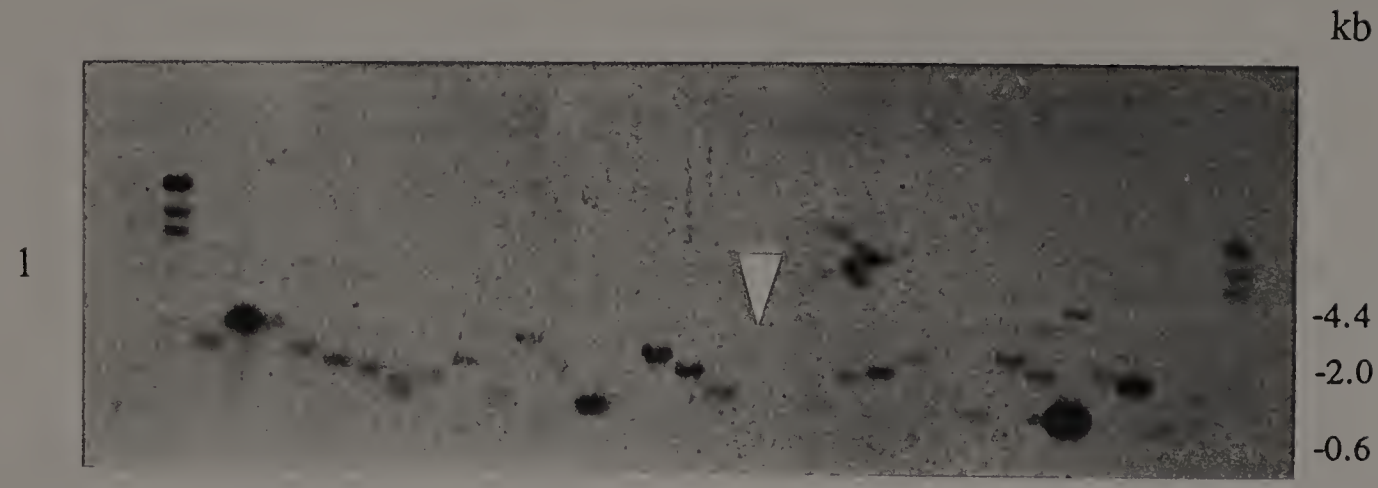

kb
1
-4.4
-2.0
-0.6
2
-4.4
-2.0
-0.6

Figure 3.2 Screening sample of PstI fragment clones of Castanea mollissima for low-copy DNA sequences.

Recombinant genomic clones were digested with PstI, separated on 1% agarose gel, and probed with total genomic DNA of C. mollissima. The clones that showed no hybridization signal were classified as low or single copy sequence clones, and the clones that showed strong or weak signal were classified as repetitive sequence clones. The arrow indicates an example of a low copy sequence clone. 1, 2, 3 indicates different blots. Molecular weights are given based on HindIII fragments of lambda DNA.

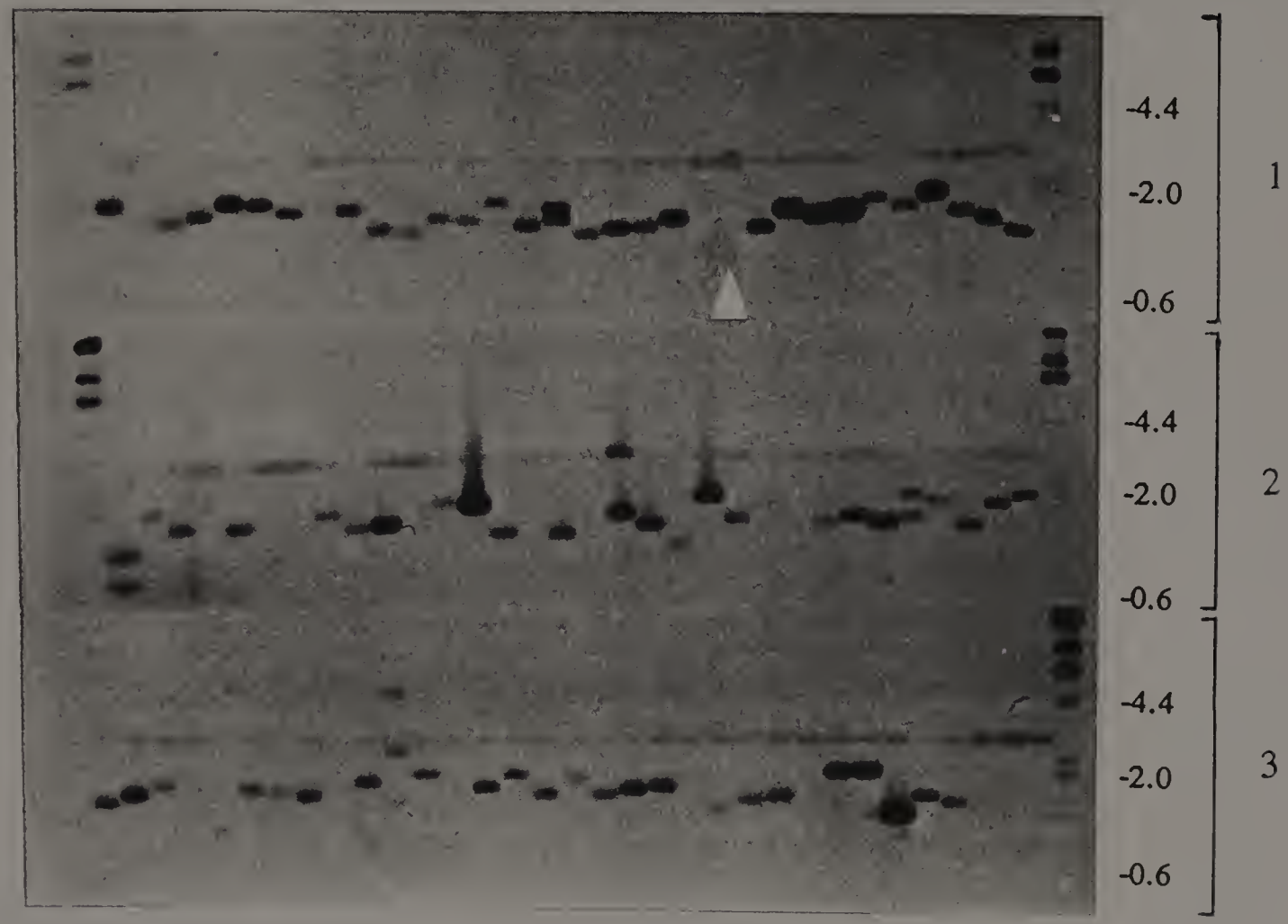
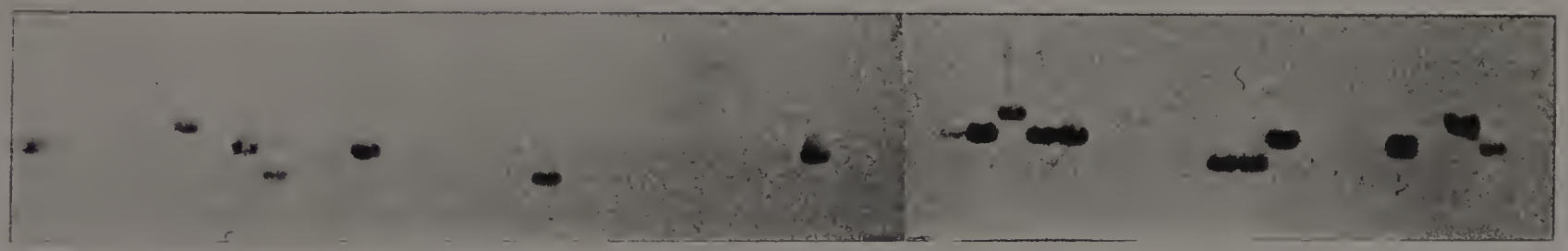

kb
-4.4
-2.0
-0.6
1
-4.4
-2.0
-0.6
2
-4.4
-2.0
-0.6
3

Figure 3.3 Screening sample of PstI fragment clones of Brassica rapa for low-copy DNA sequences.

Recombinant genomic clones were digested with PstI, separated on 1% agarose gel, and probed with total genomic DNA of B. rapa. The clones that showed no hybridization signal were classified as low or single copy sequence clones, and the clones that showed strong or weak signal were classified as repetitive sequence clones. Size markers in kilobase pairs (kb) shown at right.


kb
-2.0
-1.0

A total of 35 clones with varying intensity on the autoradiographs (28 from C. dentata, 7 from C. mollissima) were chosen, labelled with ^{32}P individually, and hybridized to C. mollissima and F_1 hybrid DNA cut with HindIII, EcoRI, and EcoRV. Among the 35 clones, 8 were low or single copy sequence clones. Four of these showed polymorphisms between C. mollissima and an F_1. Two clones (cm49, cd81) were used as markers for mapping. The hybridization signals of 27 high copy sequence clones were very intense when used as probes against C. mollissima and the F_1, and 23 clones only showed one or two bands for every enzyme digest on Southern blots. Only 4 clones showed several bands, and none of the clones hybridized to many bands under moderate conditions, which would be expected from dispersed repetitive sequences. Therefore, all cloned repetitive sequences gave patterns consistent with tandem organization of repeats.

In order to test if the cloned DNA is conserved between C. dentata and C. mollissima, total C. dentata DNA was used as a probe against the C. mollissima clones and total C. mollissima DNA as a probe against C. dentata clones. Hybridization patterns for intra- or interspecific probing were similar.

3.3.2 S_1-generated genomic clones

S_1 nuclease, capable of cleaving double-stranded DNA non-specifically in high concentrations, was used to generate a more random sample of nuclear sequences than with PstI which cleaves at specific sites. Eighty-seven plasmids were isolated from random genomic clones of C. dentata. The plasmids were cut with EcoRI and HindIII to release inserts. After screening against total C. dentata genomic DNA, 50 clones gave strong hybridization signal consistent with repeat sequence clones, whereas 37 (43%) clones were likely low or single copy sequence based on little or no signal (Figure 3.4). Among the 87 S_1 clones, 26 clones were cut into 2 or 3 pieces with EcoRI and HindIII. Only one clone contained both repeated and low-copy sequence separated by the double digest.

When those low or single copy sequence clones were used as probes on Southern blots of total chestnut DNA, they produced a very heavy background in general. It may be that long dC-dG tails caused non-specific binding to the nylon filter that is not removed during the washes. Therefore, these clones could not be reliably used as markers.

Figure 3.4 Screening sample of S_1 fragment clones of Castanea dentata for low-copy DNA sequences.

Recombinant genomic clones were digested with EcoRI and HindIII, separated on 1% agarose gel, and probed with total genomic DNA of C. dentata. The clones that showed no hybridization signal were classified as low or single copy sequence clones, and the clones that showed strong or weak signal were classified as repetitive sequence clones. The arrow indicates an example of a low copy sequence clone. 1, 2, 3 indicates different blots. Molecular weights are given based on HindIII fragments of lambda DNA.


kb
1
-4.4
-2.0
-0.6
2
-4.4
-2.0
-0.6
3
-4.4
-2.0
-0.6

3.3.3 cDNA library

One hundred and twelve plasmids were isolated from the cDNA library of C. dentata and were cut with EcoRI to release inserts. The cDNA library contained inserts larger than 300 bp probably because the cDNA used for constructing the library were passed through a Chroma spin 400 column and very small cDNA sequences were removed. Forty-eight clones with insert size of 0.5-2 kb were chosen to hybridize to F_1 and C. mollissima DNA cut with EcoRI, EcoRV and PstI. Thirty-one clones which showed polymorphism between the F_1 and C. mollissima were used for linkage analysis.

3.3.4 Genetic mapping of the clones

The clones that showed polymorphism between the F_1 and C. mollissima were used as probes on 58 individuals of the F_2 progeny. A total of 17 loci were scored. Fifteen were derived from cDNA clones, and 2 were from genomic clones (m49 from C. mollissima, d81 from C. dentata). The DNA sizes of 17 markers and the sizes of the fragments probed with the markers are listed in Table 3.1. Their linkage relationships are listed in Table 3.2. Fourteen loci correspond to 5 linkage groups and 3 loci did not show any significant linkage to the other loci (Figure 3.5). The map distances between loci were based on

Table 3.1 Size of the fragments probed with the inserts from the cDNA and genomic clones (kb)

Clones	Insert size (kb)	EcoRI cm	EcoRI F_1	EcoRV cm	EcoRV F_1	Pst I cm	Pst I F_1
192	0.5	9.5	9.5	1.0	1.0	>23.0	>23.0
			5.0				
120	1.2	7.0	7.0	8.0	8.0	7.5	7.5
				3.5	3.5	6.5	6.5
							4.0
145	0.8	5.5	5.5	7.5	7.5	>23.0	>23.0
		3.0	3.0		5.0		
		2.5	2.5				
d81	1.5	3.0	6.0	18.0	15.0		
			3.0	15.0	3.5		
				3.5			
m49	1.5	6.0	6.0	5.0	5.0		
			3.5		4.0		
048	1.0	6.0	9.0	5.0	5.0	>23.0	>23.0
			6.0		3.0		
173	0.5	8.0	12.0	18.0	18.0	18.0	18.0
			8.0	12.0	12.0		
062	1.6	12.0	12.0	10.0	10.0	12.0	20.0
		7.5		5.5	5.5		12.0
				2.0	2.0		
050	0.7	11.0	7.5			>23.0	>23.0
		7.5	6.0				
175	0.8	3.0	3.0	6.0	6.0	18.0	10.0
				5.5	5.5	10.0	2.5
					1.5		
084	2.0	9.5	5.0	10.0	15.0	7.0	9.5
		5.0	3.5	7.5	10.0	4.0	7.0
		3.5			7.5		4.0
							3.5
124	0.7	15.0	8.0	20.0	18.0	15.0	7.5
		7.5	7.0	18.0	15.0	6.5	6.5
		4.5	4.5	2.0	2.0	3.0	3.0
026	2.0	8.0	12.0	9.0	12.0	20.0	20.0
		5.0	8.0	6.0	9.0	5.5	7.0
		3.5	5.0		6.0		5.5
			3.5				

(continued)

Clones	Insert size (kb)	EcoRI cm	EcoRI F_1	EcoRV cm	EcoRV F_1	Pst I cm	Pst I F_1
143	0.7	10.0	15.0	15.0	15.0	18.0	18.0
		3.5	10.0	5.0	5.0		10.0
			3.5				
054	1.8	13.0	16.0	15.0	15.0	15.0	15.0
		9.0	9.0	13.0	13.0	12.0	12.0
		7.0	7.0		4.0	5.0	5.0
			5.0				
172	1.0	15.0	15.0	20.0	20.0	4.0	20.0
		9.0	9.0				4.0
093	0.8	10.0	17.0	10.0	10.0	>23.0	>23.0

Table 3.2 Linkage relationships among loci

	048	143	173	172	026	050	062	054	m49	124
145	31.7[a] (0.009)[b]									
192							40.1 (0.050)	24.1 (0.033)		
084		13.3 (0.000)			1.9 (0.000)					
026		10.2 (0.000)							39.1 (0.032)	
093						14.4 (0.000)				
062								26.1 (0.000)		
120	12.1 (0.000)								46.9 (0.012)	
124				22.2 (0.009)						
m49				42.3 (0.044)						
d81			8.8 (0.000)	13.4 (0.000)						
173				14.4 (0.000)						

a Recombination values in centiMorgans
b χ^2 probabilities for independant assortment

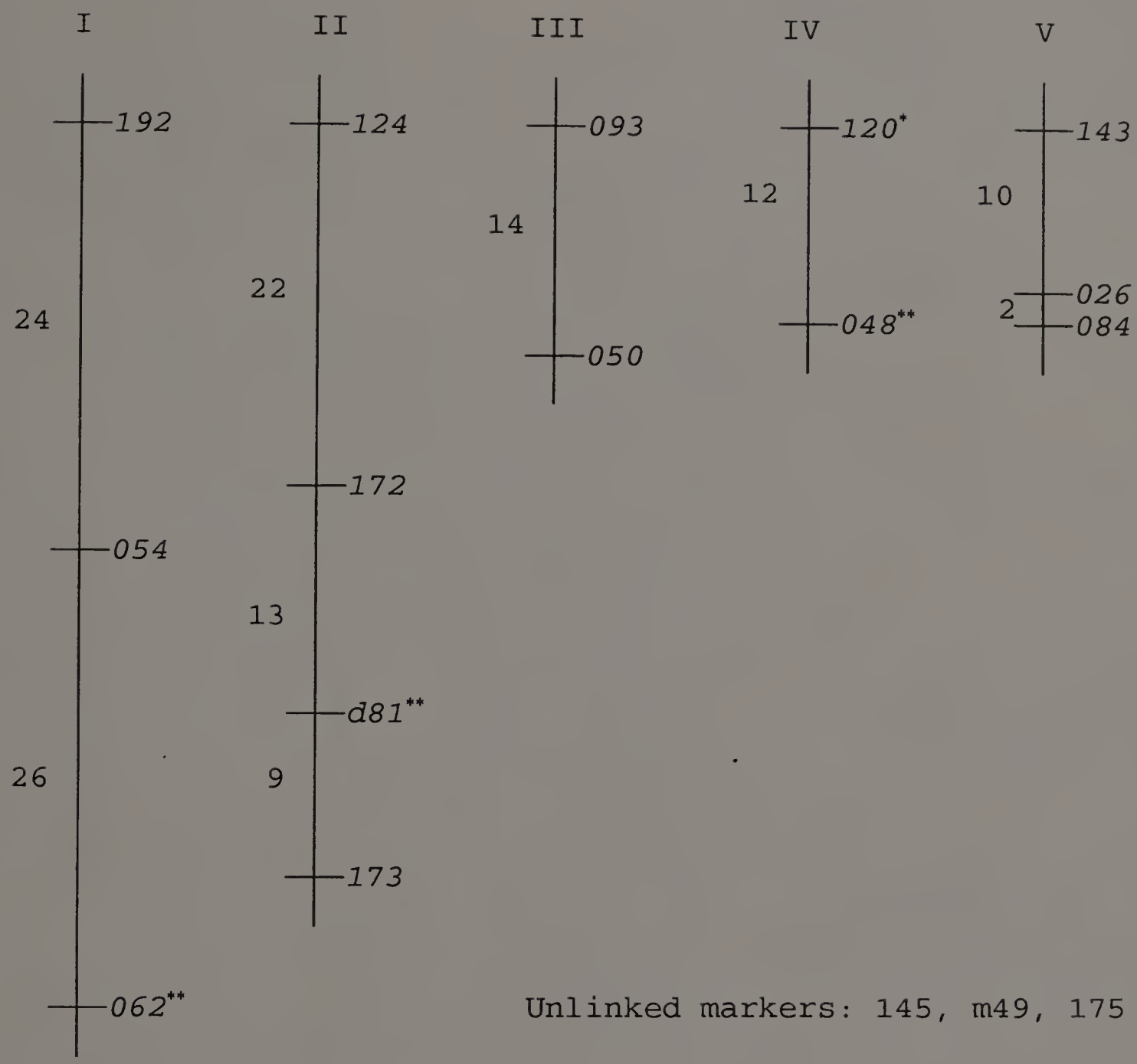


Figure 3.5 Genetic linkage map of chestnut genome.

The linkage group numbers are indicated on top. The distances between markers are shown in centiMorgans on the left. The skewed markers are labeled with * ($0.05 > P > 0.01$), ** ($P < 0.01$).

recombination values. Linkage was considered significant if the recombination value was less than 0.30.

The majority of cDNA clones (12/17) corresponded to single loci. Five of the cDNA clones may hybridize to more than one locus, since there were monomorphic fragments that were not segregating.

The segregation of 15 fragments were codominant markers in a 1:2:1 genotype ratio. One locus segregated in a 1:1 ratio since only 1 allele of C. dentata from one F_1 was scored. One locus segregated in a 3:1 ratio since only the allele from C. dentata was scored. The observed segregation ratio of all loci are listed in Table 3.3. Significant deviations ($P < 0.05$) of the segregation ratios towards one or another parental RFLP allele were observed at 4 loci (24 %). Three of the markers skewed in favor of C. mollissima alleles, while one skewed in favor of the C. dentata allele.

3.4 Discussion

3.4.1 Repetitive sequences in chestnut are under-methylated

The genomic libraries of both chestnut species were constructed with fragments derived from the restriction enzyme PstI. PstI is a methylation-sensitive enzyme. It recognizes

Table 3.3 Segregation ratios of loci (for 1:2:1 ratio)

Clone	cm/cm	cm/cd	cd/cd	χ^2
192	18	29	11	1.69
145	15	32	11	1.17
084	18	31	9	3.07
026	16	28	9	2.02
093	18	18	17	5.49
062	19	35	4	10.24**
120	23	25	9	7.74*
054[a]	15	43		0.07
124[b]	31	26		0.44
m49	15	28	15	0.07
d81	4	35	16	9.33**
048	24	26	8	9.45**
143	17	29	11	1.28
173	9	28	18	2.96
172	11	28	17	1.29
175	18	27	13	1.14
050	15	24	17	1.29

a For 3:1 ratio since only the allele from C. dentata was scored

b For 1:1 ratio since only 1 allele of C. dentata from one F_1 was scored

* $P < 0.05$

** $P < 0.01$

the sequence 5′-CTGCAG-3′, but will not cut if the cytosine at the 5′ end is methylated (Nelson and McCelland, 1987). It is expected that clones obtained from the PstI library are likely to be from hypomethylated regions of the genome. There is much evidence indicating that repeat sequences are more often C-methylated than coding sequences or regions upstream from coding regions (Burr et al., 1988). Therefore, a high proportion of single copy sequences are expected from a PstI generated library. This has been observed in several plant species. In tomato, 92% of clones from a PstI library were single copy, and 33% of the clones were single copy if the clones were derived from libraries constructed from sheared DNA or from EcoRI which is a C-methylation insensitive enzyme (Miller and Tanksley, 1990). A high proportion of low-copy sequences in PstI-generated libraries has also been observed in maize (Burr et al., 1988), common bean (Vallejos et al., 1992), and to a lesser extent (58%) in rice (McCouch et al., 1988). However, the result from chestnut is different, with only 4% and 7% of low or single copy sequence clones obtained from PstI genomic library of C. dentata and C. mollissima respectively, and 43% from S_1 random genomic library. In order to test whether this is a real difference or a result of different cloning and screening techniques, a PstI library of

Brassica rapa was constructed and screened in the same way. Forty percent of the clones were single or low copy sequences. The suspected repetitive cloned fragments of chestnut also give high intensity signals when used as probes on Southern blots of chestnut DNA. Thirty-five clones were chosen as probes to hybridize to C. mollissima and F_1 hybrids digested with HindIII, EcoRI, and EcoRV. The clones which were classified as low or single copy sequence showed low intensity signal. The clones which were classified as repetitive sequences showed high intensity signal. All clones tested showed at most three bands, and most of the clones showed only one band. This suggests that the majority of the cloned repetitive sequences are tandemly repeated.

3.4.2 Segregation distortion in chestnut

The inheritance pattern of the 17 markers were analyzed in a segregating population of 58 F_2 progeny. Four (24%) of the markers deviated significantly ($P < 0.05$) and 3 of the markers highly significantly ($P < 0.01$) from expected Mendelian segregation ratios. Segregation distortion has been observed in some other plant species, as well, e.g. in tomato (18.6%; Bernatzky and Tanksley, 1986b), rice (18.8%; McCouch

et al., 1988), potato (25.5%; Gebhardt et al., 1989), barley (10%; Heun et al., 1991).

Among 4 deviated loci, 3 loci were distorted towards C. mollissima parental allele. This suggests that a higher proportion of the genome in the F_2 is from C. mollissima and breeders will have more difficulty in restoring C. dentata with disease resistance. Therefore, these markers will be useful to select individuals with higher levels of C. dentata chromosomes and reduce the number of backcross generations required to obtain individuals with the phenotype of the C. dentata but with the blight resistance of C. mollissima.

3.4.3 RFLP markers vs other DNA markers

In this study, RFLP chromosome markers were generated. These markers will be used to attempt to tag blight disease resistant genes and as guidance for getting rid of the genetic background from C. mollissima in backcross progeny. Other approaches are also available for selection against genetic backgrounds. For example, randomly amplified polymorphic DNA (RAPD), dispersed repetitive sequences and microsatellite markers are very useful in whole genome selection, since they generally detect multiple loci per amplification or

hybridization reaction. Selection can be based on pattern similarity, and genetic mapping is not required.

CHAPTER 4

NUCLEOTIDE SEQUENCE OF THE SMALL SUBUNIT OF RIBULOSE 1,5-BISPHOSPHATE CARBOXYLASE OF CASTANEA DENTATA AND INHERITANCE AS A SINGLE LOCUS IN C. DENTATA AND C. MOLLISSIMA

4.1 Introduction

Ribulose 1,5-bisphosphate carboxylase (EC 4.1.1.39) is the enzyme responsible for the initial fixation of carbon dioxide in plants. In vascular plants and green algae, the holoenzyme consists of eight large subunits, encoded by the chloroplast genome, and eight small subunits, encoded by the nuclear genome. The products of the small subunit genes, rbcS, are translated in the cytoplasm as higher molecular weight precursors and transported into the chloroplast where they are processed and assembled with the large subunits (for review see Dean et al., 1989).

In most plants studied, rbcS is encoded by a multigene family. However, there exists variation for the number of copies of the gene. The different copies are classified into subfamilies based on degrees of sequence homology and structure, such as position of introns. The different subfamilies are generally organized as discrete loci. For

example, among dicots, there are eight copies of rbcS in Petunia hybrida that comprise three subfamilies. Two families contain one copy, the third has six and the three families reside at three loci (Dean et al., 1985). A similar situation exists in two other members of the Solanaceae, tomato and potato, where one locus contains multiple copies of rbcS and two other loci contain single genes (Sugita et al., 1987, Wolter et al., 1988). However, in another member of this family, the genus Nicotiana, there are multiple copies but the copies do not appear to be closely linked. Plants from other families also contain multiple copies organized into subfamilies and multiple chromosomal locations. The number of rbcS genes and loci reported in some plant species are summarized in Table 4.1.

In this study, the cDNA library of C. dentata was screened for rbcS and locus number was determined for C. dentata and C. mollissima.

4.2 Materials and Methods

4.2.1 Screening the cDNA library for rbcS sequences

The construction of cDNA library of C. dentata has been described in chapter 3. Recombinant plasmids DNA were isolated

according to Willimzig (1985). The plasmid DNA was digested with EcoRI, separated on a 1.0% agarose gel and blotted onto a Hybond N+ nylon membrane (Amersham). Hybridization was carried out overnight at 68°C according to Bernatzky and Schilling (1992) using the cloned rbcS-2 gene from tomato (Pichersky et al., 1986) as a probe. The probe was prepared from plasmid insert and labeled by random priming (Feinberg and Vogelstein, 1983). The filters were washed at 60°C in 2 X SSC, 0.1% SDS for 20 minutes and then in 1 X SSC, 0.1% SDS for 20 minutes. The filters were exposed to X-ray film (Kodak) for 2-5 hours with an intensifier screen.

4.2.2 DNA sequencing

Double-stranded DNA was sequenced using ^{35}S-dATP, pBluescript primers (SK,KS), and the Sequenase T_7 DNA polymerase system (U.S. Biochemical) according to manufacturer's protocol. Band compressions were resolved through repeat sequencing reactions substituting dITP for dGTP in the reaction mixture (Gough and Murray, 1983). Sequence data were analyzed using the Wisconsin Genetics Computer Group Software package.

4.2.3 Obtaining 5' *rbcS* DNA sequences

A 20-mer oligo (complementary to the sequence from base 327 to 346 in Figure 4.1) was synthesized and used as a specific primer for *rbcS*. PCR was performed on the original cDNA library (cDNA ligated to the vector) utilizing the *rbcS* primer and either the reverse or forward sequencing primer of the vector. PCR reaction conditions were as same as described in chapter 3. PCR products were separated on 1.5 % agarose gel, and isolated from agarose gel according to Tautz and Renz (1983). PCR products were cloned into pUC18 cut with *Sma*I according to Liu and Schwartz (1992). The cloned DNA was sequenced using the specific *rbcS* primer and one of the pBluescript primers as described in 4.2.2.

4.2.4 Inheritance testing

Genomic Southern analysis was performed on a *C. mollissima* parent, an F_1 hybrid between the *C. mollissima* parent and an individual *C. dentata*, and on F_2 progeny from the hybrid. DNA from the *C. mollissima* parent and the hybrid were digested with *Hind*III, *Eco*RI, and *Eco*RV. Digested DNA (approximately 1 μg/lane) was separated on 0.9% agarose and replica Southern blots were produced. The chestnut *rbcS* sequences were used as probes to hybridize to the filters

under moderate stringency conditions at 68°C for 16 hours as described in Bernatzky and Schilling (1992). The filters were washed at 68°C twice in 2 X SSC, 0.1% SDS for 10 minutes, then in 1 X SSC, 0.1% SDS for 20 minutes. One filter was further washed at 68°C in 0.5 X SSC, 0.1% SDS for 20 minutes. After autoradiography, the probe was stripped from the filters by two washes in 0.1 X SSC, 0.1% SDS at 80°C for 10 minutes. The filters were then hybridized to a labeled chestnut rbcS sequence under low stringency conditions at 60°C, and washes were performed at 60°C twice in 2 X SSC, 0.1% SDS for 10 minutes, then in 1 X SSC, 0.1% SDS for 20 minutes. The filters were then exposed to films for 10 days. A polymorphism that was observed between the C. mollissima parent and the F_1 with the enzyme EcoRI was used to produce Southern blots of 58 F_2 progeny in order to monitor segregation of sequences that hybridized to the chestnut rbcS probe. Hybridization and washes were carried out under the moderate stringency conditions described above.

4.3 Results

4.3.1 Screening and sequencing of rbcS gene

The cDNA clones with inserts approximately greater than 400 bp were isolated from 112 colonies. The plasmid DNA was digested with EcoRI, separated on a 1.0% agarose gel and blotted onto a membrane. The rbcS-2 gene from tomato was used as a probe to hybridize to the membrane. One positive clone was obtained. The clone contained an insert of 483 bp in length and was sequenced. Comparison of the sequence of rbcS-2 from tomato revealled that the chestnut clone was incomplete and the 5′ of the gene sequence was missing. PCR was performed on the original cDNA library using the rbcS primer and either the reverse or forward sequencing primer of the vector in order to obtain upstream sequence. Three major bands of approximately 370, 250, and 170 bp were obtained. The PCR products were cloned into the pUC18 vector. The clone containing the 370 bp insert was sequenced and contains information from the rbcS primer through the start site. The two other clones of PCR products containing 250 and 170 bp were also sequenced. The three clones of PCR products contained the same sequence information, differing only in length at the 5′ end. The total DNA sequenced was 759 bp in

length, and the coding sequence was 549 bp (Figure 4.1). It has 84% similarity at amino acid sequence level and 74% similarity at the nucleotide sequence level as compared with rbcS-2 from tomato.

4.3.2 Southern analysis

Genomic Southern blot of a C. mollissima parent, and F_1 hybrid DNA digested with HindIII, EcoRI, and EcoRV, was probed with the chestnut rbcS sequences. At both moderate and low stringency, only one band was obtained with EcoRI or EcoRV digestion, and three bands with HindIII digestion (Figure 4.2). A polymorphism was observed between the C. mollissima parent and the F_1 with EcoRI and HindIII digestion. Southern blots of 58 F_2 progeny DNA digested with EcoRI were used to monitor segregation of rbcS sequences. The segregation ratio of this chromosome fragment was 18:29:11, homozygous C. mollissima:heterozygote:homozygous C. dentata (Figure 4.3). This is not significantly different from an expected 1:2:1 ratio ($\chi^2 = 1.69$, $P < 0.05$).

4.4 Discussion

Southern hybridization indicates that rbcS does not belong to a multigene family in Castanea. Even at low

Figure 4.1 Nucleotide sequence of rbcS cDNA of C. dentata.

The underlined sequence indicates the rbcS primer.

Additional sequence (AGCGAA,not shown) was read 5' to the ATG start.

```
  1  ATGGCTTCCT CAATTCTTTC CTCCGCTACC GTCGCCCCCA TTAACCGGGC
 51  TACCCCTGCT CAAGCCAGCA TGGTCGCACC ATTCACTGGC CTCAAGTCTA
101  ACGCAGGCTT CCCAGTCACC CAAAAGACCA ACAAAGACAT TACCTCTCTT
151  GCTAGCAATG GTGGAAGAGT GCAATGCATG CAGGTGTGGC CTCCACTTGG
201  ATTAAAGAAG TTTGAGACCC TTTCCTACCT TCCACCACTT ACTAACGAGC
251  AATTGGCTAA GGAAGTAGAC TACCTTCTTC GCAAGGGATG GGTTCCTTGC
301  TTAGAATTTG AATTGGAGCA CCCCTTTGTG TACCGTGAGA ACAATAGGTC
351  ACCAGGGTAC TATGATGGAC GCTACTGGGT GATGTGGAAG CTTCCCATGT
401  TTGGATGCAC CGATTCCGCT CAGGTGTTGA GGGAGGTTGA GGAGGTCAAG
451  AAGGTTTACC CCAGTGCCCA TGTCCGAATC ATTGGATTCG ACAACGTACG
501  TCAAGTGCAG TGCATCAGTT TCATTGCTTA CAAGCCTCCT AGCGTCTAAG
551  ATGTTTATTA TCATTCTCTA AACGTACCCT TTTTCGAGGG TCGGTTTGTT
601  TAAATTGTAT TTTTAGGCTT TCAAAAGACA TTTCTGTTTC ATTTTCGAGA
651  CAATTCGCTC TGTTTTGAAT TTGTGTTTTT CGGATTTCCC ATGGAATGGA
701  TGAGAACTGA TTAATAAAAT TGGCTAGTTT CTTGCCAATT CCATGAACTC
751  AAA
```

Figure 4.2 Southern hybridization of rbcS gene.

Southern blot of C. mollissima (Cm) and F_1 hybrid digested with HindIII, EcoRI, and EcoRV, probed with rbcS gene. A. moderate stringency B. low stringency. Size markers in kilobase pairs (kb) shown at right.

Eco RV *Eco* RI *Hind* III

F1 Cm

kb

23.0

9.4

6.6

4.4

2.2

2.0

0.56

A 68 °C 0.5X SSC

60 °C 1.0X SSC B

Figure 4.3 Segregation pattern of rbcS sequence in an F_2 population.

Southern blot of genomic DNA of F_2 individuals digested with EcoRI, and probed with rbcS cDNA fragment. Cd: the allele from C. dentata, and Cm: the allele from C. mollissima. Size markers in kilobase pairs (kb) are shown on the left.

kb

23.0

4.4

2.0

Cm

Cd

5'

3'

stringency, additional copies of the gene are not evident. The segregation data indicates that the hybridizing fragments belong to a single locus. Based on the smallest restriction fragment(s) obtained (approximately 1 kb, EcoRV), it is likely that there is only one copy of the gene present as well. In all plants studied so far rbcS belongs to a multigene family (Table 4.1). Even in Arabidopsis thaliana, with one of the smallest plant genomes, two loci and four copies of the gene have been found.

It is commonly believed that multigene families exist for either 1) the production of high levels of gene product or 2) differential regulation of the subfamily members. Considering that Castanea likely has only one copy of the rbcS gene, it would appear that multiple copies of this gene are not strictly required for the process of carbon fixation in plants.

Table 4.1 Number of rbcS genes and loci reported in other plant species.

Species	Locus	Number of genes
Petunia	1	1
	2	1
	3	6
Tomato	1	1
	2	1
	3	3
Potato	1	1
	2	1
	3	3
Pea	1	5
Soybean	ND	>6
Arabidopsis	1	3
	2	1
Lemna gibba	ND	>6
Wheat	Nd	>12

REFERENCES

Arummuganathan, K. and Earle, E.D. 1991. Estimation of nuclear DNA content of plants by flow cytometry. Plant Mol. Bio. Rep. 9:221-231.

Bernatzky R. and Schilling, A. 1992 Methods of southern blotting and hybridization. In: Plant genomes: Methods for genetic and physical mapping, Beckmann, J.S. and Osborn, T.C. eds. Kluwer Academic Publishers, The Netherlands.

Bernatzky, R. and Tanksley, S.D. 1986a. Genetics of actin-related sequences on tomato. Theor. Appl. Genet. 72:314-321.

Bernatzky, R. and Tanksley, S.D. 1986b. Majority of random cDNA clones correspond to single loci in the tomato genome. Mol. Gen. Genet. 203:8-14.

Bernatzky, R., Pichersky, E., Malik, V.S. and Tanksley, S.D. 1988. CR1 - a dispersed repeated element associated with the Cab-1 locus in tomato. Plant Mol. Biol. 10:423-433.

Burnham, C.R. 1988. The restoration of the American chestnut. American Scientist 76:478-487.

Burnham, c.R., Rutter, P.A. and French, D.W. 1986. Breeding blight resistant chestnuts. Plant Breeding Reviews 4:347-397.

Burr, B., Burr, F.A., thompson, K.H., Albertson, M.C. and Stuber, C.W. 1988. Gene mapping with recombinant inbreds in maize. Genetics 118:519-526.

Clapper, R.B. 1954. Relative blight resistance of some chestnut species and hybrids. J. For. 50:453-455.

Condit, R. and Hubbell, S.P. 1991. Abundance and DNA sequence of two-base repeat regions in tropical tree genomes. Genome 34:66-71.

Dean, C., van den Elzen, P., Tamaki, S., Dunsmuir, P. and Bedbrook, J. 1985. Linkage and homology analysis divides the eight genes for the small subunit of Petunia ribulose 1,5-bisphosphate carboxylase into three gene families. Proc. Natl. Acad. Sci. USA. 82:4964-4968.

Dean C., Pichersky, E. and Dunsmuir, P. 1989. Structure, evolution and regulation of rbcS genes in higher plants. Ann. Rev. Plant Physiol. Plant Mol. Biol. 40:415-439.

Dvorak, J., McGuire, P.E. and Cassidy, B. 1988. Apparent sources of the A genomes of wheats inferred from polymorphism in abundance and restriction fragment length of repeated nucleotide sequences. Genome 30:680-689.

Feinberg, A.P. and Vogelstein, B. 1983. A technique for radio-labeling DNA restriction fragments to high specific activity. Anal. Biochem. 132:6-13.

Figdore, S.S., Kennard, W.C., song, K.M., Slocum, M.K. and Osborn, T.C. 1988. Assessment of the degree of restriction fragment length polymorphism in Brassica. Theor. Appl. Genet. 75:833-840.

Forde, B.G., Kreis, M., Bahramian, M.B., Matthews, J.A. and Miflin, B.J. 1981. Molecular cloning and analysis of cDNA sequences derived from poly A+ RNA from barley endosperm: identification of B hordein clones. Nucl. Acids Res. 9:6689-6707.

Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P. and Firoozabady, E. 1983. Rapid flow cytometric analysis of the cell cycle in intact plant tissues. Sciences 220:1049-1051.

Ganal, M.W., Lapitan, N.L.V. and Tanksley, S.D. 1988. A molecular and cytogenetic survey of major repeated DNA sequences in tomato (Lycopersicon esculentum). Mol. Gen. Genet. 213:262-268.

Gebhardt, C., Ritter, E., Debener, T., Schachtschabel, V., Aakemeier, B, Uhrig, H. and Salamini, F. 1989 RFLP analysis and linkage mapping in Solanum tuberosumm. Theor. Appl. Genet. 78:65-75.

Gough, J.A. and Murray, N.E. 1983. Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol. 166:1-19.

Gupta, M., Shepherd, N.S., Bertram, I. and Saedler, H. 1984. Repetitive sequences and their organization on genomic clones of Zea mays. EMBO J 3:133-139.

Hake, S. and Walbot, V. 1980. The genome of Zea mays, its organization and homology to related grasses. Chromosoma 79:251-270.

Hammatt, N., Blackhall, N.W. and Davey, M.R. 1991. Variation in the DNA content of Glycine species. J. Exp. Bot. 42: 659--665.

Harris, B. and Dure, L.S. III 1978. Developmental regulation in cotton seed germination: Polyadenylation of stored messenger RNA. Biochem. 17:3250-3256.

Heun, M., Kennedy, A.E., Anderson, J.A., Lapitan, N.L.V., Sorrells, M.E. and Tanksley, S.D. 1991 Construction of a restriction fragment length polymorphism map for barley (Hordeum vulgare). Genome 34:437-447.

Hoang-Tang, Dube, S.K., Liang, G.H. and Kung, S.D. 1991. Possible repetitive DNA markers for Eusorghum and Parasorghum and their potential use in examining phylogenetic hypotheses on the origin of Sorghum species. Genome 34:241-250.

Jaynes, R.A. 1962. Chestnut Chromosomes. For. Sci. 8:372-377.

Kalm, L.V., Vize, P. and Smyth, D.R. 1986. Structure of ribosomal RNA genes in Lilium species. In: Reddy, G.M. and Coe, E.H. Jr. (eds) Gene Structure and Function in Higher Plants. Oxford & IBH Publishing CO., New Delhi.

Kelly, R.J., Johnson, R.D. and Siegel, A. 1990. Heterogeneity and organization of the ribosomal RNA genes of Cucurbita maxima. Plant Mol. Biol. 14: 927-933.

Krebbers, E., Seruinck, J., Hdrdies, L., Cashmore, R. and Timko, M.P. 1988. Four genes in two diverged subfamilies encode the ribulose-1, 5-bisphosphate carboxylase small subunit polypeptides of Arabidopsis thaliana. Plant Mol. Biol. 11:745-759.

Leutwiler, L.S., Hough-Evans, B.R. and Meyerowitz, E.M. 1984. The DNA of Arabidopsis thaliana. Mol. Gen. Genet. 194:15-23.

Liu, Z.G. and Schwartz, L.M. 1992. An efficient method for blunt-end cloning of PCR products. Biotechniques 12:7-9.

Maniatis, T., Fritsch, E.F. and Sambrook, J. 1982. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, New York.

McCouch, S.R., Kochert, G., Yu, Z.H., Wang, Z.Y., Khush, G.S., Coffman, W.R., Tanksley, S.D. 1988. Molecular mapping of rice chromosomes. Theor. Appl. Genet. 76:815-829.

McLean, M., Baird, v., Greats, A.G.M. and Meagher, R.B. 1988. Determination of copy number and linkage relationships among five actin gene subfamilies in Petunia hybrida. Plant Mol. Biol. 11:663-672.

McMurphy, L.M. and Rayburn, A.L. 1991. Genome size variation in maize populations selected for cold tolerance. Plant Breeding 106:190-195.

Melchinger, A.E. 1990. Use of molecular markers in breeding for oligogenic disease resistance. Plant Breeding 104:1-19.

Miller, J.C. and Tanksley, S.D. 1990. Effect of different restriction enzymes, probe source, and probe length on detecting restriction fragment length polymorphism in tomato. Theor. Appl. Genet. 80:385-389.

Mitra, R.K. and Bhatia, C.R. 1986. Repeated DNA sequences and polyploidy in cereal crops. In: Dutta, S.K. (ed) DNA Systematics. CRC Press, Boca Raton, Fla.

Murray, M.G., Peters, D.L. and Thompson, W.F. 1981 Ancient repeated sequences in the pea and mung bean genomes and implications for genome evolution. J. Mol. Evol. 17:31-42.

Nelson, M. and McClelland, M. 1987. The effect of site-specific methylation on restriction-modification enzymes. Nucleic Acids Res. 15:219-230.

Ohri, D. and Khoshoo, T.N. 1986. Plant DNA: Contents and systematics. In: Dutta, S.K.(ed) DNA Systematics. CRC Press.

Pichersky, E. Bernatzky, R., Tanksley, S.D. and Cashmore, A. R. 1986. Evidence for selection as a mechanism in the concerted evolution of *Lycopersicon esculentum* (tomato) genes encoding the small subunit of ribulosse-1,5-bisphisphate carboxylase/oxygenase. Proc. Natl. Acad. Sci. U.S.A. 83:3880-3884.

Pichersky, E., Malik, V., Bernatzky, R., Hoffman, N., Tanksley, S.D.. and Cashmore, A.R. 1987. Tomato Cab-4 and Cab-5 genes encode a second type of CAB polypeptide localized in photosystem II. Plant Mol. Biol. 9:109-120.

Pichersky, E., Brock, T.G., Nguyen, D., Hoffman, N.E., Piechulla, B., Tanksley, S.D. and Green, B.R. 1989. A new member of the CAB gene family: structure, expression and chromosomal location of Cab-8, the tomato gene encoding the type III chlorophyll a/b-binding polypeptide of photosystem I. Plant Mol. Biol. 12:257-270.

Price, H.J. 1988. Plant genome size and the DNA C-value paradox. The Plant Genetics News Letter 4:18-24.

Pruitt, R.E. and Meyerowitz, E.M. 1986. Characterization of the genome of Arabidopsis thaliana. J. Mol. Biol. 187:169-183.

Rayburn, A.L. 1990. Genome size variation in Southwestern Indian maize adapted to various altitudes. Evol. Trends in Plants. 4:53-57.

Rayburn, A.L. and Auger, J.A. 1990. Genome size variation in Zea mays ssp. mays adapted to different altitudes. Theor. Appl. Genet. 79:470-474.

Rayburn A.L., Auger,J.A. Benzinger E.A. Hepburn A.G. 1989. Detection of intraspecific DNA content variation in Zea mays(L) by flow cytometry. J. Exp. Bot. 40:1179-1184.

SAIKI, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Konr, G.T., Mullis, K.B. and Erlich, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Sambrook, J., Fritsch, E.F. and Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press, New York.

Schweizer, G., Ganal, M., Ninnemann, H. and Hemleben, V. 1988. Species-specific DNA sequences for identification of somatic hybrids between Lycopersicon esculentum and Solanum acaeele. Theor. Appl. Genet. 75:679-684.

Sugita, M., Manzara, T., Pichersky, E., Cashmore, A. and Gruissem, W. 1987. Genomic organization, sequence analysis and expression of all five genes encoding the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from tomato. Mol. Gen. Genet. 209:247-256.

Suiter, K.A., Wendel, J.F. and Case, J.S. 1983. Linkage-1: a PASCAL computer program for the detection and analysis of genetic linkage. J. Hered. 74:203-204.

Tanksley, S.D. and Rick, C.M. 1980. Isozymic gene linkage map of the tomato: Applications in genetics and breeding. Theor. Appl. Genet. 57:161-170.

Tanksley, S.D. and Pichersky, E. 1988. Organization and evolution of sequences in the plant nuclear genome. In: Gottlieb, L.D.and Jain, S.K. (eds) Plant Evolutionary Biology. Chapman and Hall Ltd., London.

Tautz, D. and Renz, M. 1983. An optimized freeze sequence method for the recovery of DNA fragments from agarose gels. Anal. Biochem. 132:14-19.

Vallejos, C.E., Sakiyama, N.S. and Chase, C.D. 1992. A molecular marker-based linkage map of Phaseolus vulgaris L. Genetics 131:733-740.

Willimzig, M. 1985. LiCl-boiling method for plasmid mini-preps. Trends in Genetics 1:158.

Wilson, D.R. and Larkins, B.A. 1984. Zein gene organization in maize and related grasses. J. Mol. Evol. 20:330-340.

Wolter, F.P., Fritz, C.C., Willmitzer, L. and Schell, J. 1988. rbcS genes in Solanum tuberosum: conservation of transit peptide and exon shuffling during evolution. Proc. Natl. Acad. Sci. USA 85:846-850.

Zamir, D. and Tanksley, S.D. 1988. Tomato genome is comprised largely of fast evolving, low copy-number sequences. Mol. Gen. Genet. 213:254-261.

Zhang, H.B. and Dvorak, J. 1990. Isolation of repeated DNA sequences from Lophopydrum elangatum for detection of Lophopydrum chromatin in wheat genomes. Genome 33:283-293.

Zhao, X., Wu T., Xie, Y. and Wu, R. 1989. Genome-specific repetitive sequences in the genus Oryza. Theor. Appl. Genet. 78:201-209.