4543

CONSTRUCTION OF AN AFLP LINKAGE MAP AND ANALYSIS OF QTLS FOR ECONOMIC TRAITS IN PAPAYA (*CARICA PAPAYA* L.)

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI'I IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

HORTICULTURE

DECEMBER 2004

By

Rajeswari K. Srinivasan

Dissertation committee:
Richard Manshardt, Chairperson
Ray Ming
Robert Paull
Adelheid Kuehnle
Monto Kumagai

Acknowledgement

I am grateful to Dr. Richard Manshardt, for his guidance, encouragement and support throughout the Ph.D. program. I would like to express my appreciation to Dr. Ray Ming, Hawaii Agriculture Research Center (HARC), for proving me excellent lab facilities for carrying out AFLP, as well as for supporting me in statistical analysis. I would also like to extend my appreciation to my committee members, Dr. Robert Paull, Dr. Adelheid Kuehnle, and Dr. Monto Kumagai for their useful discussions and continuous support. I wish to thank Dr. Gernot Presting, for his constructive comments on my dissertation.

I would like to thank Dr.Clifford Morden and Dr. Sterling Keeley for allowing me to use their lab facilities in my earlier trials of AFLP. I would also like to thank Dr.Leo Cheung (Cancer Research Center), for his useful suggestions in statistical analysis. My appreciation goes to Dr. Hao Ma and Mike Melzer for their help in linkage mapping and cloning LCY-B gene, respectively. I like to acknowledge the people in Dr. Ray Ming's lab, Dr. Qing Yi, Denise Steiger, Heather Pearl, and Minna Kim for various assistance received during my experiments at HARC.

I like to thank Susan Magita and Mark Hanzawa, at UH agriculture experiment station at Poamoho, for good maintenance of the papaya trees.

The financial support given by USDA and East-West Center were greatly appreciated.

Last but not least, I like to thank my family for their unconditional support throughout my study.

ABSTRACT

A genetic linkage map for papaya was constructed using 513 AFLP markers, an isozyme marker, and a morphological marker. The map, consisted of 15 linkage groups that covered 2066 cM of the papaya genome. Two hundred and eighty one plants from a F_2 population derived from Kapoho X Saipan Red were used to create the map. Two qualitative traits (PGM and fruit-flesh color) as well as three quantitative traits (node number at floral conversion, carpellody, and fruit weight) were analyzed. The fruit-flesh color locus was located on LG-14 with a linked marker at 4.1 cM distance. Two noteworthy features (blocks of consecutive markers with dominant alleles arising from a single parent and unidirectional segregation distortion) were noticed in the mixed-phase linkage map. To create a reliable marker order for QTL analysis, two coupling-phase maps, one for each parent, were created. A single QTL explaining 64.3 % of the trait, node number at floral conversion, was identified on LG-4 by the interval mapping method. The AFLP markers identified for flesh color (M1527AKF) and node number at floral conversion (M2104EK5 and M0832AK2) can be used in marker assisted selection and map based cloning. More than 50% of the markers that showed significant association with carpellody in the nonparametric Kruskal-Wallis test were located on four linkage groups (LG-2, LG-3, LG-4 and LG-7). A block of markers on LG-7 showed significant association with carpellody in each of the two seasons studied. Although the broad sense heritability was high (74.0%) for mean fruit weight, only five markers were identified as having significant association with the trait by single marker analysis, and no QTL peak was identified by interval mapping. In another study, an AFLP marker that consistently co-segregates in coupling phase with the normal phenotype at the *diminutive*

mutation locus in papaya var. Sun Up, was detected using bulked segregant analysis. Lastly, a 281-bp fragment of lycopene β-cyclase (LCY-B) gene was isolated from papaya (Accession No. AY753202), which can be used as a probe in investigating the possibility of polymorphism in LCY-B gene as the molecular basis for fruit-flesh color difference in papaya.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page No.**

LIST OF FIGURES

Figure **Page No.**

LIST OF ABBREVIATIONS

AFLP	Amplified fragment length polymorphism
BSA	Bulked segregant analysis
BSA	Bovine serum albumin
BLAST	Basic local alignment search tool
CCS	Capsanthin capsorubin synthase
CTAB	Hexadecyl trimethyl ammonium bromide
CYC-B	Chromoplast specific lycopene β-cyclase
DFR	Dihydroflavonol 4-reductase
HPLC	High pressure liquid chromotography
LB	Luria-Bertani
LCY-B	Lycopene β-cyclase
LCY-E	Lycopene ε-cyclase
MPLC	Medium pressure liquid chromatography
NCBI	National center for biotechnology information
NXS	Neoxanthin synthase
PCR	Polymerase chain reaction
PGM	Phosphoglucomutase
QTL	Quantitative trait loci
RAPD	Random amplified polymorphic DNA
RFLP	Restriction fragment length polymorphism
RIL	Recombinant inbred line

SCAR	Sequence characterized amplification region
SNP	Single nucleotide polymorphism
SOC	Super optimal catabolite
SSR	Simple sequence repeat
STR	Short tandem repeats
VNTR	Variable number tandem repeats

CHAPTER 1

GENERAL INTRODUCTION

Papaya is a tropical fruit tree grown for its leaves, flowers, and ripe and unripe fruit. Ripe fruit is used as a dessert, while unripe fruit are used in cooking. Unripe fruits also produce papain, a proteolytic enzyme used in tenderizing meat. Papaya is the second most important fruit crop in Hawaii after pineapple. In recent years, the papaya industry has added $12-18 million to the Hawaiian economy. The economic importance necessitates constant improvement of papaya for yield and consumer preference. Papaya breeding has been going on in Hawaii since the early 20th century that started with the introduction of the famous smaller-sized, hermaphrodite solo from Barbados (Storey, 1969) and culminated with the development of genetically engineered, ringspot virus-resistant varieties 'SunUp' and 'Rainbow' (Manshardt, 1999).

To develop an improved variety, papaya breeders raise large populations of trees to study the inheritance of different traits and to select good trees. This requires large areas to grow the plants, and involves money for labor and other maintenance costs. Also, breeders spend considerable time in raising the tree to maturity, if they need to study the fruit characteristics. DNA markers help overcome these problems. Once a marker is developed for a particular trait, the population can be screened at the seedling stage for desirable segregants, and only selected plants need to be grown in the field. Thus DNA markers are a powerful tool for increasing the efficiency of papaya breeding. Breeders also benefit from knowledge about the linkage between different traits of interest, as it

helps in selection and mapping. Mapping with molecular markers will help improve understanding of the number of loci controlling traits of interest and their chromosomal location. Hence, development of a genetic linkage map for papaya would be very useful. A few maps have been constructed earlier (Hofmeyr, 1939; Sondur *et al.*, 1996 and Ma *et al.*, 2004), though not many traits of economic importance are included. Among different DNA marker technologies, amplified fragment length polymorphism (AFLP) is highly reproducible, an essential characteristic for developing an useful marker. Thus an AFLP linkage map would be very helpful to papaya breeders.

Many commercially important traits are not simply inherited. Economically important traits like flowering, fruiting and yield are controlled by multiple genes called quantitative trait loci (QTL). Analyzing quantitative trait loci (QTLs) will help us understand the genetics underlying complex traits.

Commercially important traits are early flowering and smaller fruit size. Carpellody is a fruit malformation in papaya, the occurrence of which makes the fruit unmarketable. Early flowering correlates to lower node number at flowering (Nakasone, 1955). Finding QTLs linked to carpellody, node number at flowering and fruit weight will be of immense value to the breeders in developing new varieties.

Mapping is necessary to identify markers linked to important QTLs, but it is not necessary or an efficient means to identify markers for simply inherited traits controlled by major genes. Bulked segregant analysis is a faster approach to find a marker closely

linked with specific regions of the genome that are linked with a single trait (Michelmore, 1991).

Usually AFLP markers are not robust, that is, markers identified from an AFLP linkage map are useful only in the population from which the map was produced. The more closely linked the marker is to the gene of interest, the more likely it is to be useful in other related populations. Therefore, it would be most generally useful if the marker were developed in the gene controlling the trait itself. An investigation was undertaken to develop markers for papaya, with the following objectives:

I. Use AFLP markers to create a linkage map of the papaya genome that includes major genes and QTLs that for some important economic traits.

II. Isolate a marker closely linked to the mutant '*diminutive*' locus using bulked segregant analysis.

III. Isolate a fragment of lycopene β-cyclase (LCY-B) gene from papaya that can be used as a probe to investigate the potential polymorphism in LCY-B as the molecular basis for flesh-color difference in papaya

CHAPTER 2

LINKAGE MAPPING AND QTL ANALYSIS FOR ECONOMIC TRAITS IN PAPAYA

2.1 Introduction

Papaya is the second most valuable fruit crop in Hawaii, after pineapple. The papaya industry adds approximately $12-18 million to the Hawaiian economy every year. This economic importance of papaya necessitates constant breeding work to improve cultivars. To develop new cultivars, breeders need to grow large populations to select plants with desirable traits. When fruit traits are studied in papaya, breeders will need to wait for at least a year from planting to evaluate the fruit. To expedite the selection process for desirable plants, phenotypic, enzymatic or genetic markers linked to the traits of interest need to be identified. Markers linked to major genes involved in the expression of simple as well as complex traits can be identified by using linkage maps. Such linked markers are required for marker assisted selection and would be expected to expedite a breeding program. Linkage map were created for papaya as early as in 1939 by Hofmeyr, the map covered only 41 cM. Using molecular marker technology, two additional maps have been reported for papaya (Sondur *et al.*, 1996; Ma *et al.*, 2004). The traits on the maps however do not include many commercially desirable traits. Developing a linkage map of these other traits would help us improve our knowledge of the papaya genetics.

While simple traits are controlled by a single gene, complex traits are controlled by many interacting loci, the effects of which are often influenced by the environment. QTL maps help us find the number of genetic loci involved in the expression of the quantitative trait, approximate chromosomal locations of these loci, and the relative contributions of each locus to trait expression (Stuber, 1992). QTL mapping has been used extensively in other crops to dissect complex traits like disease resistance (Taran *et al.*, 2003; Flandes-Galvez *et al.*, 2003), flowering (Fazio *et al.*, 2003; Li *et al.*, 2003) and fruit characteristics (Fazio *et al.*, 2003; Frary *et al.*, 2003; Knapp and Tanksley, 2003). In papaya, only one QTL map (Sondur, 1996) has been created so far. In the present investigation, a linkage map was created for papaya using AFLP markers, and QTL analyses were performed for three quantitative traits: node number at floral conversion, carpellody and fruit weight. In addition, an AFLP marker in linkage at 4.1 cM with fruit-flesh color was identified. A single QTL that is responsible for 64.3% of the phenotypic variability in node number at floral conversion was identified using the interval mapping method. Using the nonparametric Kruskal-Wallis test, four linkage groups, LG-2, 3, 4, and 6, were found to be associated with the expression of carpellody; a fruit malformation. In spite of having high broad-sense heritability (74%), no QTL was identified for fruit weight by the interval mapping method, suggesting that this trait is controlled by many genes with small and individual effects.

2.2 Literature Review

2.2.1 Taxonomy and Origin

Papaya (*Carica papaya* L) is a diploid, dicotyledonous tree belonging to the Caricaceae, a family which until recently included four other genera: *Cylicomorpha, Horovitzia, Jacaratia* and *Jarilla* (Badillo, 1971). According to a recent taxonomic revision (Badillo, 2000, 2001), the commercial papaya is the only species in the genus *Carica,* and all twenty other species formerly in *Carica* have now been assigned to the genus *Vasconcellea.* Papaya is native to the low-lands of Central America in the region between Southern Mexico and Nicaragua (Storey, 1969). The chromosome number of papaya is 2n = 18 (Storey, 1953). Even though the phenotypic variations between different papaya varieties are high, genetic diversity is low (Stiles *et al.*, 1993 ; Kim *et al.*, 2002).

2.2.2 Flower types

Papaya is a polygamous plant with three different classes of flowers: male (staminate); female (pistillate) and hermaphrodite (bisexual). Typically, a hermaphrodite flower has five petals, 10 stamens, and a single ovary. Stamens are arranged in two whorls one set opposite to the petals and another set alternate with the petals. The ovary has the carpels with parietal placentation. Petals are fused to form a corolla tube. In the male flower, the ovary is reduced to a pistillode and the flower size is reduced. In female flowers, the stamens are missing and the petals are separate. Hermaphrodite flowers produce fruits, that are commercially preferred.

Hermaphrodite trees are usually andromonoecious, which produce solitary flowers or short cyme inflorescences with a few flowers, consisting mainly of fertile bisexual flowers. These hermaphrodite trees also produce a few staminate flowers at basal nodes of the cyme.

Some cultivars are dioceous, consisting of male and female trees, while some are gynodioceous consisting of female and hermaphrodite trees. Though the basic tree types of papaya are male, female and hermaphrodite, a number of intermediate forms are observed. A number of researchers classified papaya in different ways based on flower types and sex. While Higgins and Holt (1914) classified papaya trees into 13 different classes based on flower types, Hofmeyr (1938) reported nine tree categories. Later, Storey (1941) classified papaya into five main categories based on flower types. Again, in 1958, Storey reclassified papaya into 32 types with 15 forms of male, 15 forms of hermaphrodite and two forms of female. Storey elaborated on the flower types in his paper "Theory of the derivation of the unisexual flowers of Caricaceae" (1967). The five principal flower types reported by Storey (1941 and 1967), based on the sex forms and the shape of the fruit that they produce, are:

Type I. Pistillate or female flower: This is an extreme of the flower types lacking stamens. The petals are fused inconspicuously at the base, appearing as if they are free. The pistil is five carpellate and is circular or undulate in cross-section. The fruit are spherical to ovoid in shape with five distinct petal scars at the base. Storey (1967) hypothesized that the female flowers arose from the hermaphrodite flowers by loss of all

ancestral gynoecium elements except the ventral carpel bundle system, and replacement of the rest of the carpel structure by 5 transmuted ancestral stamens.

Type II. Pentandria bisexual flowers: This type of flower falls between type I and type III flowers. Pentandria flowers were earlier reported by Higgins and Holt (1914) and Hofmeyr (1938). In this type of flower, all five inner stamens are carpellodic. The remaining five stamens alternate with the petals. The basal parts of the petals are fused among themselves and also with the pistil. Stamen filaments are fused with the petals and pistil at the base of the corolla tube. The globular or slightly oblong fruit, which arises from the partly inferior ovary, has five deep longitudinal grooves. A glossy disc is present at the base of the fruit, surrounded by a continuous ring, marking the petal scar. Some of the pentandria flowers produce pistils with 10 stigmatic rays. The fruit has a central locule containing ovules, surrounded symmetrically by five ovule-filled locules. These types of flowers are called multi-loculate pentandria.

Type III. Intermediate or carpellodic bisexual flower: This is the common carpellodic type with visibly distorted flowers. Stamens, varying in number from two to ten, show different degrees of fusion to the pistil and the petals (Storey, 1967). One or more of the stamens from the inner set of five stamens, and more rarely those from the outer set, may turn out to be carpelloid. These stamens are usually carpel-like in structure, showing different degrees of stamen-carpel transition. The carpels, varying in number from five to 10, are often incompletely fused and the pistil is distorted. When a carpellodic stamen is completely joined with the pistil, it forms a rounded lobe on the pistil, enclosing a locule with ovules. In the regions of the flower where carpellodic stamens and their adnations have occurred, the petals are irregular in shape and the corolla tube is shortened. The ten

stamens in the normal flowers are reduced by the number of carpellodic stamens. The fruit is misshapened due to adnation of carpelloid structures to the ovary. Often, anther-like tissues are found at the distal end of the fruit.

Type IV. Elongata or bisexual flower: This is the most preferred flower type, producing symmetrical elongate fruits. In this flower type, the petals are fused obscurely with the pistil at the base forming an unbroken, circular petal scar. The flowers have 10 stamens, arranged in two whorls, an inner set opposite to the petals and sub-sessile, while the outer set alternates with the petals and have short filaments. In some genotypes, anthers and stigmas are in contact at anthesis, while in other genotypes, anthers lie below the stigma. Usually the ovary has five carpels, though the number of carpels may vary from 1 to 7 or 8. The fruit develops from a superior ovary.

Type IV+. During certain seasons, type IV flowers do not produce a functional pistil, bearing only a vestigial ovary without a stigma. These barren hermaphrodite flowers appear structurally similar to type IV flowers, and are slightly smaller in size than type IV flowers.

Type V. Staminate or male flower: This is the other extreme flower type, characterized by the presence of functional stamens and the absence of a functional pistil. Staminate flowers arose from hermaphrodite flowers by the loss of the gynoecium and retention of the androecium. These flowers have long, slender corolla tubes. The number and arrangement of the petals and the stamens are the same as those of type IV flowers. These flowers have a rudimentary pistil called pistillode.

2.2.2.1 Effect of environment on carpellody expression

The fate of a flower to be elongata or carpellodic is determined during stamen and ovary differentiation. Arkle and Nakasone (1984) reported that stamen differentiation begins around eight weeks before anthesis and is completed in three weeks. Ovary differentiation begins around seven weeks prior to anthesis and is completed in three weeks. The degree of stamen adnation to the carpels is related to the time at which stamen transmutation is initiated. Cold temperatures seven to eight weeks before the anthesis interfere with the normal development of stamens. Similarly, warm temperatures six to eight weeks before anthesis inhibit the normal development of ovaries. If the conducive conditions for transmutation of stamen to carpel prevail in the early stage, before carpels begin to form, then the stamens will join completely with the carpels. If stamens transform into carpels when the original carpels are already in the process of development and fusion, the transformed carpels fuse incompletely or show partial super-imposed adnation to the original carpels (Storey, 1958). Awada and Ikeda (1957) investigated the environmental forces that stimulate carpellody and reported that higher irrigation increased the nitrogen absorption, which in turn increased plant growth in terms of elongation of trunk and rate of emergence of new leaves, leading to significant increase in the production of carpellodic fruits. Later, Awada (1958) confirmed the significant correlation between carpellody and the rate of increase in stem circumference. Also, Awada's study revealed that when the daily minimum temperature is below 21°C, the rate of carpellodic flower production increased, and at 16°C and below, almost all the flowers produced are carpellodic in nature. Storey (1958) classified hermaphrodite papaya trees into five categories based on the season of the carpellody expression:

a) trees producing symmetrical elongata flowers year round, b) trees producing carpellodic flowers only during the spring, c) trees producing carpellodic flowers during the winter and the spring, d) trees producing normal flowers only during the summer and e) trees producing carpellodic flowers throughout the year. Storey (1958) reported that two sets of genetic factors, one that affects female sterility (carpel abortion) and another that affects transmutation of stamens to carpels, determines carpellody.

2.2.2.2 Carpel abortion

In contrast to carpellody, some hermaphrodite plant, under certain environmental conditions, produce flower that is intermediate between elongata and staminate flowers. During high temperatures, hermaphrodite tree produce a range of flowers with fewer than five carpels by individual carpel abortion (Nakasone and Lamoureux, 1982). The number of carpels ranges from two to four. The flower with reduced number of carpels produces fruit of different size and shape than normal elongata fruit. Fruit is long and cucumber-shaped. Sometimes, differential development of carpels is noticed. The fruit is curved when one or two carpels are poorly developed. As an extreme of the reduced carpel nature, female sterile flowers (Type IV) without any functional ovary or stigma are also produced in papaya. Usually, the axillary flowers tend to be female sterile. Female sterility is epistatic to stamen carpellody. Low nitrogen and moisture stress lead to high female sterility (Awada and Ikeda, 1957). The fruit is cylindrical, long and cucumber-shaped with the number of carpels ranging from one to four.

2.2.3 Papaya genetics

2.2.3.1 Sex

Papaya is a polygamous flowering plant, producing male, female and hermaphrodite sexes. Hofmeyr (1938) and Storey (1938) individually reported that a single gene with three alleles controls the sex expression in papaya. The dominant alleles *M1* and *M2* are responsible for male and hermaphrodite development, respectively and the recessive allele *m* is responsible for female development. They also reported that all combinations of dominants, *M1M1*, *M2M2* and *M1M2* are lethal. The genotype *M1m* is male, *M2m* is hermaphrodite and *mm* is female. In 1941, Hofmeyr proposed a genetic balance hypothesis. He assigned an arbitrary value of 47 to the autosomal chromosome complement and values of 21, 22 and 25 to *M1, M2* and *m* respectively to the sex chromosomes. The hermaphrodite genotype *M2m* deviates from the autosomal value of 47 by zero. Female (*mm*) and male (*M1m*) deviate from the autosomal value by three and one respectively. The closer the value is to a threshold value, the more unstable the genotype is in the sex expression. So, male is more unstable, and influenced by modifying factors, and female is stable. Later, in 1953, Storey proposed that sex expression is influenced by a set of genes: Mp-male pendulate, l-zygotic lethal, Sa-suppressor of the androecium, Sg-suppressor of the gynoecium and c-hypothetical factor for suppression of crossing over. Recently, Liu *et al.* (2004) have physically mapped the sex chromosome in papaya. Male and hermaphrodite chromosomes share a nearly identical male-specific Y region that has diverged substantially from the same region in the female due to suppression of crossing-over. The findings suggest that female plants evolved from hermaphrodite by a male sterility mutation, and non-recombinant plants

were selected naturally. Later, male plants evolved from hermaphrodite by a separate mutation for female sterility.

2.2.3.2 Node number at flowering

Earlier flowering and low bearing height are commercially preferred characteristics in papaya. The height at which the first flower is produced by the tree is determined by two independent sets of multiple factors: length of internode and number of nodes from cotyledons to the first flowering node (Nakasone, 1955). Usually, the earliest flowering plants have produced the fewest nodes at first flowering. Nakasone (1955) found that correlation of height of the first flower with the number of nodes to flowering is higher than correlation of the height of first flower with internodal length. Subadrabandhu and Nontaswatsri (1997) reported that the gene action for height of first flower is largely additive.

2.2.3.3 Flesh color

Papaya has two different fruit-flesh colors: yellow and red. Hofmeyr (1938) reported that flesh color is controlled by a single gene. Yellow flesh color is dominant over red flesh color. Yamamoto (1964) extracted carotenoid pigments from the fruit-flesh with acetone and methanol and analyzed carotenoid composition using HPLC (High pressure liquid chromatography). He reported that ζ-carotene and cryptoxanthin are abundant carotenoids in both red and yellow-fleshed fruits. In addition, red-fleshed papaya had high amounts of lycopene, which was completely absent in yellow-fleshed

fruits. Recently, Chandrika (2003) analyzed carotenoids in red and yellow-fleshed fruits using MPLC (medium pressure liquid chromatography), and her study confirmed that red-fleshed fruits have lycopene, which is absent in yellow-fleshed fruits. Storey (1969) reported that although red, and yellow are basic fruit-flesh colors, each color has a range of shades due to modifying genes acting upon their expression.

2.2.3.4 Fruit weight

Different varieties of papaya show a wide range of fruit weight. Storey (1953) reported that inheritance of fruit weight in papaya is additive in nature. Fruits from F_1 hybrid plants show weight close to the mean of fruit weights of the parents, and F_2 plants show a range of fruit weights in between the parental fruit weights. A diallel study by Subadrabandhu and Nontaswatsri (1997) confirmed that the gene action for fruit weight in papaya is additive.

2.2.3.5 Phosphoglucomutase (PGM)

Phosphoglucomutase is a monomeric enzyme in the cirtic acid cycle, having two electrophoretic zones of activity, *Pgm*-1 and *Pgm-2,* for papaya. The alleles at the *Pgm-1* locus in papaya are inherited in Mendelian fashion as co-dominant markers (Morshidi, 1998).

2.2.4 Markers

Markers are the mutations in the genome that can be mapped to specific locations on the chromosomes. The mutations can be observed at the morphological level, like altered leaf phenotype, or observed at the DNA level by different techniques like restriction fragment length polymorphism (RFLP) amplified fragment length polymorphism (AFLP).

2.2.4.1 Morphological markers

Morphological markers are the visible changes in phenotype that are associated with changes in functional gene expression. Usually, morphological markers follow simple Mendelian inheritance, and each morphological marker is associated with a single gene (Liu, 1998). Some common morphological markers include dwarf plants and pigmentation variants. Generally, morphological marker mutations are recessive in nature.

Tanksley (1983) distinguished morphological markers and molecular markers as follows: a) Phenotypes of the morphological markers can be distinguished only at the whole plant level, while molecular markers can be found at the whole plant, tissue and cellular level. b) The number of available morphological markers in nature is small compared with the number of alleles detected at the molecular level. c) Most morphological marker phenotypes are recessive in nature, often associated with deleterious and undesirable effects, whereas most molecular markers do not carry any

deleterious effects on the plant. d) The mode of gene action for morphological markers is dominant or recessive while some molecular markers are co-dominant in nature, enabling distinction between homozygotes and heterozygotes, thus making them useful in any segregating generation. e) Morphological markers may be accompanied by strong epistatic effects, whereas epistatic effects are rarely associated with molecular markers.

2.2.4.2 Protein markers

2.2.4.2.1 Isozymes

Isozymes are markers that make use of the principle that different allelic variants of the same enzyme can show different electrophoretic movement in a starch gel. Isozymes often represent 'allozymes', different forms of enzymes produced by different alleles of the same locus (Prakash *et al.,* 1969). After electrophoresis, the starch gel is soaked in a staining solution that utilizes the specific enzymatic activity to mark the location of the isozyme banding in the gel by imparting a color change at that site. Protein mobility differences among isozymes are caused by changes in protein size or charge. Protein sizes differ because of glycosylation, phosphorylation or amino acid substitution caused by point mutation. Isozymes are co-dominant markers that can be unequivocally identified (Tanksley and Rick, 1982), but are restricted to coding sequences of the genome. Isozyme markers have been used in plant breeding in many ways, including estimation of out-crossing rates, introgression of genes from wild species, measurement of genetic variability, varietal patenting and protection, and gene

tagging. However, the number of useful isozymes in nature is low. It is impossible to saturate the genome with isozyme markers (Tanksley, 1983).

2.2.4.3 DNA markers

2.2.4.3.1 Restriction fragment length polymorphism (RFLP)

RFLP is one of the earliest DNA marker systems, used originally in adenovirus (Grodzicker *et al.*, 1974). Botstein (1980) used this method to create a human genetic linkage map, and it has subsequently been applied in many different organisms. In this method, genomic DNA is first cut with restriction enzymes, and the restricted fragments are separated on an agarose gel. The DNA is transferred from an agarose gel onto a nylon or nitrocellulose membrane by southern blotting and visualized by hybridization with a labeled probe. Hybridization probes can be constructed from cDNA or from genomic DNA. Probes can be radioactively-labeled (Botstein, 1980) or chemically labeled. Polymorphisms in RFLP marker system arise whenever there is a mutation in the restriction site. Point mutation leads to the loss of a restriction site or creation of a new one. Deletion or insertion of a segment of DNA within a fragment will alter the size. These genotypic changes lead to altered electrophoretic mobility of restricted fragments on an agarose gel (Botstein, 1980).

2.2.4.3.2 Rapid Amplified Polymorphic DNA (RAPD)

RAPD is a PCR- based marker system developed independently by Williams (1990) and Welsh and McCleland (1990). In this method, the genomic DNA is amplified

with 9- or 10-bp single primers of arbitrary nucleotide sequence. RAPD markers work based on the principle that the probability of finding the same priming sites on complementary strands is low. The criteria for RAPD primers are that they must contain 40% or more G+C content and have a minimum annealing temperature of 36°C. When the PCR reaction is complete, the amplified fragments are separated by agarose gel electrophoresis, stained with ethidium bromide and visualized under ultra-violet light. Polymorphism results whenever there is a deletion of a priming site or when insertions or chromosomal rearrangements make the priming sites too distant to support amplification. Sometimes RAPD detects even single base pair mutations at the priming sites. These randomly amplified DNA fragments are inherited in Mendelian fashion and are dominant or recessive in nature. Whenever there is an insertion or deletion point mutation in the amplified region between two inbred lines, the resulting mobility changes is visualized as co-dominant markers. RAPD markers are useful in creating linkage maps. RAPD markers can be cloned and used as RFLP probe, or can be sequenced and used as Sequence Characterized Amplified Region (SCAR) markers. In papaya, Sondur (1996) has created a genetic linkage map using RAPD markers.

2.2.4.3.3 AFLP

Based on selective PCR amplification of restriction fragments generated from total genomic DNA, Vos *et al.*(1995) found another type of DNA marker, called amplified fragment length polymorphism (AFLP). Restricted fragments for amplification are generated by two restriction enzymes, a rare cutter with a 6-bp recognition site and a

frequent cutter with a 4-bp recognition site. Double-stranded oligonucleotide adapters are ligated to the ends of the DNA fragments to generate priming sites for PCR amplification. The adapter sequence, together with the adjacent restriction site, serves as a primer binding site during the amplification. DNA amplification is performed in two steps. Adding selective nucleotides on the 3' end of the AFLP primers reduces the number of amplified bands four-fold with each additional selective base. During selective amplification, restricted fragments with the rare cutter palindrome on one end and frequent cutter palindrome on the other end are the predominant class amplified. This technique uses stringent amplification conditions. When touch-down PCR (Don *et al.*, 1991) is used during selective amplification, the number of spurious bands is reduced significantly. Selective amplification of AFLP is carried out in two steps. In the first step, 12 cycles of PCR are performed starting with an annealing temperature of 65°C in the first cycle. In the subsequent cycles, annealing temperature is reduced by 0.7°C every cycle, leading to an annealing temperature of 55°C at the last PCR cycle. Any difference in annealing temperature between genuine and spurious annealing will give two-fold advantage per 0.7°C (per cycle) to the genuine product. If the genuine band has annealing temperature 5.6°C higher than that of the spurious band, the genuine band has 2^8-fold advantage over the spurious band in amplification.

During selective amplification, the primers can be labeled with either radioactive isotopes, silver stain (Vos *et al.*, 1995), or fluorescent dyes (Myburg, 2001). After selective amplification, the fragments are separated electrophoretically in denaturing polyacrylamide gel. DNA bands are visualized autoradiographically or on a gel

fluorescence scanner. The sources of polymorphism in AFLP are mutations in the restriction site or in the selective extension nucleotide binding region, or chromosomal rearrangements. A polymorphism in mobility of the band is also observed between two inbred genotypes if there is a deletion or insertion point mutation inside the amplified fragment in one genotype. The AFLP band of genotypes with insertion mutation moves slightly slower than that of the other genotype. When both bands can be detected on a single gel, the polymorphism is considered as co-dominant marker. This method combines the reliability of RFLP with the power of PCR. A limited set of primers can be used on a wide variety of species. AFLP markers are reliable and repeatable (Pejic *et al.*, 1998) and the method can be used for the identification of markers linked to specific traits. This identification is carried out by bulked segregant analysis (Xu *et al.*, 2000; Jia *et al.*, 2001), for fingerprinting varietal genotypes (Kim *et al.*, 2002), for genetic mapping (Kang *et al.*, 2001; Lehmensiek *et al.*, 2001; Buerstmayr *et al.*, 2002) and other plant breeding objectives. Recently, Ma *et al.* (2004) have created a high density map of the sex-determining region in papaya using AFLP markers. In sorghum, integrated genetic and physical maps have been created using AFLP markers (Klein *et al.*, 2000).

2.2.4.3.4 Mini- and micro-satellites

Jeffreys *et al.* (1985) reported that mini-satellites, which are tandem repetitive sequences of 10-100 bp, can be used as probes to detect polymorphic loci among different individuals. These markers are also called variable number tandem repeats (VNTR) by Nakamura (1987). Later, Litt and Luty (1989) and Weber and May (1989)

showed independently that polymorphisms also exist among individuals for tandem repeats of shorter sequences (1-4 bp) called microsatellites. Microsatellites are also called simple sequence repeats or SSR (Jacob *et al.*, 1991) and short tandem repeats or STRs (Edwards *et al.*, 1991). To detect VNTRs, genomic DNA is digested with restriction enzymes and the resulting fragments are probed with tandem repeats. A single probe can hybridize to many loci, giving a DNA fingerprint for an individual (Jeffreys *et al.*, 1985). Alternatively, after probing the restricted DNA with tandem repeats, a single hybridized band can be isolated and sequenced. Later on, Hanotte *et al.*, 1991 demonstrated that the repeated core sequences, along with the flanking region, could be used as locus-specific probes. By using the flanking regions of the repeated sequences as primers, SSRs can be amplified and fragment length polymorphisms detected (Litt and Luty, 1989). Two approaches can be used to identify SSR loci (Lavi *et al.*, 1995): a) identification of microsatellites in sequence databases and the generation of primers to the conserved flanking region of the microsatellites; b) isolation of positive clones from a genomic library, after hybridization to a microsatellite probe and sequencing of these clones to find the flanking regions of the SSRs. In both mini- and microsatellites, polymorphism results from allelic differences in the number of core sequence repeats, which arise by mitotic or meiotic unequal crossing-over, sister chromatid exchange or by DNA slippage during replication (Jeffreys *et al.*, 1985, and Litt and Luty, 1989).

2.2.4.4 Selection of marker system

The type of marker system selected for a particular study depends on the objective, nature of the population, genetic diversity of the plants under study and available time and money (Staub *et al.*, 1996). Different marker systems are advantageous under different situations. RFLP and SSR are co-dominant markers usable across populations, but marker development takes time. RFLP needs high amount of starting DNA and involves radioactive labeling of the probe. Mini- and microsatellites are suitable for automated analysis, but are typically limited to interspecific applications. RAPD markers are easy to produce, require very small amount of DNA and no prior knowledge about the sequence, and a set of universal primers can be used for a wide variety of species. However, RAPD markers are very sensitive to the PCR reaction conditions and require verification to be reliable. Like RFLPs, RAPDs are useful only for the particular cross used to produce these marker (Staub *et al.*, 1996). Further, as RAPD and AFLP markers are usually dominant or recessive, it is not possible to identify whether an amplified fragment comes from a homozygous or heterozygous plant. AFLP requires small quantity of DNA (250 ng) as it is PCR based, uses arbitrary primers, and produces some co-dominant markers. Among RFLP, RAPD, AFLP and SSR, AFLP reveals the maximum number of polymorphisms in a single assay (Staub *et al.*, 1996; Russel *et al.*, 1997). This reduces the number of assays that are needed to generate required polymorphism. Further, AFLP, including marker scoring can be automated and multiplexed (Myburg, 2001). AFLP is the most efficient to generate maximum number of markers in shorter time for mapping analyses.

2.2.5 Linkage maps

Linkage maps are developed by ordering molecular markers and/or genes on a chromosome and expressing the relative distances of one marker from the others based on recombination frequency between the markers. Linkage maps act as a bridge between conventional plant breeding and molecular biology by giving essential information for a) easy identification and incorporation of useful genes into cultivars through marker assisted selection in plant breeding programs, b) a better understanding of complex inherited traits (QTL) important in crop improvement, c) positional cloning of useful genes, d) finding orthologous genes through comparative mapping of related and unrelated taxa (Lee, 1995).

Generally, the following steps are involved in creating linkage maps: a) creating a segregating population and scoring it for both phenotypic and genotypic data, b) identification of linkage groups using maximum likelihood methods, c) ordering of markers within the linkage groups by three point and multipoint analyses.

2.2.5.1 Mapping population

The basic assumption of using marker loci to detect polygenes is that linkage disequilibrium exists relative to the alleles of the linked polygenes. Linkage disequilibrium is defined as the non-random association of alleles at different loci in the population. The primary cause of linkage disequilibrium in the segregating populations derived by controlled matings is physical linkage of different loci (Tanksley, 1993).

Segregating populations are usually generated from two parental lines that are homozygous for different alleles at loci which affect variation in the trait of interest. The parental alleles are then shuffled by creating a large segregating mapping population for mapping, in which the phenotype and the genotype of each individual are measured (Mauricio, 2001).

Different types of segregating populations can be used for linkage mapping (Liu, 1998). In the back-cross design, the mapping population is derived by crossing F_1 progeny to the parent that is homozygous for the recessive allele of the trait under analysis. In F_2 intercross, F_1 plants are self-pollinated to generate the mapping population. If a co-dominant marker system is used, F_2 populations give double the amount of information, compared to backcross (Lander and Botstein, 1989). Backcross and F_2 population can be reproduced only if the plant is clonally propagated.

This problem can be overcome in species, in which haploid plants can be produced by wide hybridization, anther culture, or ovary culture. Generating diploids from these haploids is accomplished by colchicine treatment (Raina, 1997).

Recombination estimation is the most accurate in this doubled haploid method because all alleles are in homozygous condition. Doubled haploid plants can be selfed to get genetically identical plants which helps to evaluate the effect of environment on the trait.

Another kind of immortal population is composed of recombinant inbred lines (RIL), which are produced by selfing F_2 through F_8 generation using single seed descent method. Each recombinant inbred line (RIL) plant is homozygous and represents one F_2 plant (Burr *et al.*, 1988). Long-generation heterozygous perennials, like forest trees, self-incompatible, or clonally propagated plants, can also employ F_2-like strategies or BC-like strategies for creating mapping population (Muranty, 1996). The natural populations are characterized by studying segregation of genotypic frequencies and the phenotypic frequencies. To create pseudo F_2 population, plants that are phenotypically different are selected as parents. In pseudo backcross, parents suggesting heterozygous at the loci that control the trait under investigation are crossed.

2.2.5.1.1 Selective genotyping

Genotyping is determining allelic composition of individual plants at marker loci. Generally, due to time and money constraints, it is not possible to genotype all the plants in a population, and a subset of plants is chosen randomly for genotyping. For a genotyping population of a given size, the resolving power of the map can be increased by selecting plants for genotyping from the phenotypic distributional extremes (Lander and Botstein, 1989). In a normally distributed backcross population, phenotypes more than one standard deviation from the mean comprise about 33% of the population but contribute 81% of the total linkage information. Similarly, progenies with phenotypes more than two standard deviation from the mean comprise about 5% of the total population, but contribute about 28% of the total linkage information. If the phenotypic

extremes of a population have significantly different frequencies of positive and negative alleles at a particular marker, then a QTL controlling the phenotype is likely to be linked to the marker (Tanksley, 1993). Thus selective genotyping is an efficient way to decrease the number of plants that are needed for genotyping. The disadvantages of selective genotyping are: a) in order to find the plants at the distributional extremes, large populations still need to be scored for the phenotypic traits; b) generally, it is not possible to analyze more than one trait at a time because it is difficult to find plants that show extreme phenotypes for all the traits under study; c) individuals at the distributional extremes tend to accumulate either positive or negative alleles for all the QTLs associated with a particular trait, and thus the mapping population is deficient in plants that have a mixture of positive and negative alleles at different QTLs. This latter complication will lead to identification of false positives in single marker analysis and inflation of QTL effects in interval mapping.

2.2.5.2 Identification of marker segregation pattern

The segregating population is scored for genetic markers, and the marker allele frequencies are checked to determine whether they follow the expected segregation pattern for the particular population. In a backcross, dominant markers as well as co-dominant markers are expected to segregate in 1:1 ratio. In an F_2 population, dominant markers are expected to segregate in 3:1 ratio and co-dominant markers are expected to segregate in 1:2:1 ratio. Chi-square test is used to identify markers that show segregation distortion in the population.

2.2.5.3 Identification of linkage groups

Linkage groups are identified by estimating pair-wise recombination for all the markers by two-point analysis, and any linked markers are assigned to the same linkage group (Doerge, 2002). Linkage groups are created based on a given minimum LOD score and maximum recombination frequency allowed between a pair of markers. Maximum likelihood is the widely accepted statistic for finding the recombination frequency between two markers. In this approach, all possible values of recombination frequency (θ) are tried in the following binomial model to find the value of θ for which the observed segregation data maximize the likelihood function.

$$L(\theta/NR) = \theta^{R}(1-\theta)^{N-R}$$

where L = Likelihood

N = Total number of observed data

R = Number of recombinants

θ = Recombination frequency

The significance of linkage between a pair of markers is tested by LOD score which is the $\log_{10}$ of likelihood ratio.

$$\text{Likelihood ratio} = \frac{\text{Maximum likelihood } (\theta = \text{value that maximizes L})}{\text{Likelihood at } \theta = 50}$$

2.2.5.4 Ordering of markers within the linkage groups

Within a linkage group, the markers are ordered based on three-point analysis and multi-point analysis. The order of markers which gives minimum map length is

considered to be the best order (Ott, 1991). The precision of the order of markers within the linkage group is decreased whenever dominant markers in repulsion phase are encountered in a F_2 population.

2.2.5.5 Mapping dominant markers in F_2 population

Dominant markers give incomplete information in heterozygotes, which leads to increased estimates of recombination rate (Sybenga, 1996; Mester *et al.*, 2003a). The presence of repulsion phase markers in F_2 population does not affect grouping of markers into linkage groups, but it does affect the ordering of markers (Knapp *et al.*, 1995; Mester *et al.*, 2003a). When dominant markers are used in F_2 population, the maximum likelihood estimator is biased (Knapp *et al.*, 1995). When the population has no double recessive phenotype, and the frequency of double dominant phenotypes is less than two thirds, the recombination frequency estimates become zero irrespective of the number of recombinants observed.

The maximum likelihood estimator of recombination frequency, θ is given by,

$$\hat{\theta}_{ML} = 2^{\frac{1}{2}}\left\{2\hat{f}_1 - \hat{f}_2 - \hat{f}_3 - 1 + \left[8\hat{f}_4 + \left(1 - 2\hat{f}_1 + \hat{f}_2 + \hat{f}_3\right)^2\right]^{\frac{1}{2}}\right\}^{\frac{1}{2}}$$

where $\hat{f}_1, \hat{f}_2, \hat{f}_3, and\ \hat{f}_4$, are observed frequencies of phenotypes A-B-, A-bb, aaB- and aabb respectively.

when $\hat{f}_4 = 0, \hat{\theta}_{ML} = 2^{\frac{1}{2}}\left\{\left(3\hat{f}_1 - 2\right) + \left[\left(2 - 3\hat{f}_1\right)^2\right]^{\frac{1}{2}}\right\}^{\frac{1}{2}}$

when $\hat{f}_4 = 0, and \quad \hat{f}_1 \leq \frac{2}{3}, \hat{\theta}_{ML} = 0$

when $\hat{f}_4 = 0, and \hat{f}_1 > \frac{2}{3}, \hat{\theta}_{ML} = \left(3\hat{f}_1 - 2\right)^{\frac{1}{2}}$, where recombination frequency can be under-estimated or over-estimated.

The probability of having no double recessive plants in the population decreases as population size increases (Knapp *et al.*, 1995; Mester *et al.*, 2003a). To increase the precision of ordering, dominant markers in F_2 population can be divided into two groups, the coupling-phase markers and the co-dominant markers (Knapp *et al.*, 1995). Complementary coupling-phase maps using AFLP markers in F_2 population have been created in a few crops including wheat (Peng *et al.*, 2000), *Medicago* (Thoquat, 2002) and *Mimulus* (Fishman *et al.,* 2001).

2.2.5.5.1 Mapping function

The distance between two ordered markers on a linkage group is measured by the percentage of recombination that occurs between them during meiosis. The recombination fraction is estimated by the ratio of recombinants to total number of offspring and is expressed as map units (centiMorgans or cM). One cM is defined as the distance over which one percent crossing over is observed. When three or more adjacent loci are concerned, recombination in one region may give physical hindrance to

recombination in the adjacent region by the phenomenon called interference. Interference is maximum in closely linked (<10 cM) markers. If there is no interference, for three consecutive loci A, B and C, the distance between A and C is defined as

$$r_{12} = r_1 + r_2 - 2r_1r_2$$

where, r_1 = recombination fraction between AB

r_2 = recombination fraction between BC

r_1r_2 = Double cross-over frequency

If there is complete interference,

$$r_{12} = r_1 + r_2$$

If interference is less than complete, whenever a new locus is found in between two existing loci, the distance between them should be reduced. This problem can be overcome by converting recombination frequency into map distance by a mapping function. Map distances found using a mapping function, are fixed even when a new locus is found. Two common mapping functions are Haldane (1931) and Kosambi (1944). Haldane mapping function assumes that there is no interference, while Kosambi takes some interference into account.

2.2.5.5.2 Relationship between map distance and physical map length

The relationship between the genetic map distance and physical map length is not the same throughout the genome, as recombination rate is not uniform throughout the genome. Some regions are relatively inert in crossing over. In a chromosome, crossing over occurs less frequently in heterochromatic regions than in euchromatic regions

(Roberts, 1965). Recombination rate is suppressed in the centromere region, which leads to a compression of the physical distance when estimated by genetic means and results in an apparent clustering of markers. Thus, the actual number of base pairs represented by 1 cM in the centromere will be higher than that represented by 1 cM in other parts of the chromosome. In tomato, the physical distance represented by 1 cM ranges from 50 Kb to >4000 Kb (Paterson, 1996).

2.2.6 QTL mapping

A QTL is a region of a chromosome that contains a single gene or a group of genes having a significant effect on a quantitative trait (Tanksley, 1993). QTLs are defined by the flanking markers. QTLs are identified by checking the probability of linkage between the trait and individual markers by single marker analysis or by scanning the ordered linkage groups for the presence of QTLs associated with quantitative traits by interval mapping or by composite interval mapping.

The principles behind finding linkage between QTL and markers are: a) if the genome coverage of the map is large, the chances are good that some of the markers and some of the QTLs controlling the traits of interest will be linked, and b) from a segregating population, linkage relationships can be found by studying the marker segregation pattern and corresponding trait variation (Liu, 1998). The possibility of identifying minor QTLs depends on the distance between marker and the QTL, size of

segregating population, heritability of the trait, and threshold level used for declaring a QTL (Tanksley, 1993).

2.2.6.1 Single marker analysis

In 1923, Sax showed that in the bean *Phaseolus vulgaris*, seed size (a polygenic trait) and seed pigmentation (a monogenic trait) are associated with each other. Later, Thoday (1961) showed that location of genes on the chromosomes can be determined based on the markers. Single marker analysis, in which each marker is tested one at a time for linkage with the trait of interest, is the simplest method of detecting QTL. For single marker analysis, there is no need to create linkage maps. Simple statistical tests (like t-tests, ANOVA, simple linear regression, likelihood ratio test and maximum likelihood ratio) test are used to determine the significance of the linkage between quantitative trait phenotypes and the marker genotypes. Often, the linear regression model is used for single marker analysis. For a backcross progeny, the model is

$$Yj = \beta o + \beta_1 X_j + \varepsilon_j$$

where Y_j = dependent variable or phenotypic value for the jth individual in the population

X_j = Independent variable or marker genotype value

βo = Intercept for the regression or population mean for the trait

β_1 = Slope of the regression line or regression co-efficient

ε_j = Random error for the j^{th} individual

For F_2 population, the model is

$$Y_j = \beta o + \beta_1 X_{1j} + \beta_2 X_{2j} + \varepsilon_j$$

where, Y_j = Dependent variable or trait value for the jth individual in the population

X_{1j} = Variable for the marker additive effect

X_{2j} = Variable for the marker dominance effect

ε_j = Random error for the j^{th} individual

β_1 = Slope for the additive regression line

β_2 = Slope for the dominant regression line

Usually, the null hypothesis tested in these statistical analyses is that $\beta 1 = 0$ or the mean of the trait value does not depend on the genotype at a particular marker. The null hypothesis is tested using a F statistic, which is the ratio between the regression mean square and residual mean square. When the test statistic is larger than a crucial value, the null hypothesis is rejected and a QTL is declared (Doerge, 2002). The main disadvantages of single point analysis are: a) the further a QTL is from the marker gene, the less likely it is to be detected statistically due to crossover events between the marker and the QTL; b) the magnitude of the effect of any detected QTL will normally be underestimated due to recombination between the marker locus and the QTL (Tanksley 1993); c) the position and the effect of the QTL are confounded (Doerge, 2001); d) the location of QTL on the chromosome cannot be found unless a linkage map is created (Mauricio, 2001).

2.2.6.2 Interval analysis

In 1989, Lander and Botstein introduced a method called interval analysis, in which a genetic map is generated first, followed by detection of QTL location. In this method, a QTL is located in between a pair of flanking markers. Ordered pairs of markers on the linkage groups are scanned for the presence of a QTL at small intervals (example, 2 cM). Interval maps can be created using linear regression, non-linear regression, or maximum likelihood. In single marker analysis by maximum likelihood method, the relationship between the phenotype ϕ_j and genotype g_j for the j^{th} individual is described by the equation,

$$\phi j = a + bg_j + \varepsilon$$

where, g_j = number of alternate alleles in the genotype

ε = random normal variable with mean 0 and genotypic variance σ^2

a, b and genotypic variance σ^2 = unknown parameters

The maximum likelihood analysis estimates the values of unknown parameters L ($\hat{a}$, $\hat{b}$, $\hat{\sigma}^2$), which maximize the probability that the observed phenotype would have occurred. In interval mapping, the QTL genotype g_j is unknown and the probability of g_j being a particular genotype is calculated using the flanking markers.

The likelihood map plots LOD score against the chromosomal map position, measured in recombination units (Mauricio, 2001). LOD scores shows the likelihood with which QTL exists in a particular location in the genome. Whenever LOD score exceeds a threshold value determined by the genome size and marker density in the map,

a QTL associated with the trait is declared. Usually, a LOD score between 2 and 3 is required as a threshold to keep the false positive rate below 5% (Lander and Botstein, 1989).

The advantages of the interval mapping method are that it can be performed at any position covered by markers on the map and that the approximate location of the QTL can be found by the support interval, if there is only one QTL on the chromosome. The support interval is given by the region on the linkage group covered by the width of the peak, when the likelihood drops 10 fold (a LOD drop of 1.0). Though interval analysis is statistically more stringent than single marker analysis, QTLs that exist on the same LG, but outside the interval, can influence the estimated position and effect of QTL within the interval (Tanksley, 1993). Further, interval mapping uses only one pair of markers at a time, thus ignoring information from other markers. Sometimes "ghost" QTLs, which are spurious QTLs that arise from a co-segregation between QTL and non-adjacent or distant markers, are identified (Knapp *et al.*, 1995; Zeng *et al.*, 1993; Doerge and Churchill, 1996; Mauricio, 2001).

2.2.6.3 Composite Interval mapping

This mapping procedure increases the resolution of QTL locations by using markers other than those adjacent to the target QTL to decrease the noise level. This is an improved interval mapping procedure developed by Jansen (1993) and Zeng (1993, 1994). This method combines interval mapping and multiple regression to avoid ghost

QTLs. The precision of the interval mapping is increased by separating multiple QTL linked on the same chromosome during the statistical test. Because markers flanking other QTLs in the same linkage group are used as co-factors, the effects of QTLs in intervals other than the one under study are eliminated. Composite interval mapping is useful for highly heritable traits because most of the residual variation can be removed from genetic variation.

2.2.6.4 Threshold value

Under the null hypothesis that there is no relationship between genotype and the phenotype, the value of the tested statistics should exceed the threshold, if a significant association exists between the phenotype and the QTL, with the probability not to exceed the minimum level of $\alpha = 0.05$ (Doerge and Churchill, 1996). In type I error, a QTL is declared when no significant effect is present (false positive). In type II error, the test fails to detect a QTL when one is present (false negative). Estimating an appropriate significance threshold value is necessary to avoid these errors. The threshold values suggested by Lander and Botstein (1989) and Lander and Kruglyak (1995) for interval mapping had assumed that the map is densely covered with markers and the quantitative trait phenotype follows a normal distribution. Churchill and Doerge (1994) suggested a permutation-based method to estimate the threshold value. The phenotypic data are permuted with respect to marker data many times to effectively sample randomly from the distribution of the test statistic under a null hypothesis that no QTL exists. In this method, the phenotypic data of *n* individuals are shuffled 1000 times over the marker

data to generate a sample that has only random association between the phenotype and their genotype. An empirical threshold value for declaring a major QTL (α=0.05) is calculated. This method does not require that the phenotype follow a normal distribution. This method is also applicable in smaller sample conditions.

2.2.6.5 Non-parametric methods of QTL analysis

Most of the traditional QTL mapping methods including a) single marker analysis by ANOVA, t-test, and linear regression b) interval mapping method using maximum likelihood statistic or c) composite interval mapping using multiple regression assume that the phenotypic data follow normal distribution (Kruglyak and Lander, 1995). However, many phenotypes in nature are not normally distributed and often a proper transformation method to convert the data to normal distribution is not available. In this situation, application of a nonparametric or distribution free method will be statistically valid. Kruglyak and Lander (1995) suggested the nonparametric Wilcoxon rank-sum statistic for analyzing QTL in backcross populations. This method tests for the presence of the QTL, but does not estimate the QTL effect. This method has been incorporated in the software Mapmaker\QTL v 1.9. For F_2 populations, nonparametric QTL analysis is done by the Kruskal-Wallis rank sum test (Lehmann, 1975; Van Ooijen *et al.*, 1993). The Kruskal-Wallis test can be used even when the phenotypic data is ordinal and even when the marker segregation is distorted. By this method, the F_2 plants are first ranked based on their phenotypic values or index. Regardless of the marker genotypes, plants with a same phenotypic index (called a tie) are given the same rank called "mid-rank".

Mid-rank equals the average of ranks of plants in the tie. Then, the mean rank of plants in each genotype is calculated. If a QTL is linked to a particular marker, the genotypic classes will show large difference in their mean rank values. A test statistic based on this difference is used to test the presence of a QTL. The null-hypothesis used in this test assumes no difference in the frequency distribution between different genotypic classes, and the statistic is chi-square distributed with genotypic classes, minus one degrees of freedom. The softwares MAPQTL, and R/QTL allows Kruskal-Wallis test for F_2 population.

2.2.7 Linkage and QTL maps in papaya

As early as in 1939, Hofmeyr reported a linkage map for papaya with the traits sex type, flower color and petiole color. The association of the traits was,

M---$_{25cM}$----Y--$_{16cM}$---P

where, M = sex locus, Y = flower color, and P = petiole color. But this linkage map covered only 41cM and did not include many economically important traits. More than 50 years later, after the discovery of DNA markers, Sondur (1996) developed a more detailed map in papaya using 62 RAPD markers in an F_2 population developed from Sunrise X UH356. Traits like plant sex, plant vigor, node at which first flowering occurs, and stem diameter were mapped using the software MAPMAKER/EXP 3.0. Their resulting linkage map was 999.3 cM long with 11 linkage groups. The locus that determines plant sex type, SEX1, flanked by markers 7 cM away, was located in linkage group 1. QTLs controlling plant vigor, node at which first flowering occurs and stem

diameter were also found on the first linkage group at LOD of 3.0. He also identified 3 QTLs for carpellody, two on linkage group 5 and one on linkage group7. He had also mapped three QTLs each for fruit number and mean fruit weight, and two QTLs for number of fruits per node.

Recently, Ma *et al.* (2004) created a linkage map for papaya with 1498 AFLP markers, the papaya ringspot virus coat-protein marker, morphological sex types and flesh color using 54 F_2 plants derived from the parents Kapoho and SunUp. The map consisted of 12 linkage groups which varied in length from 86.2 to 865.2 cM. The total length of the map was 3294.2 cM with an average distance of 2.2 cM between the adjacent markers at a LOD score of 5.0 and recombination frequency 0.25. The papaya sex locus was located on linkage group LG-1 with 227 co-segregating AFLP markers. The papaya ring spot virus coat-protein locus was mapped to linkage group 7, with flanking markers at a distance of 3.7 and 4.7 cM. Fruit-flesh color was also mapped to linkage group 7 with flanking markers 3.4 and 3.7 cM away.

2.3 Materials and methods

2.3.1 Mapping population

The population used for mapping was a F_2 generation derived from the parents Kapoho and Saipan Red, that have contrasting characteristics (Table 2.1). Kapoho is a commercial variety from Hawaii, and Saipan Red is a breeding line obtained from Saipan in the Northern Mariana Islands and maintained at the University of Hawaii. The parents

Table 2.1 Traits of the parents used in the mapping population

Traits	**Kapoho**	**Saipan Red**
Bearing height	High	Low
Flesh color	Yellow	Red
Carpellody	Absent	Moderately carpellodic
PGM locus	Homozygous for fast-moving allele	Homozygous for slow-moving allele

were grown at Dole Food company's experimental plots near Poamoho, Oahu during 1997-1999. F_1 plants, grown at the University of Hawaii experimental station, Poamoho, Oahu in 1999 were self-pollinated to obtain F_2 seeds. The F_2 seeds were treated with GA_{4+7} (100μM) and matriconditioning material Micro-Cel E, as recommended by Andreoli and Khan (1993) to improve the germination. The seedlings were transplanted at the UH Agriculture Experiment Station at Poamoho in March 2001. All 281 F_2 plants were scored for sex, flesh color, node number at floral conversion, carpellody and PGM isozyme variation.

2.3.2 Phenotypic observation

2.3.2.1 Node number at floral conversion

The trees were scored for node number at floral conversion by counting the number of nodes from the cotyledon to the node, at which the axillary meristem number changed from one (vegetative) to two (vegetative and reproductive).

2.3.2.2 Carpellody

Trees in the F_2 population were rated for carpellody during February 2002, on a scale from 0-3 with **0** being no carpellody, **1** being 0 to 33% carpellody, **2** being 34 to 67% carpellody and **3** being 68 to 100% carpellody.

The rating was repeated in August 2002 on a more detailed scale ranging from 0-3. The scores and corresponding percentage of affected fruits were: **0**-0% carpellody,

0.5-0 to16.5%, **1**-16.5 to 33.0%, **1.5**-33.0 to 49.5%, **2**-49.5 to 67.0%, **2.5**-67.0 to 83.5%, and **3**-83.5 to 100% respectively.

2.3.2.3 Fruit weight

Mean fruit weight of each tree was found by calculating the average weight of 2-5 fruits per tree. For seven trees fruit weight was available for only one fruit. Mean fruit weight was calculated only for the plants in the genotyping population (n=84).

2.3.2.4 Flesh color

Fruit-flesh color was scored as red or yellow in the F_2 population.

2.3.2.4 Isozyme

Isozyme variation at the *Pgm-1* locus was determined by starch gel electrophoresis as described by Lebot *et al.* (1991). Leaf extracts were prepared using Bosquet's buffer (Bousquet *et al.,* 1987) with pH 7.5, which contained Tris, 0.1 M; sucrose, 0.2 M; ethylenediaminetetraacetic acid (disodium EDTA), 0.5 mM; dithiothreitol (DTT), 5 mM; cysteine HCL, 12 mM; ascorbic acid, 25 mM; sodium metabisulfite, 0.02 M; diethyl dithio carbamic acid (DIECA sodium salt), 0.005M; bovine serum albumin (BSA), 0.1%; polyethylene glycol, MW 20,000 (PEG), 1%; polyoxyethylenesorbitan monolaurate (Tween-80), 2%; dimethyl sulfoxide (DMSO), 10%; β-mercapto ethanol, 1%; polyvinylpolypyrrolidone (PVPP), 8 g/100 ml of buffer. Leaf extracts were run on a starch gel. Histidine citrate was used as both gel buffer and

tray buffer. Gel buffer contained 0.016 M histidine (free base) and 0.002 M citric acid (anhydrous). Tray buffer contained 0.065 M histidine and 0.007 M citric acid. Both gel and tray buffers were adjusted to pH 6.5. Electrophoresis was carried out at 4°C, 15V/cm and 40-50 mA for six hours. After electrophoresis, the gels were sliced and stained for phosphoglucomutase (PGM) as per Soltis *et al.* (1983) and the distance that PGM had moved was scored.

2.3.2.5 Testing normality of distributions

Normality of the phenotypic data distributions were tested using SAS v.8 . (SAS Institute, 1995; 1999) A frequency histogram showing the distribution of the observed data against the expected normal distribution was produced. Statistical tests for normality, Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling were also performed. The null hypothesis (H_0) used in the test is that the observed data follow normal distribution. Skewness and Kurtosis values were also obtained. For normal distribution, values of both Skewness and kurtosis should be zero.

2.3.3 Genotyping

For genotyping, 84 plants that were in the phenotypic distributional extreme of the quantitative trait stamen carpellody were chosen. We also deliberately included a few plants in the distributional extreme of node number at floral conversion.

2.3.3.1 DNA extraction

2.3.3.1.1 DNA extraction of the parents

DNA of the parents was extracted by CTAB buffer and cleaned by CsCl as described by Saghai-Maroof *et al.*, (1984) with the modification suggested by Doyle and Doyle (1987). Leaf tissue (0.5-1.5 g) was ground with 7.5 ml CTAB buffer and then incubated at 60-65°C for 15 min. CTAB buffer contained 100 mM Tris-HCL (pH 8.0), 1.4 M NaCl, 20 mM EDTA, 2% CTAB, 0.2% β-mercaptoethanol, and 2% CTAB. The ground leaf tissue was extracted once with 24:1 chloroform: isoamyl alcohol. The aqueous phase was extracted once with isopropanol, spun in the centrifuge and the DNA was precipitated as a pellet. DNA was washed with wash buffer and resuspended in 4.0 ml of TE (Tris-HCl 10mM, pH 8.0 and EDTA 1mM). Cesium chloride and ethidium bromide were added to the DNA and the mixture was spun at 55K rpm overnight. The DNA band was pulled and ethidium bromide was removed by water-saturated butane extractions. Cesium chloride was removed by dialysis. DNA was washed with 70% ethanol and finally dissolved in 0.5ml TE.

2.3.3.1.2 DNA extraction of the mapping population

Leaves of the F_2 plants were collected and dried with silica gel. DNA was extracted from 20 mg of dried leaf with DNeasy plant mini kit from Qiagen Inc (Valencia, CA, USA) as per the instructions of the manufacturer.

2.3.3.1.3 DNA quantification

The DNA was subjected to electrophoresis on 1.2% agarose gel, along with uncut Lambda DNA 10, 20, 40, 60, 80, and 100 ng/µl. The gel was stained with ethidium bromide and the DNA was visualized with a UV illuminator. DNA was quantified by comparing the intensity of the band with that of the standards.

2.3.3.2 AFLP

2.3.3.2.1 Restriction and ligation

The restriction and ligation reactions of DNA were performed using the AFLP analysis system II-core reagent kit from Invitrogen Inc. (Carlsbad, CA). All reactions were performed as per the instructions of the manufacturer, except that the reactions were carried out in half of the suggested volume. All the reagents were also reduced by half in volume. After ligation of the adapter, the samples were diluted 1:10 with TE (Tris 10 mM and EDTA 0.1 mM).

2.3.3.2.2 Pre-amplification

Pre-amplification was performed using single nucleotide, extended standard EcoRI and MseI primers. The samples were amplified with E+A and M+C, E+G and M+C, E +A and M+G, or E+A and M+A. The PCR reactions were carried out in 20 µl volume with 5 µl of diluted DNA, 0.3 µm of EcoRI primer, 0.3 µm of MseI primer, 0.2 µm of dNTPs, 1.5 mM $MgCl_2$, 1X Mg-free PCR buffer and 1U Taq polymerase (Promega Inc., Madison, WI). The thermocycling profile used for pre-amplification consisted of 20

cycles with denaturation at 94°C for 30 sec, annealing at 56°C for 60 sec, extension at 72°C for 60sec and storage temperature at 4°C. After pre-amplification, the samples were diluted 20 times with TE.

2.3.3.2.3 Selective amplification

During the selective amplification, multiplexed PCR reactions were performed using standard EcoRI and MseI primers with two or three extension nucleotides. Two µl of the diluted pre-amplification products were selectively amplified in 10 µl volume. The reaction mixture consisted of 1X Mg-free PCR buffer, 0.1 mM dNTP, 1.5 mM MgCl2, 0.4 µM of MseI primer, 5 ng of EcoRI Infra Red (IR) dye-labeled primer that fluoresces Taq polymerase (Promega Inc., Madison, WI). The cycling parameters were as follows: one cycle of 94°C for 30 s; 65°C for 30 s; 72°C for 60 s, followed by 12 more cycles with one cycle of 94°C for 30 s; 65°C for 30 s; 72°C for 60 s, followed by 12 more cycles with 0.7°C decrease in annealing temperature per cycle. This was followed by 23 cycles of 94°C for 30 s; 56°C for 30 s; 72°C for 60 s with soak temperature of 4°C. At the end of PCR, 10 µl of sequence loading buffer (98% formamide, 10 mM EDTA and 0.025% bromophenol blue) were added to each sample, which were then denatured at 90°C for three minutes and placed on ice. The samples were run on polyacrylamide gel in Li-Cor IR^2 automated sequencer (Li-Cor Inc., Lincoln, NE). Fifteen ml of 10% ammonium per sulfate and 1.5 ml of TEMED were added to 20 ml of 6.5% polyacrylamide gel matrix from Li-Cor and the mixture was used to cast the gel, which was 25 cm long with 0.25 mm spacers and a 48 well comb. The polyacrylamide gel was reused twice, thus making a single gel useable for three times. In each run, one lane was loaded with a mixture

Table 2.2 Adapter sequences used in AFLP

EcoRI-Forward	5'- CTCGTAGACTGCGTACC-3'
EcoRI-Reverse	5'-AATTGGTACGCAGTCTAC-5'
MseI-Forward	5'-GACGATGAGTCCTGAG-3'
MseI-Reverse	5'-TACTCAGGACTCAT-3'

of 50-700 bp sizing standard labeled with IR 700 dye and 50-700 bp sizing standard labeled with IR-800 dye. The gel was pre-run for 15 min with the electrophoretic electrophoretic options chosen as follows: 1500V, 40 mA, 40 W and 50°C. After pre-run, 1.0 µl of the sample was loaded in the gel and the run was carried out for three hours at 1500 V, 35 mA, 35 W and 46°C. Fragments ranging in size from 50-700 bp were obtained as real-time images. The images were visualized through the software 'e-seq', provided with the Li-Cor IR^2 sequencer, and scored manually. The markers were checked to determine whether they followed Mendelian segregation ratios of 3:1, 1:2:1 or 2:1 using chi-square test at P=0.05.

2.3.3.2.4 Marker nomenclature

Each marker name has 8 characters. First character represents the size of the marker (small: < 80bp, medium: 80-350 bp or large >350 bp). Next two characters are codes used for EcoRI primer and fourth and fifth characters are codes used for MseI primers. Sixth character represents the order by which the marker was identified in the gel for a particular primer pair. Seventh character represents which parent contributes the dominant form of the marker (K-Kapoho; S-Saipan Red; C-co-dominant). Eighth character represents the marker segregation pattern (2-2:1 ratio; 3-3:1 ratio; 5-both 2:1 and 3:1 ratio; C-1:2:1; 0-neither of these ratios).

2.3.4 Linkage mapping

2.3.4.1 Construction of the linkage map

The linkage map was created with 513 AFLP markers and flesh color and PGM using the software Mapmaker/Exp 3.0 (Lander *et al.*, 1987; Lincoln *et al.*, 1992a). Though only 84 F_2 plants were genotyped, phenotypic data of all 281 plants were recorded for QTL mapping, and missing genotypic values were noted for ungenotyped plants. Mapmaker used the maximum likelihood test statistic for creating the map.

The markers were divided into different linkage groups at LOD 5.0 and recombination frequency 0.25. For ordering markers within a linkage group, seed order with five markers was found by using the command "order" at LOD 2.0. The rest of the markers were added by the command "build". The threshold for addition of markers was decreased gradually to LOD score 0.01. Kosambi mapping function (Kosambi, 1944) was used, and the map distances were given in centiMorgan (cM). The linkage groups were named LG-1 to LG-15, based on the length of the linkage group.

QTL analysis was performed using two coupling-phase complementary linkage maps. The markers were divided into two groups, each containing co-dominant markers and dominant markers from one parent. One map was created for dominant markers from Kapoho and another map was created for dominant markers from Saipan Red. In coupling-phase maps, except in LG-4 and LG-5, markers that did not fall into 3:1 or 1:2:1 were removed. Because marker distortion in LG-4 and LG-5 is caused by a biological mechanism rather than by chance, distorted markers in this linkage group were included

Table 2.3 Codes used for AFLP-EcoRI primers

EcoRI primers							
20 E-AA	21 E-AG	02 E-AAG	01 E-AAC	08 E-AGG	07 E-AGC	09 E-AAT	10 E-AGT
13 E-AGA	14 ATA	15 ATG	03 ACA	11 ATC	04 ACC	12 ATT	05 ACG
06 ACT	18 GG	19 GT	17 GC	16 GA	34 GTA	35 GTC	36 GTG
37 GTT	30 GGA	31 GGC					

Table 2.4 Codes used for AFLP-MseI primers

MseI primers							
01 M-CAA	02 M-CAC	03 M-CAG	04 M-CAT	05 M-CTA	08 M-CTT	06 M-CTC	07 M-CTG
10 M-CC	11 M-CG	09 M-CA	12 M-CT	30 M-GC	31 M-GG	29 M-GA	32 M-GT
26 M-AC	27 M-AG	25 M-AA	28 M-AT				

in coupling-phase maps. Faint markers were also removed from the coupling-phase map. The markers were divided into groups by the same thresholds used in the mixed-phase map (LOD 5.0 and recombination frequency 0.25). Coupling phase linkage groups were named KLG or SLG followed by the number relative to the mixed-phase map. For example, KLG-1 referred to the Kapoho coupling-phase linkage group with markers from LG-1 of the mixed-phase map. Within each linkage group the markers were ordered using the command 'order' and 'build'. The first five markers were obtained using the command 'sorder' at LOD 2.0, and the rest of the markers were added using 'build' command with lowest threshold of LOD 2.0. In both mixed- and coupling-phase maps, the final order of the markers was tested using the "ripple" command on a McIntosh computer. In this command, for each set of three consecutive loci, all six permutations of the markers are tested while keeping the order of the remaining loci fixed. In coupling-phase linkage maps, in the linkage groups KLG-5b, SLG-4a and SLG-15, the order of markers did not meet LOD 2.0. However, since the number of markers in these groups was not more than five, the best order was found by using the command 'compare', which analyses all possible combinations of the markers and chooses the best order.

2.3.4.2 Significance of non-random distribution of dominant markers

In some linkage groups, blocks of consecutive markers came from the same parent. Assuming a random binomial model, the probability of observing such blocks by chance was calculated as,

$$P=1-(1-0.5^{x+1})^{n+1}$$

where, x = Number of markers in the block

n = Number of markers in the entire linkage group (Burnside, 1959).

To check whether the block of contiguous markers with dominant alleles from one parent is an artifact created in the process of ordering of markers in Mapmaker, two linkage groups, one with such a block (LG-6) and another without such a block (LG-5) were selected. To see whether the pattern of markers was consistent, markers were ordered 5 times for each linkage group, and LOD 2.0 was used as threshold for ordering.

2.3.5 QTL mapping

2.3.5.1 Single marker analysis

2.3.5.1.1 Node number at floral conversion and fruit weight

Single marker analysis was carried out by the linear regression method using the software WinQTL cartographer v 2.0 (Basten *et al.* 2001). Whenever the calculated F statistic was significant at P = 0.005 or less, the marker was considered to be linked with the trait.

2.3.5.1.2 Carpellody

Since the phenotypic distribution of carpellody was clearly non-normal, the nonparametric Kruskal-Wallis rank test was used to analyze whether each individual marker is linked to a QTL influencing the trait. The data was analyzed in the software MAPQTL v 5.0 (Van Ooijen, 2004). Whenever the test statistic K was significant at P = 0.005 or less, the marker was considered to be linked to the trait.

2.3.5.2 Interval mapping

An interval map was constructed using the software Mapmaker\QTL v 1.0 (Lincoln *et al.*, 1992b). This software uses maximum likelihood test statistics to find the presence of a QTL. The whole genome was scanned every 2 cM for the presence of a QTL. A QTL map has the linkage map in cM on the X-axis and LOD score value on the Y- axis. The LOD score shows the log-likelihood for the presence of the QTL in a particular location. Whenever the LOD score exceeds the threshold value, determined by 1000 permutations, which is equivalent to P=0.05 (Churchill and Doerge, 1994) a QTL was declared. The possible locations of the QTL are given by the support interval. The support interval for the QTL was given by the region on the chromosome covered by the width of the peak, when the likelihood drops 10 fold (a LOD drop of 1.0).The regions showing the QTL peaks were again scanned for additive, dominant or recessive gene effects.

2.4 Results

2.4.1 Phenotypic observations

2.4.1.1 Quantitative traits

2.4.1.1.1 Node number at floral conversion

The number of nodes at floral conversion ranged from 22 to 49 over 280 plants in the F_2 population with a mean of 31.7 (SD = 4.2). In the genotyping population, number of nodes ranged from 24 to 49 with a mean of 32.07 (SD = 5.1). The phenotypic distribution in the F_2 and genotyped population are given in Fig. 2.1 and Fig. 2.2.

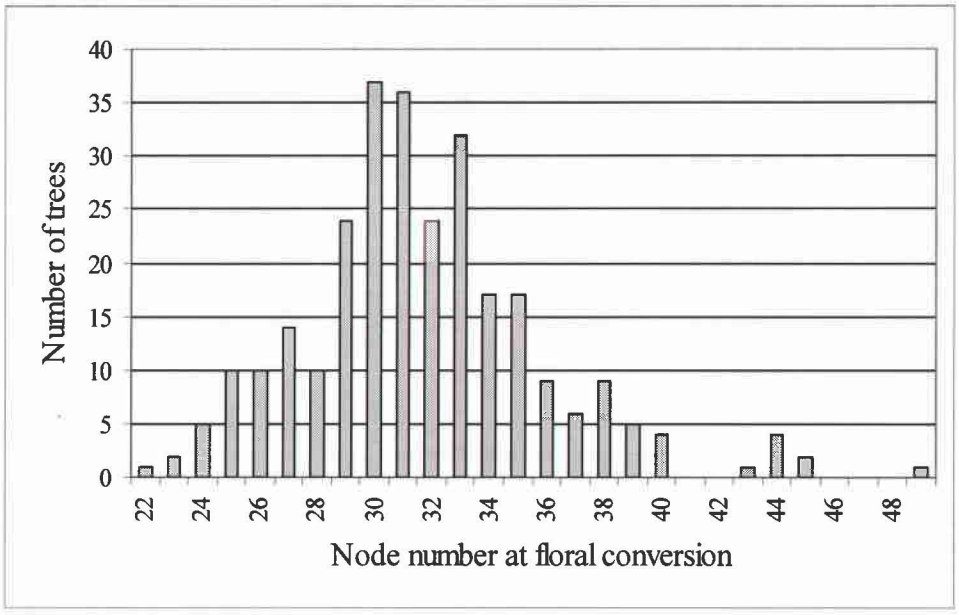


Fig. 2.1 The frequency distribution of node number at floral conversion in Kapoho X Saipan Red F_2 population (n = 280).

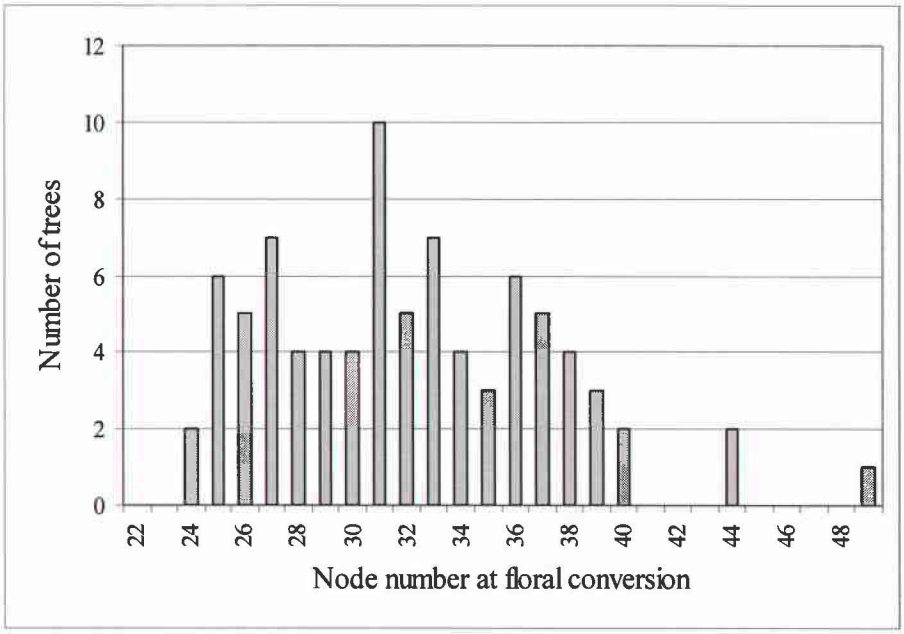


Fig. 2.2 The frequency distribution of node number at floral conversion in sub-F_2 population of Kapoho X Saipan Red, used for genotyping (n = 84).

respectively. The graphical (Fig. 2.3 and Fig. 2.4) as well as statistical tests of normality (Table 2.5 and 2.6) showed that at P=0.05, phenotypic distribution of F_2 deviates from normality, while the genotyped population was normal.

2.4.1.1.2 Carpellody

For the 214 hermaphrodite F_2 plants that were rated for carpellody, the mean was 0.9 (SD = 0.6) during Feb '02 and 1.05 (SD = 0.8) during Aug '02. For the genotyped population, the mean were 1.31 (SD = 0.7) and 1.95 (SD = 0.8) for Feb '02 and Aug '02 respectively. The frequency distribution of carpellody over two seasons in the F_2 population and genotyped population are given in Fig. 2.5 and 2.6, respectively. Since carpellody was scored on an ordinal scale, a test of normality could not be done.The phenotypic distribution graphs (Fig. 2.5 and 2.6) showed clearly that the distributions were not normal.

2.4.1.1.3 Fruit weight

In the genotyped population, average fruit weight for individual trees ranged from 365.5g to 1573.0g with a mean of 734.7 (SD = 216.8). The frequency distribution of mean fruit weight in the genotyped population is given in Fig. 2.7. The distribution was normal by graphical method (Fig. 2.8), as well as by statistical tests of normality at P=0.05 (Table 2.7).

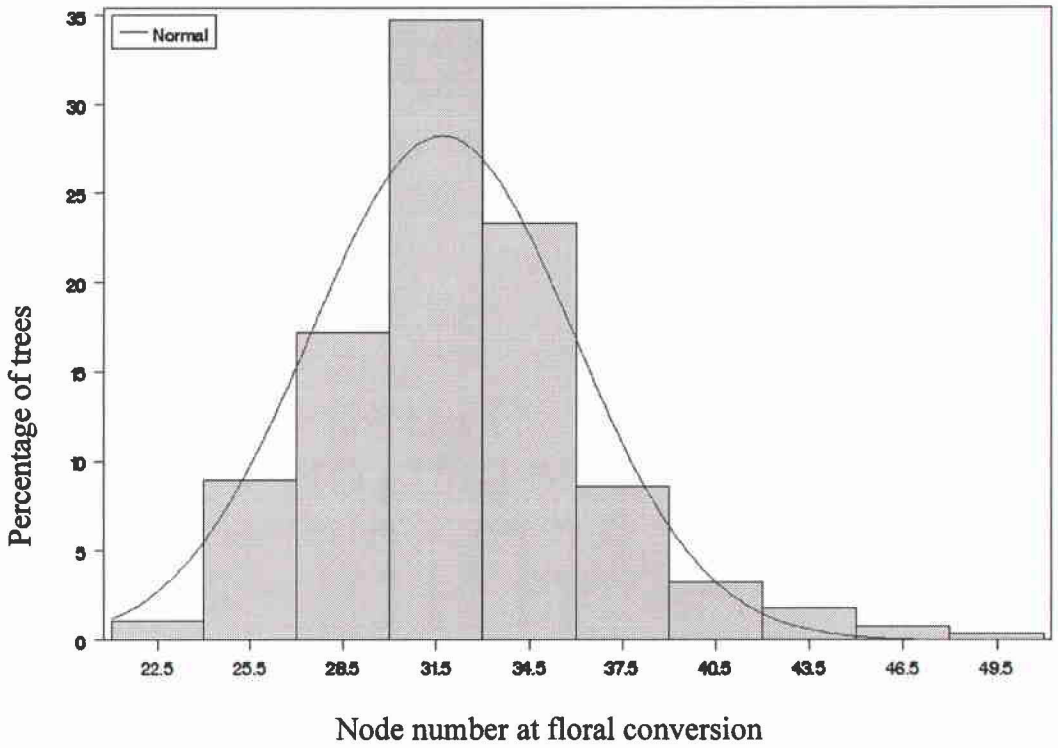


Fig. 2.3 Expected and observed phenotypic data distribution for node number at floral conversion in Kapoho X Saipan Red F_2 population (n=280).

Table 2.5 Results of tests of normality for node number at floral conversion in Kapoho X Saipan Red F_2 population.

Test	Probability
Kolmogorov-Smirnov	P=<0.010
Cramer-von Mises	P=<0.005
Anderson-Darling	P=<0.005
Skewness	0.754
Kurtosis	1.478

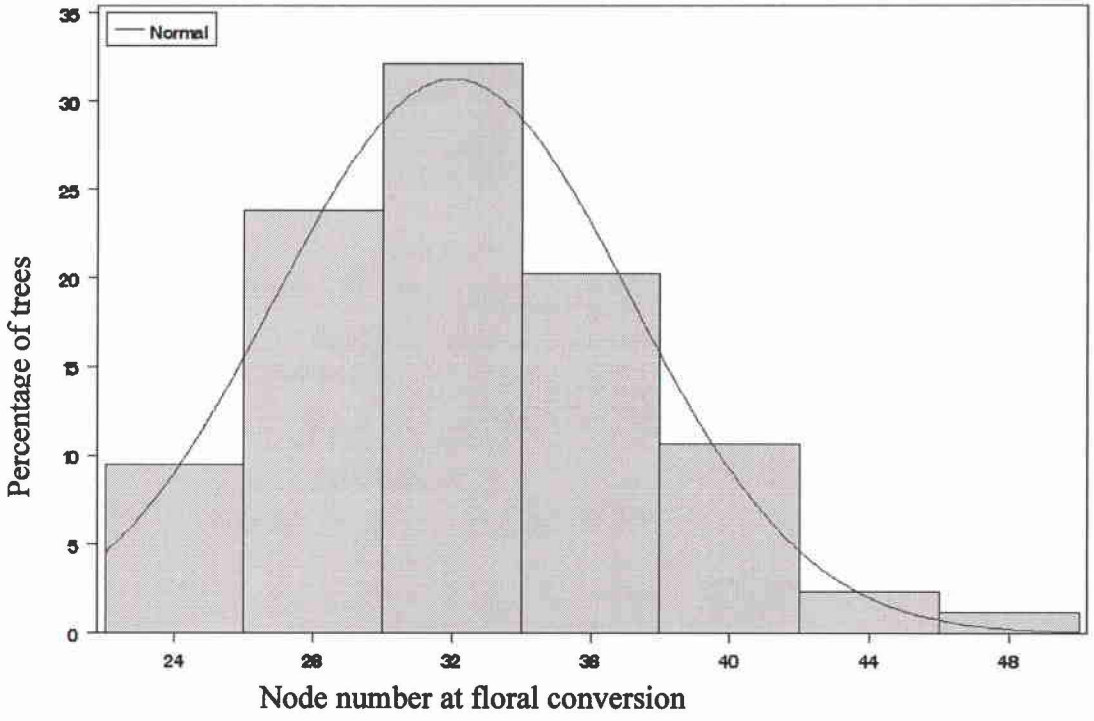


Fig. 2.4 Expected and observed phenotypic data distribution for node number at floral conversion in sub-F_2 population of Kapoho X Saipan Red, used for genotyping (n=84).

Table 2.6 Results of tests of normality for node number at floral conversion in sub-F_2 population of Kapoho X Saipan Red, used for genotyping (n=84).

Test	
Kolmogorov-Smirnov	P=0.083
Cramer-von Mises	P=0.109
Anderson-Darling	P=0.054
Skewness	0.630
Kurtosis	0.414

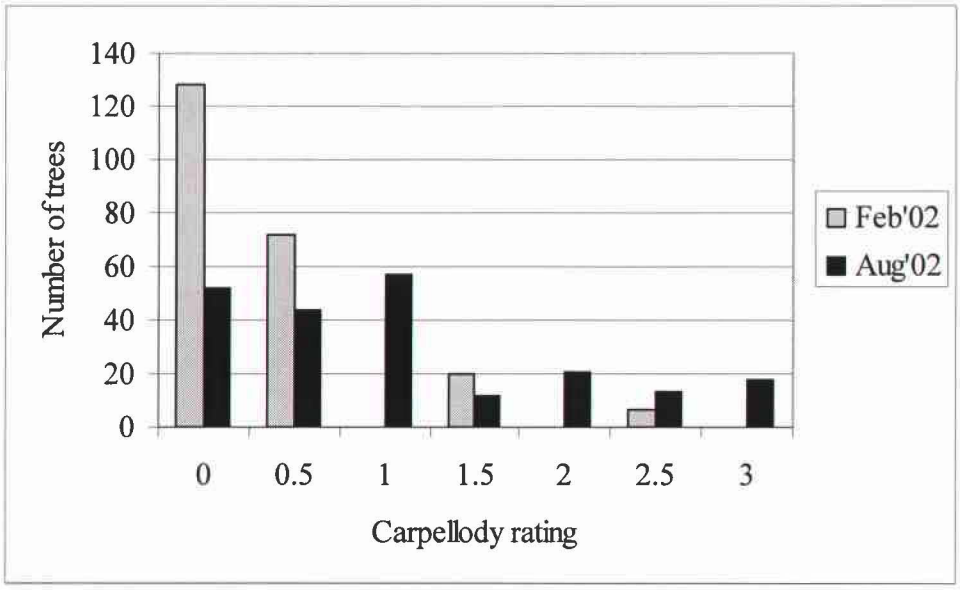


Fig. 2.5 The frequency distribution of carpellody over two seasons in hermaphrodite plants of Kapoho X Saipan Red F_2 population (n=214).

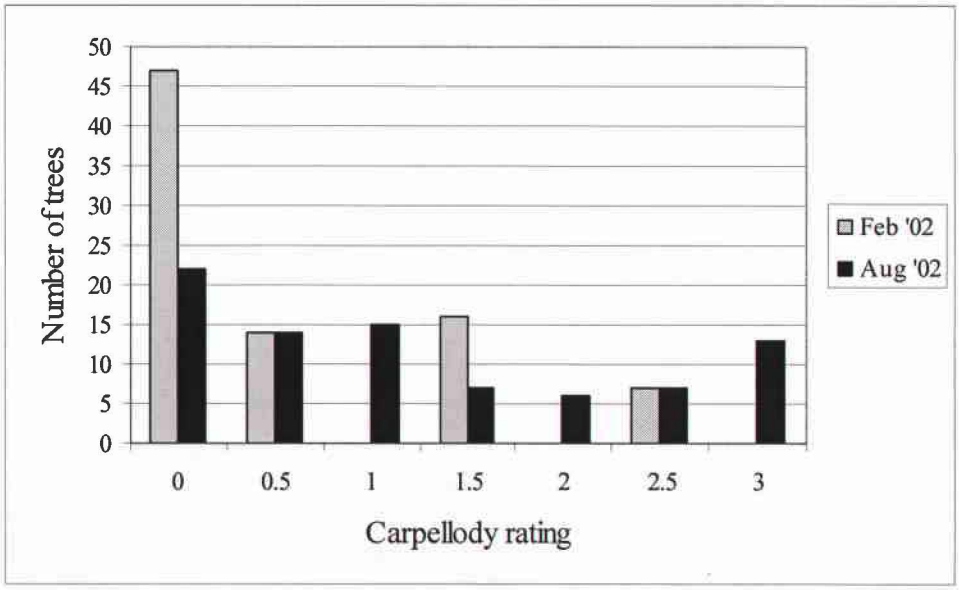


Fig. 2.6 The frequency distribution of carpellody over two seasons in sub-F_2 population of Kapoho X Saipan Red, used for genotyping (n=84).

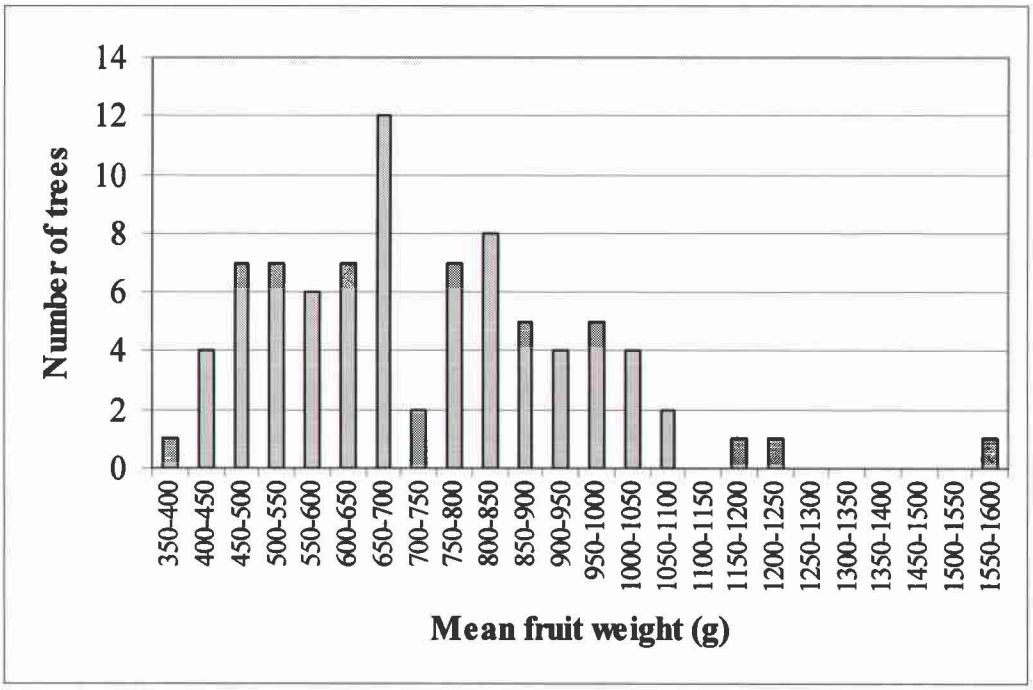


Fig. 2.7 The frequency distribution of mean fruit weight in sub-F_2 population of Kapoho X Saipan Red, used for genotyping (n=84, weight represents means of 5 fruits per tree).

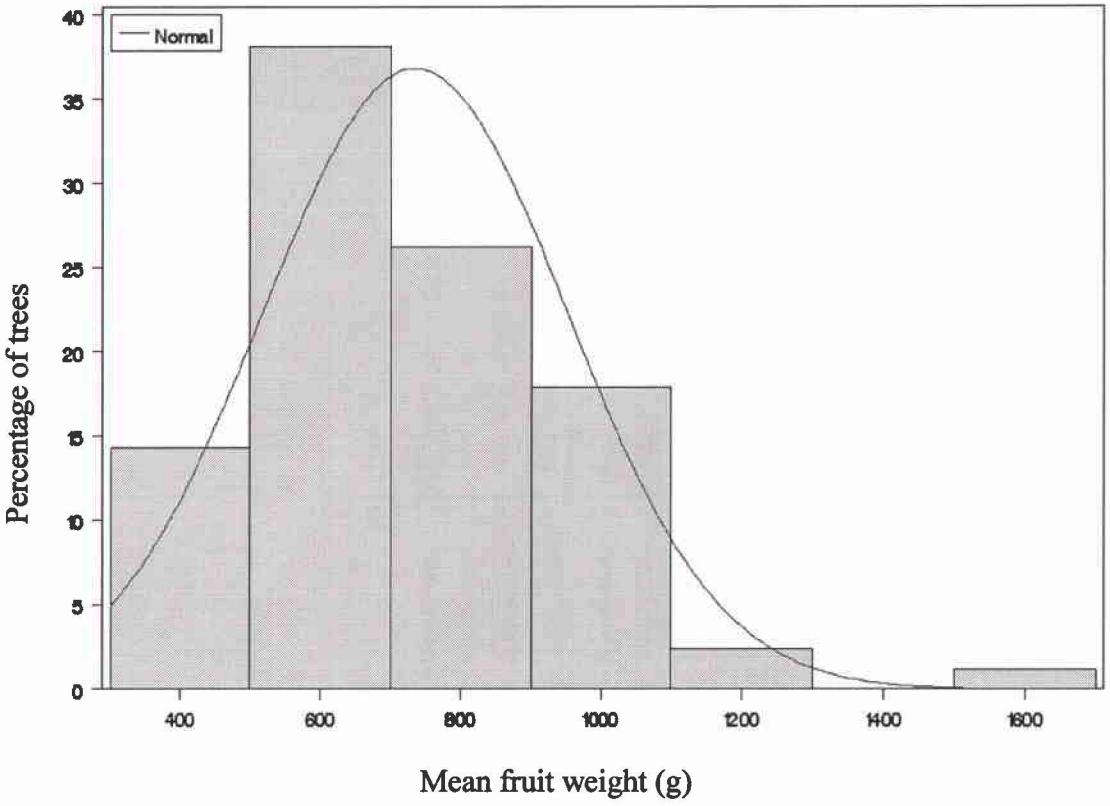


Fig. 2.8 Expected and observed phenotypic data distribution for mean fruit weight in sub-F_2 population of Kapoho X Saipan Red, used for genotyping (n=84).

Table 2.7 Results of tests of normality for mean fruit weight in sub-F_2 population of Kapoho X Saipan Red, used for genotyping (n=84).

Test	
Kolmogorov-Smirnov	0.068
Cramer-von Mises	0.185
Anderson-Darling	0.124
Skewness	0.837
Kurtosis	1.440

2.4.1.2 Qualitative traits

2.4.1.2.1 Phosphoglucomutase (PGM)

The segregation ratio of PGM alleles in the F_2 population did not follow the expected Mendelian ratio of 1:2:1 at P = 0.05 (Table 2.8). This failure was due to the slow migrating electromorph (S) from Saipan Red being under-represented in the F_2 population.

Relative to the number of plants that were homozygous for the fast migrating Kapoho allele (F/F), the number of F/S plants was 9% (142 instead of 156) less than expected and the number of S/S plants was 36% (50 instead of 78) less than expected.

In the F_2 population, phenotypic observation suggested a nonrandom relationship between PGM genotypes and the quantitative trait, node number at floral conversion. It appeared that the F allele of PGM was linked in coupling-phase to a QTL conferring more nodes of vegetative growth before flowering (Fig. 2.9). Regression analysis confirmed that this relationship was highly significant (P=0.005) in the F_2 population (Table 2.9).

2.4.1.2.2 Flesh color

In the F_2 population 71.53% of the trees showed yellow fruit-flesh color and 28.47% of the trees showed red fruit-flesh color, following the expected Mendelian segregation pattern of 3:1 at P=0.05

Table 2.8 Chi-square test for the segregation ratio of PGM alleles in F_2 population.

	PGM allele distribution		
	F/F	**F/S**	**S/S**
Expected	67.5	135.0	67.5
Observed	78.0	142.0	50.0
	P (X^2)= 0.038		

Table 2.9 ANOVA of regression analysis between PGM and node number at floral conversion in Kapoho X Saipan Red F_2 population.

Source of variation	**df**	**SS**	**MS**	**F**	**P (F)**
Regression	1	143.88	143.88	8.10	0.0048
Residual	263	4670.28	17.76		
Total	264	4814.17			

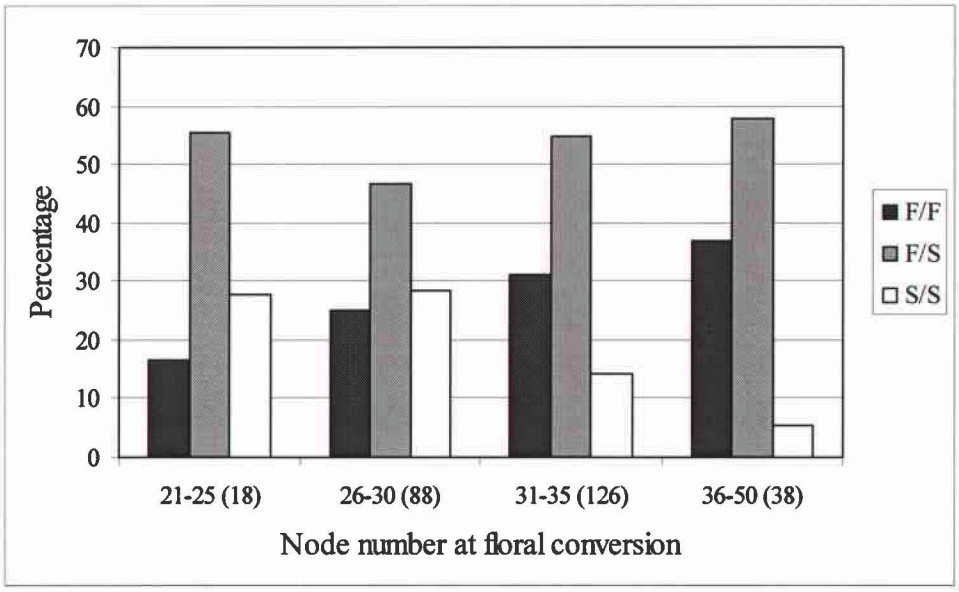


Fig. 2.9 Relationship between PGM genotypes and node number at floral conversion in Kapoho X Saipan Red F_2 population. The numbers in parenthesis represent the number of plants in each category.

Field observations suggested that one of the major genes controlling carpellody might be linked to the locus determining fruit-flesh color. The percentage of trees with red-fleshed fruit was higher than expected among carpellodic individuals and lower than expected among noncarpellodic individuals (Fig. 2.10). But the correlation coefficient was significant in neither Feb'02 (r = -0.1099) nor in Aug'02 (r = -0.0213).

2.4.2 AFLP markers

2.4.2.1 Polymorphism

The parents Kapoho and Saipan Red, along with 84 plants selected from the F_2 for marker genotyping, were screened with 158 AFLP primer sets. A total of 513 loci showed polymorphism between the parents and also segregated among the F_2 plants (Exampl: Fig. 2.11). On an average, 73.8 bands were produced per primer set and 3.3 markers were observed per primer combination. Over all, 4.4% of the AFLP bands were polymorphic. Most of the amplified marker fragments were large-sized (26.1% > 350 bp) or medium-sized (80 bp < 69.4% < 350 bp) and only 4.5% markers were small-sized (<80 bp). Out of 513 polymorphisms, 38 markers were co-dominant and the rest were dominant. Among the dominant markers, 254 (53.5%) came from Kapoho, and 221(46.53%) came from Saipan Red. The primer set E-ATA + M-CT produced the highest number of polymorphisms (11), while 12 other primer sets did not produce any polymorphism. The number of bands ranged from 24 to 150 per gel. The number of bands per gel and the number of markers per gel were significantly correlated (r = 0.256).

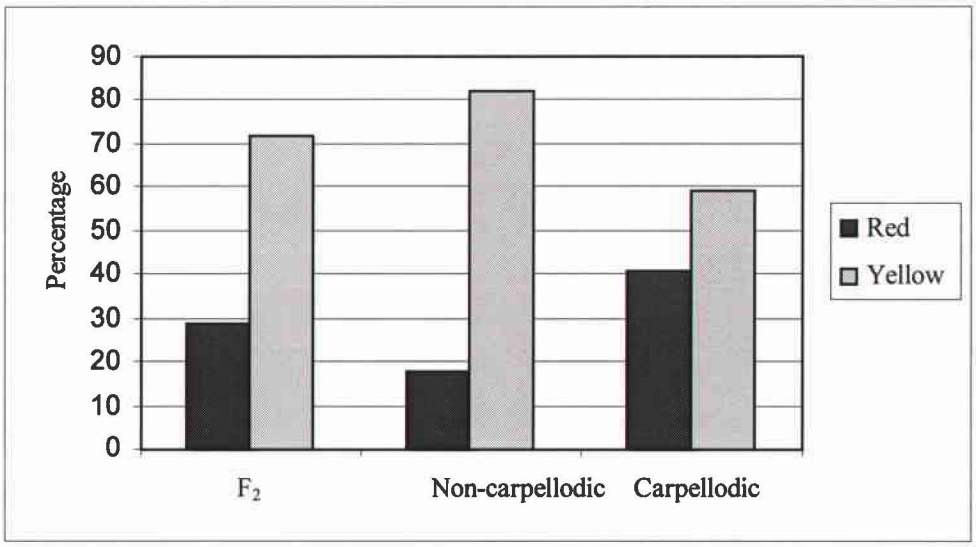


Fig. 2.10 Flesh color frequency in extreme non-carpellodic (n=28) and carpellodic (n=22) trees.

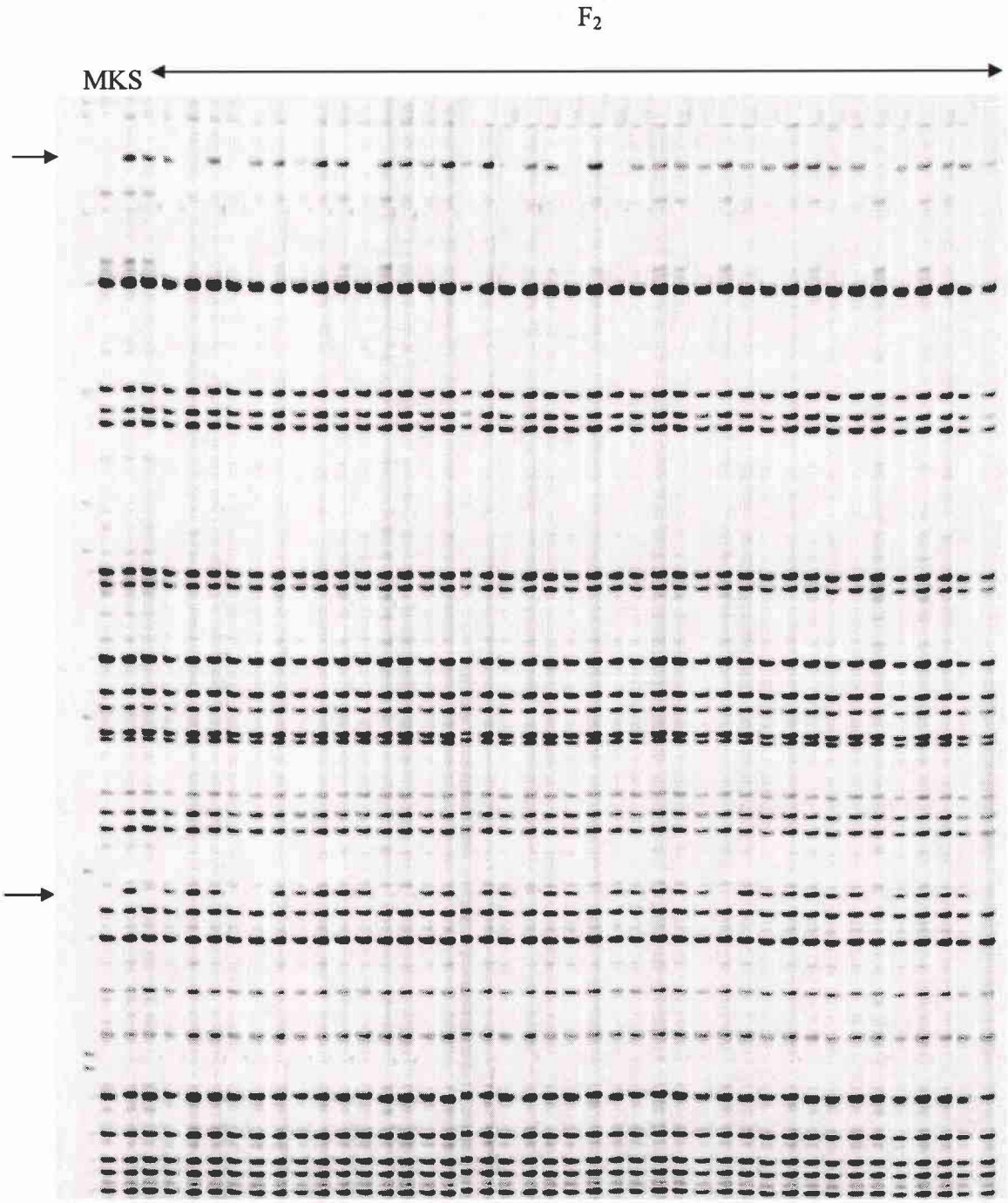


Fig. 2.11 Part of the AFLP gel obtained with primer combination E-ACC and M-CC. The arrows show the polymorphisms. The lanes M, K and S refer Molecular standard, Kapoho and Saipan Red respectively.

2.4.2.2 Effect of nucleotide extension

The primer set E-A*+M-C** and E-A**+M-C* produced the most markers (3.9) per primer set while E-G*+M-C** and E-G**+M-C* produced the fewest markers (1.9) per primer set (Table 2.10). Primer sets containing the E-A combination averaged more bands (79.6) and more markers per primer set (3.5), compared to primer sets containing the E-G combination (average number of bands = 41.5 and average number of markers = 2.0). Primer sets with E-** + M *** combinations (n = 33) produced 3.21 markers on an average and primer sets with E-***+ M** (n = 125) combinations produced 3.26 markers on an average. Thus the number of extension nucleotides in E-primers made no difference in the number of markers produced.

2.4.2.3 Segregation distortion

Out of 513 AFLP markers, 403 markers (78.4%) segregated in either 1:2:1 or 3:1 ratio (Table 2.11). Among the distorted markers, 64.5% of the markers showed 2:1 segregation patterns. As the genotyping population had only hermaphrodite plants, linkage of these markers within the sex determination region, which had a well-ocumented segregation distortion caused by an associated lethal factor, was not explained by the observed deviations from the expected 3:1 ratios. Among the 38 co-dominant. markers, 7 markers were distorted. Among the Kapoho dominant markers 14.8% were distorted, and among the Saipan Red dominant markers 14.9% were distorted.

Table 2.10 Number of bands and number of markers produced by different primer combinations.

Primer set	# of bands	# of markers
E-A* + M-C** and E-A** + M-C*	74.07	3.93
E-A** + M-G*	78.89	3.26
E-A** + M-A*	97.24	2.48
E-G* + M-C** and E-G** + M-C*	41.46	1.86

* represents any one of the bases A, T, G or C

Table 2.11 Segregation patterns of AFLP markers.

	Total	Distorted	Faint	Follow Mendelian ratio	Chi-test		
					2:1	3:1	2:1 & 3:1
PGM		1					
Flesh color				1			
Co-dominant	38	7	2	29			
Dominant from Kapoho	254	36	19	199	22	46	153
Dominant from Saipan Red	221	33	13	175	27	53	122

2.4.3 Linkage map

2.4.3.1 Mixed-phase linkage map

At a LOD score of 5.0 and a recombination frequency of 0.25, 513 AFLP markers, along with loci controlling flesh color and PGM isozyme variation, were mapped into 25 groups with 16 unlinked markers. Fourteen groups containing 4 or more markers were considered for ordering of markers within the linkage group. In addition, a group with three markers, which contained the flesh color locus, was also considered for ordering. The 15 linkage groups with 413 markers covered 2066.6 cM of the genome (Fig. 2.12). The linkage groups were labeled LG-1 to LG-15 based on their size. The individual groups ranged in size from 16.7 cM to 319.0 cM. The number of markers ranged from 3 to 104 in each group (Table 2.12). Although the average marker interval is 5.0 cM in the whole map, average marker intervals in individual groups ranged from 2.6 cM in LG-3 to 8.8 cM in LG-8 and LG-9. LG-7 contains the PGM locus with flanking markers 11.0 and 8.3 cM away. The flesh color locus fell on the end of LG-14 with a linked marker at 4.1 cM distance.

2.4.3.1.1 Segregation distortion

The linkage groups LG-4 and LG-5 accounted for 49.4% of the distorted markers. In LG-4, all the markers in the central region of the linkage group were dominant markers from Kapoho and showed a segregation ratio of 1:2. In LG-5, almost all the markers showed unidirectional segregation distortion, with a predominance of the Kapoho allele.

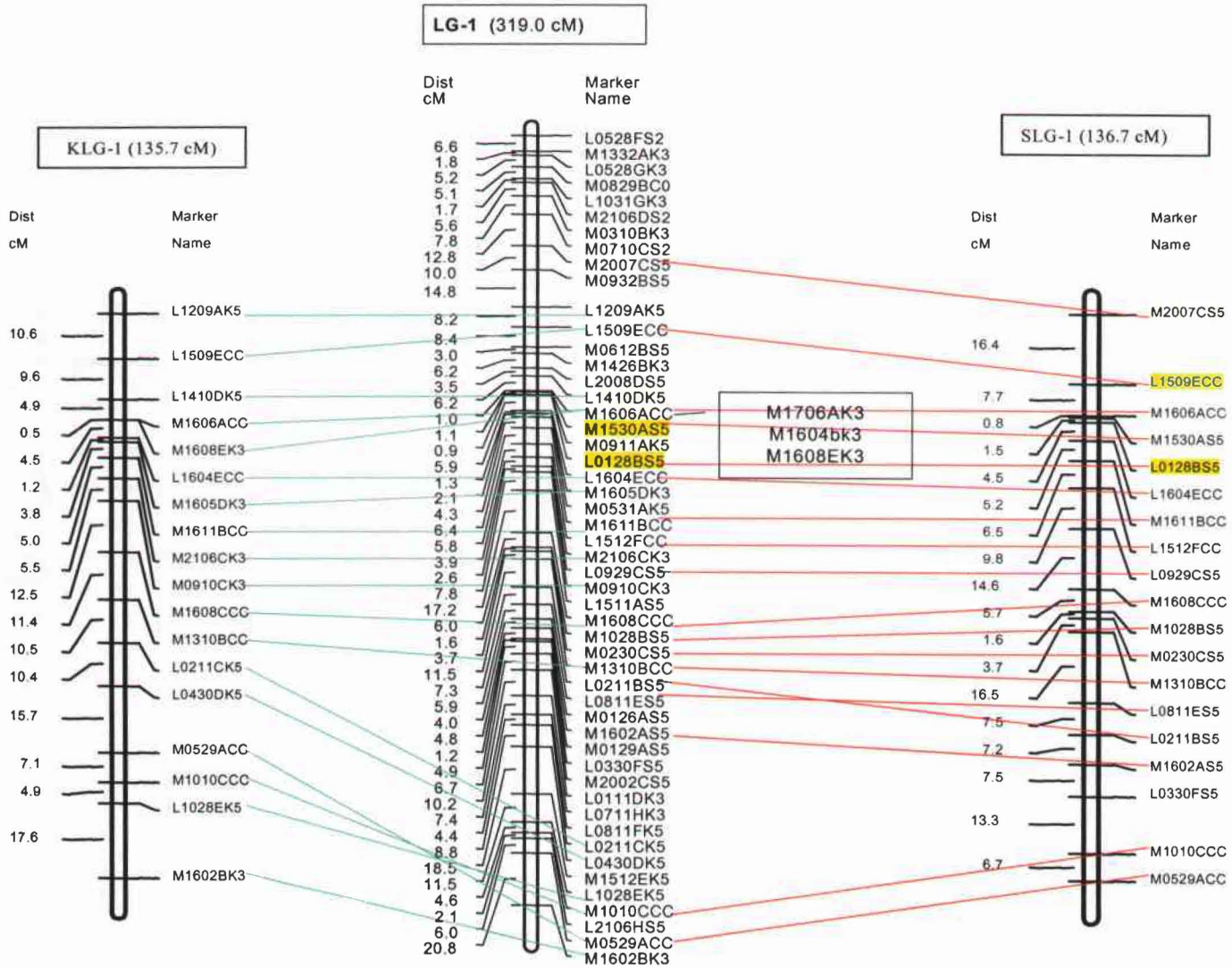
Table 2.12 Summary of mixed-phase and coupling-phase linkage maps.

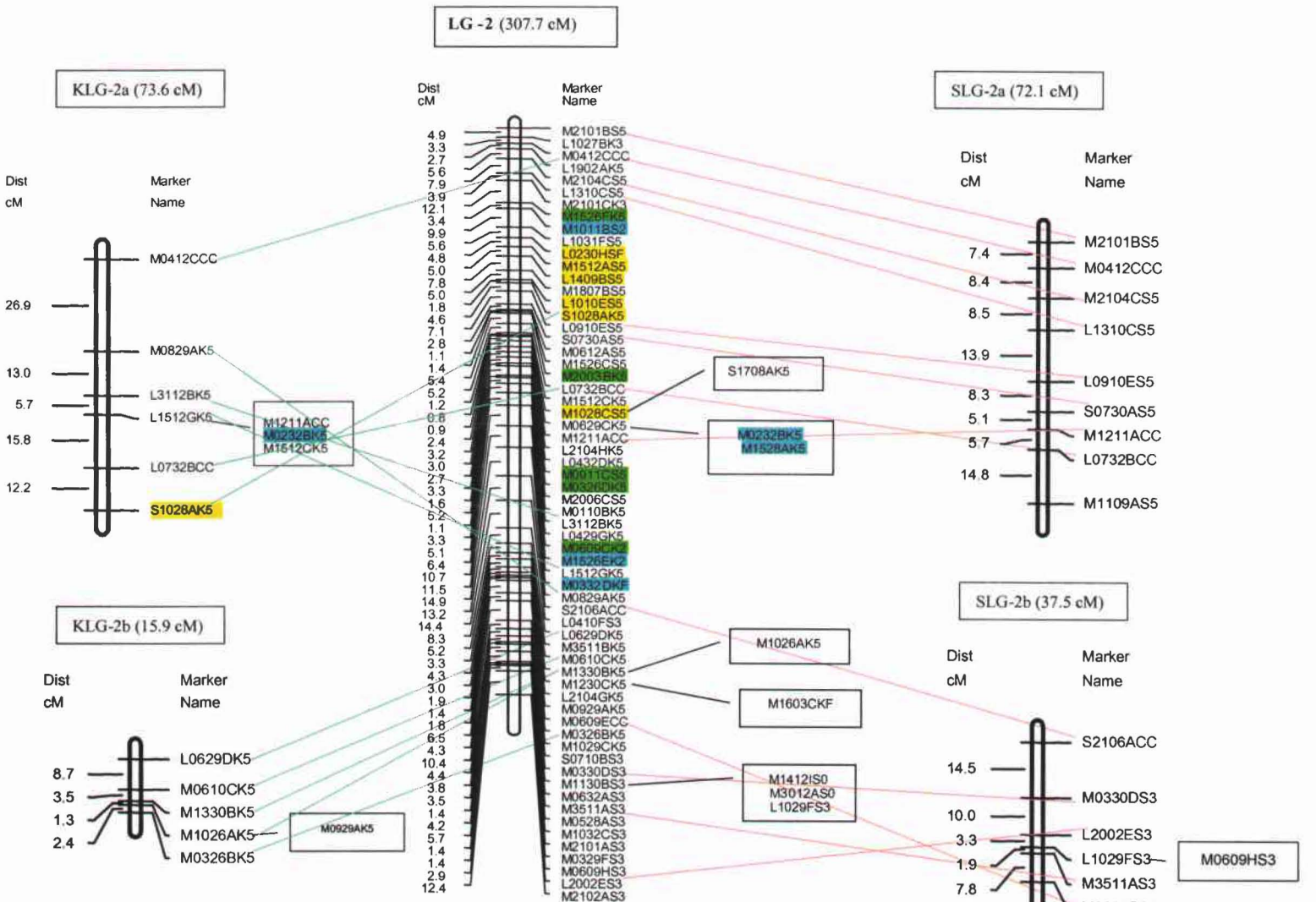
Mixed-phase map					Coupling-phase Map for Kapoho markers					Coupling-phase Map for Saipan Red markers				
Linkage group	Number of markers mapped	Number of frame-work markers	Map distance (cM)	Average marker interval (cM)	Linkage group	Number of markers mapped	Number of frame-work markers	Map distance (cM)	Average marker interval (cM)	Linkage group	Number of markers mapped	Number of frame-work markers	Map distance (cM)	Average marker interval (cM)
LG1	54	51	319.0	5.9	KLG1	18	18	135.7	7.6	SLG1	19	19	136.7	7.2
LG2	71	63	307.7	4.3	KLG2a	9	6	73.6	8.2	SLG2a	9	9	72.1	8.1
					KLG2b	6	5	15.9	2.7	SLG2b	7	6	37.5	5.4
LG3	104	77	269.0	2.6	KLG3	22	18	101.3	4.6	SLG3	16	9	68.8	4.4
LG4	32	29	256.8	8.0	KLG4	14	14	124.9	9.0	SLG4a	4	4	38.8	9.9
										SLG4b	5	5	37.7	7.7
LG5	30	29	221.6	7.4	KLG5a	7	7	37.2	5.3	SLG5	11	11	117.4	10.7
					KLG5b	3	3	16.3	5.4					
LG6	46	41	198.5	4.3	KLG6	13	9	47.4	3.7	SLG6	7	7	45.7	6.6
LG7	27	26	167.3	6.2	KLG7	9	8	66.3	7.4	SLG7	8	8	80.0	10.2
LG8	7	7	61.3	8.8						SLG8	4	3	10.6	2.7
LG9	6	5	52.8	8.8	KLG9	5	4	47.7	9.7					
LG10	9	9	48.4	5.6	KLG10	3	3	18.2	6.1	SLG10	3	3	10.9	3.6
LG11	7	7	48.9	7.0	KLG11	6	5	46.4	4.4					
LG12	6	6	44.7	7.5	KLG12	4	4	28.0	7.0					
LG13	6	6	35.4	5.9						SLG13	3	3	20.2	6.7
LG14	3	3	18.5	6.2	KLG14	2	2	19.9	10.0					
LG15	5	5	16.7	3.3						SLG15	5	5	16.7	3.3
Total	413	364	2066.6	5.0	Total	121	106	778.8	6.4	Total	101	92	693.0	6.9

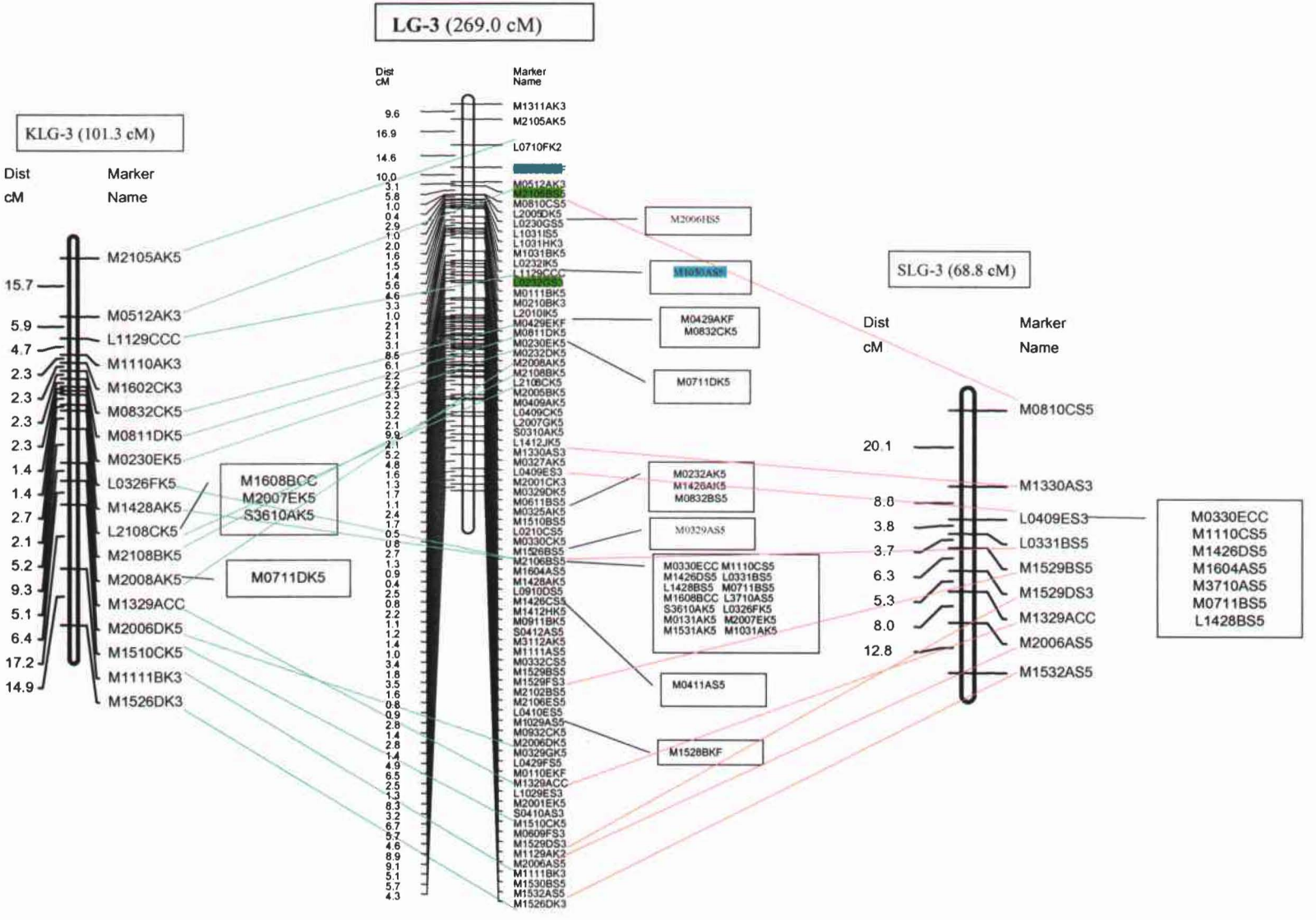
Fig. 2.12 Mixed- and coupling-phase linkage maps of papaya generated from Mapmaker/Exp V3.0 using 513 AFLP markers, an isozyme marker and a morphological marker. Co-segregating markers are listed in boxes. Each marker name has 8 characters. First character represents the size of the marker (small: < 80bp, medium: 80-350 bp or large >350 bp). Next two characters are codes used for EcoRI primer and fourth and fifth characters are codes used for MseI primers. Sixth character represents the order of the marker in the gel for the particular primer pair. Seventh character represents which parent contributes the dominant form of the marker (K-Kapoho; S-Saipan Red; C-co-dominant). Eighth character represents the marker segregation pattern (2=2:1 ratio; 3=3:1 ratio; 5-both 2:1 and 3:1 ratio; C-1:2:1; 0-neither of these ratios)

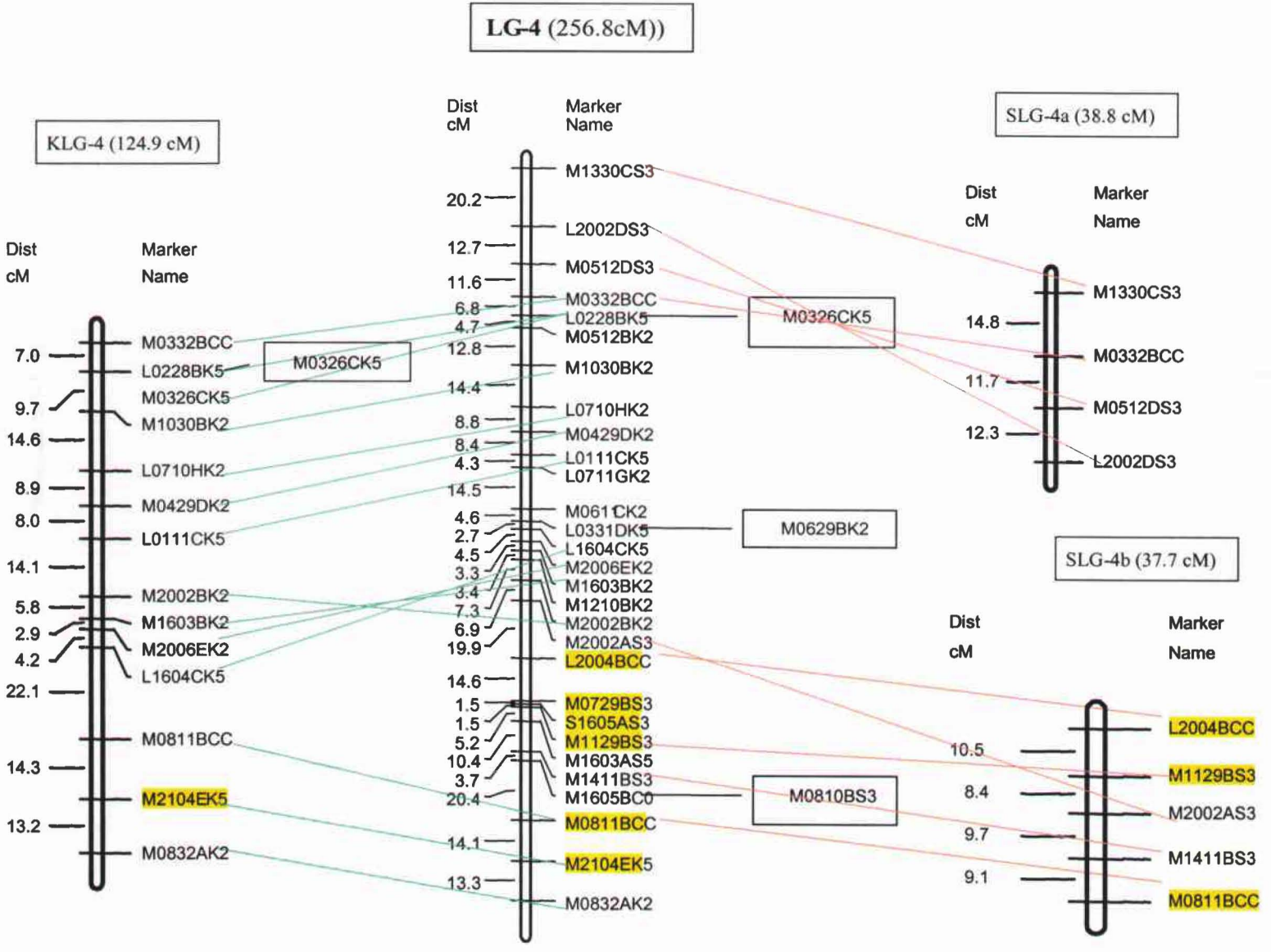
Maps in the center are mixed-phase maps. Maps on the left side are coupling-phase maps for dominant markers from Kapoho. Maps on the right side are coupling-phase maps for dominant markers from Saipan Red. Position of framework markers of coupling-phase maps have been indicated on mixed-phase maps by connecting lines.

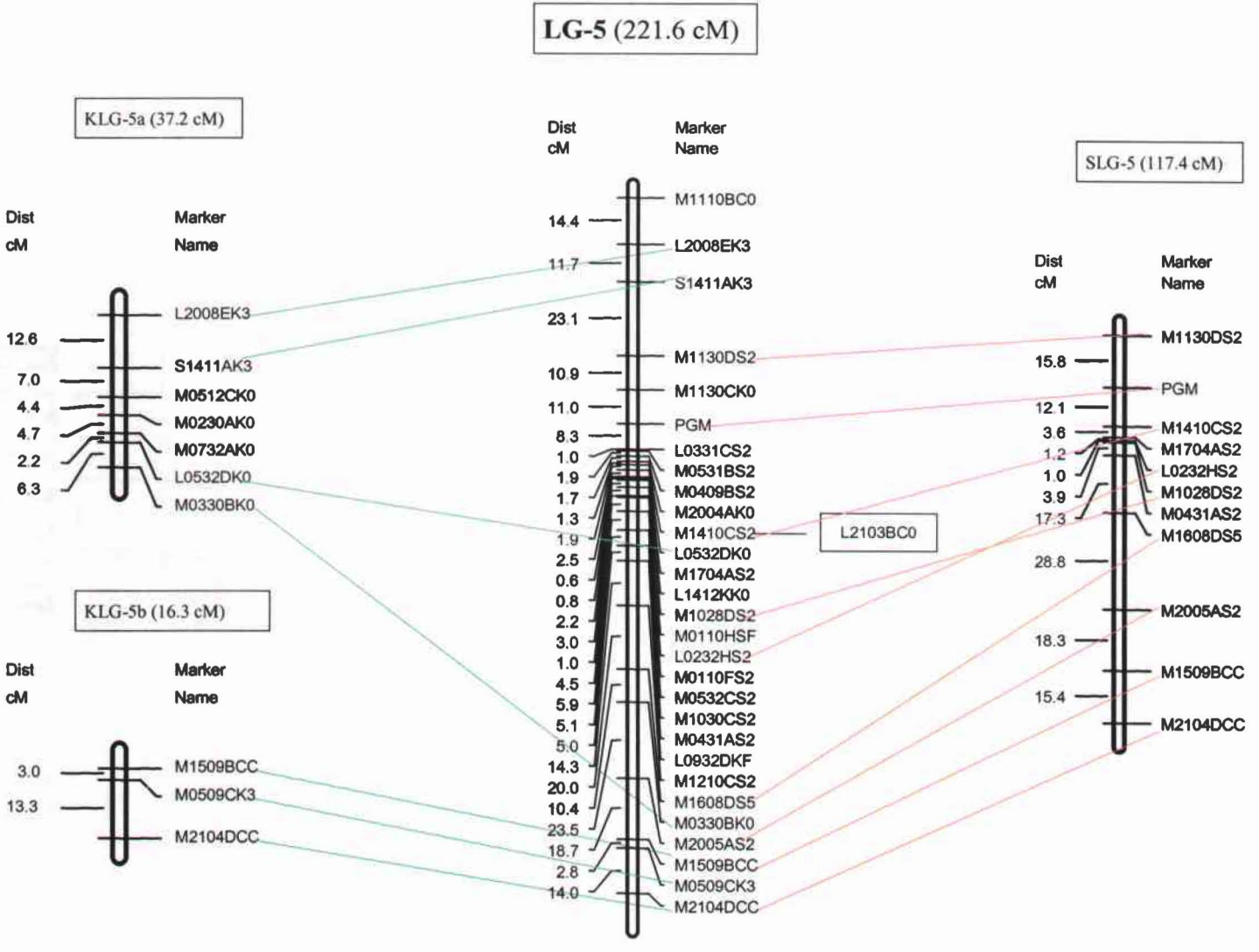
In the mixed-phase maps, markers that show significant association with carpellody in Feb'02, Aug'02 and both seasons are highlighted in blue, yellow and green color, respectively.

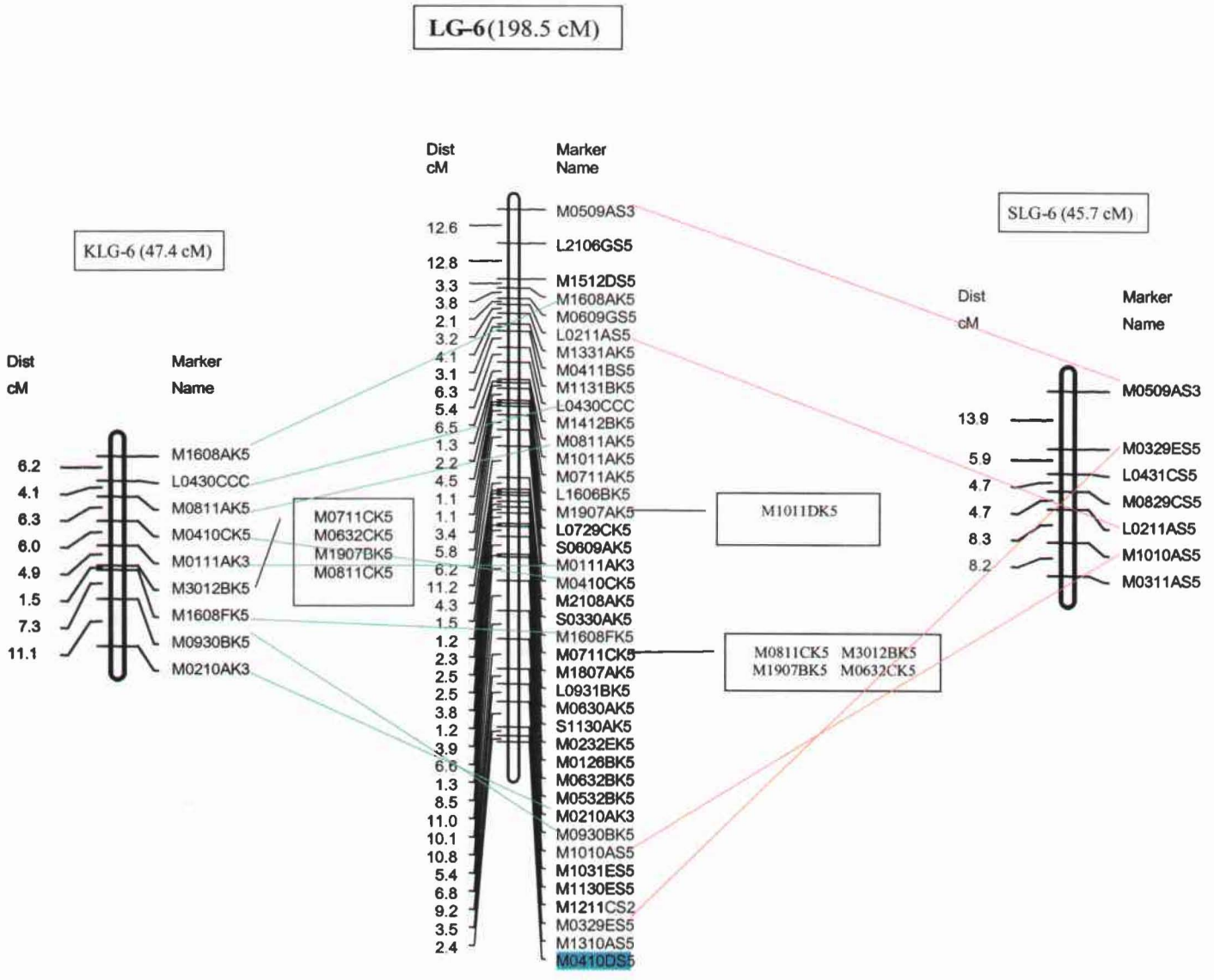

LG-1 (319.0 cM)
Dist
cM
Marker
Name
KLG-1 (135.7 cM)
Dist
cM
Marker
Name
SLG-1 (136.7 cM)
Dist
cM
Marker
Name
M1706AK3
M1604bk3
M1608EK3

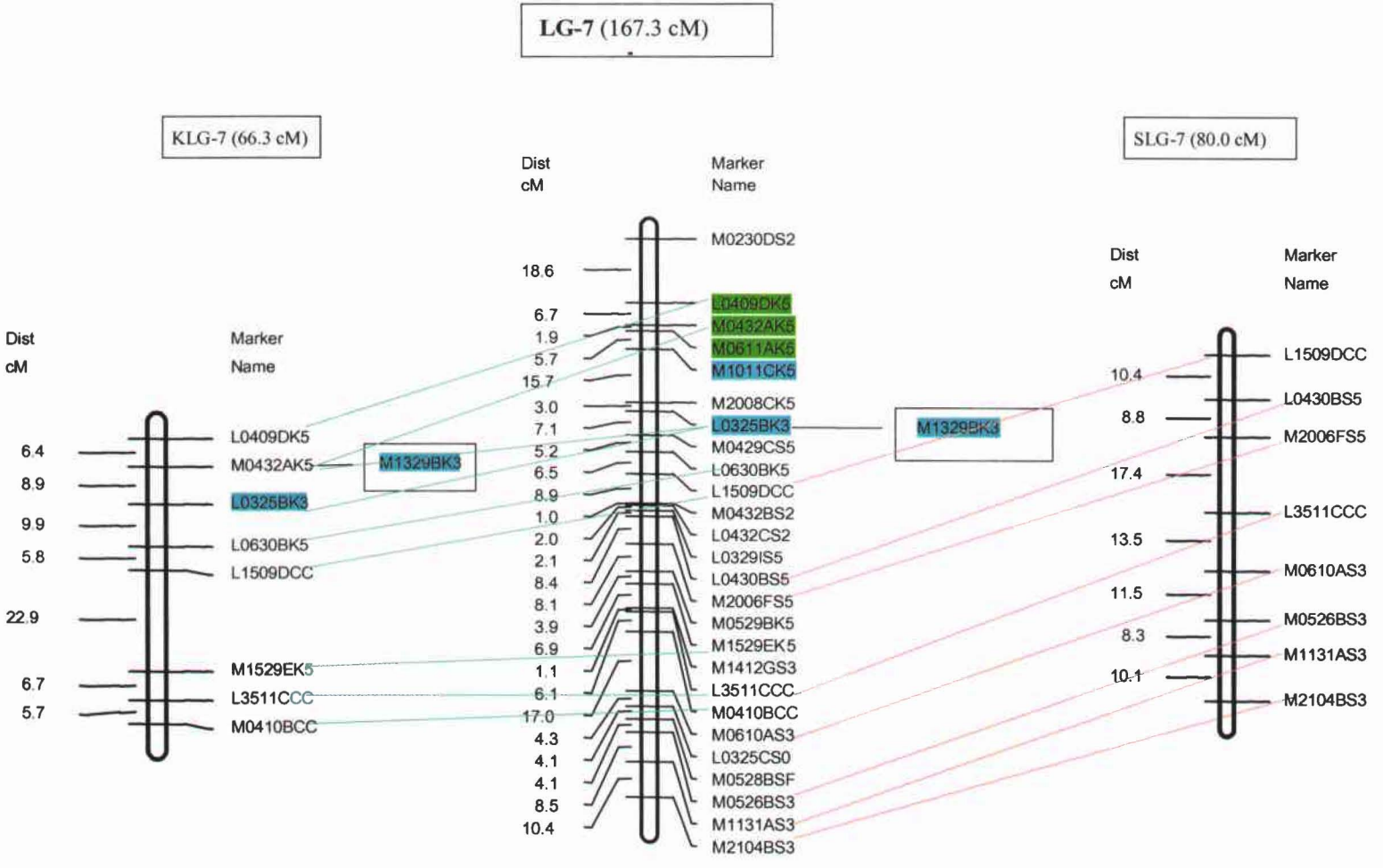

LG -2 (307.7 cM)
KLG-2a (73.6 cM)
Dist
cM
Marker
Name
M0412CCC
26.9
M0829AK5
13.0
L3112BK5
5.7
L1512GK5
15.8
L0732BCC
12.2
S1028AK5
M1211ACC
M0232BK5
M1512CK5
KLG-2b (15.9 cM)
Dist
cM
Marker
Name
8.7
L0629DK5
3.5
M0610CK5
1.3
M1330BK5
2.4
M1026AK5
M0326BK5
M0929AK5
Dist
cM
Marker
Name
4.9
M2101BS5
3.3
L1027BK3
2.7
M0412CCC
5.6
L1902AK5
7.9
M2104CS5
3.9
L1310CS5
12.1
M2101CK3
3.4
M1011BS2
9.9
L1031FS5
5.6
L0230HSF
4.8
M1512AS5
5.0
L1409BS5
7.8
M1807BS5
5.0
L1010ES5
1.8
S1028AK5
4.6
L0910ES5
7.1
S0730AS5
2.8
M0612AS5
1.1
M1626CS5
1.4
L0732BCC
5.2
M1512CK5
1.2
M1028CS5
0.8
M0629CK5
0.9
M1211ACC
2.4
L2104HK5
3.2
3.0
2.7
3.3
1.6
5.2
M2006CS5
1.1
M0110BK5
3.3
L3112BK5
5.1
6.4
10.7
11.5
14.9
13.2
14.4
8.3
5.2
3.3
4.3
3.0
1.9
1.4
1.8
6.5
4.3
10.4
4.4
3.8
3.5
1.4
4.2
5.7
1.4
1.4
2.9
12.4
M0829AK5
S2106ACC
L0410FS3
L0629DK5
M3511BK5
M0610CK5
M1330BK5
M1230CK5
L2104GK5
M0929AK5
M0609ECC
M0326BK5
M1029CK5
S0710BS3
M0330DS3
M1130BS3
M0632AS3
M3511AS3
M0528AS3
M1032CS3
M2101AS3
M0329FS3
M0609HS3
L2002ES3
M2102AS3
S1708AK5
M0232BK5
M1528AK5
M1026AK5
M1603CKF
M1412IS0
M3012AS0
L1029FS3
SLG-2a (72.1 cM)
Dist
cM
Marker
Name
7.4
M2101BS5
8.4
M0412CCC
8.5
M2104CS5
13.9
L1310CS5
8.3
L0910ES5
5.1
S0730AS5
5.7
M1211ACC
14.8
L0732BCC
M1109AS5
SLG-2b (37.5 cM)
Dist
cM
Marker
Name
14.5
S2106ACC
10.0
M0330DS3
3.3
L2002ES3
1.9
L1029FS3
7.8
M3511AS3
M0609ECC
M0609HS3

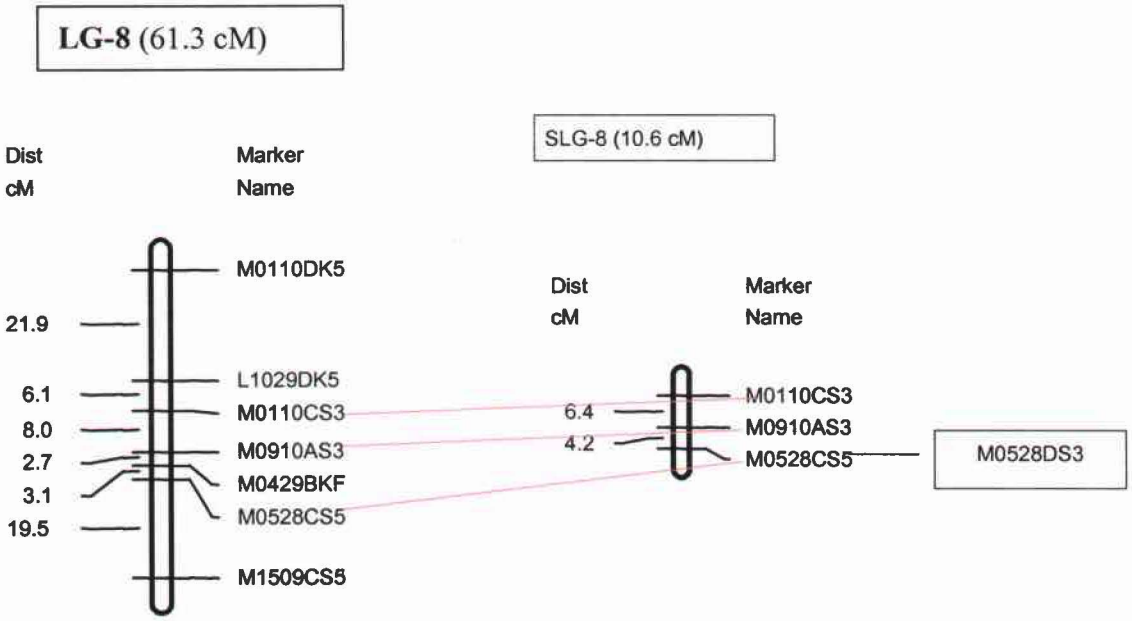

LG-3 (269.0 cM)
KLG-3 (101.3 cM)
Dist
cM
Marker
Name
M2105AK5
M0512AK3
L1129CCC
M1110AK3
M1602CK3
M0832CK5
M0811DK5
M0230EK5
L0326FK5
M1428AK5
L2108CK5
M2108BK5
M2008AK5
M1329ACC
M2006DK5
M1510CK5
M1111BK3
M1526DK3
M1608BCC
M2007EK5
S3610AK5
M0711DK5
M1311AK3
M2105AK5
L0710FK2
M0810CS5
M2006HS5
M0429AKF
M0832CK5
M0711DK5
M0232AK5
M1426AK5
M0832BS5
M0329AS5
M0330ECC M1110CS5
M1426DS5 L0331BS5
L1428BS5 M0711BS5
M1608BCC L3710AS5
S3610AK5 L0326FK5
M0131AK5 M2007EK5
M1531AK5 M1031AK5
M0411AS5
M1528BKF
SLG-3 (68.8 cM)
M0810CS5
M1330AS3
L0409ES3
L0331BS5
M1529BS5
M1529DS3
M1329ACC
M2006AS5
M1532AS5
M0330ECC
M1110CS5
M1426DS5
M1604AS5
M3710AS5
M0711BS5
L1428BS5

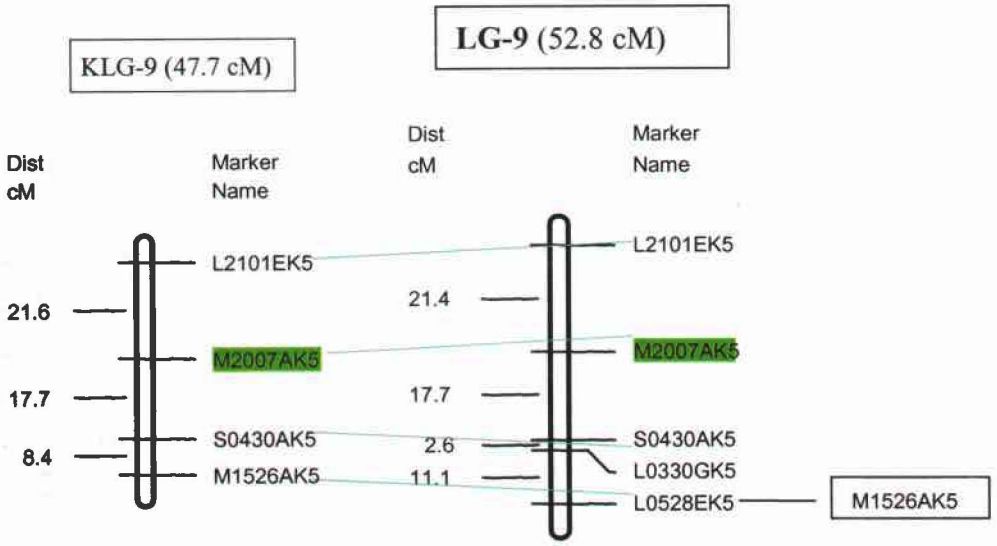

LG-4 (256.8cM))
KLG-4 (124.9 cM)
Dist
cM
Marker
Name
M0332BCC
L0228BK5
M0326CK5
M1030BK2
L0710HK2
M0429DK2
L0111CK5
M2002BK2
M1603BK2
M2006EK2
L1604CK5
M0811BCC
M2104EK5
M0832AK2
7.0
9.7
14.6
8.9
8.0
14.1
5.8
2.9
4.2
22.1
14.3
13.2
M0326CK5
M0629BK2
M0810BS3
M1330CS3
L2002DS3
M0512DS3
M0332BCC
L0228BK5
M0512BK2
M1030BK2
L0710HK2
M0429DK2
L0111CK5
L0711GK2
M0611CK2
L0331DK5
L1604CK5
M2006EK2
M1603BK2
M1210BK2
M2002BK2
M2002AS3
L2004BCC
M0729BS3
S1605AS3
M1129BS3
M1603AS5
M1411BS3
M1605BC0
M0811BCC
M2104EK5
M0832AK2
20.2
12.7
11.6
6.8
4.7
12.8
14.4
8.8
8.4
4.3
14.5
4.6
2.7
4.5
3.3
3.4
7.3
6.9
19.9
14.6
1.5
1.5
5.2
10.4
3.7
20.4
14.1
13.3
SLG-4a (38.8 cM)
M1330CS3
M0332BCC
M0512DS3
L2002DS3
14.8
11.7
12.3
SLG-4b (37.7 cM)
L2004BCC
M1129BS3
M2002AS3
M1411BS3
M0811BCC
10.5
8.4
9.7
9.1

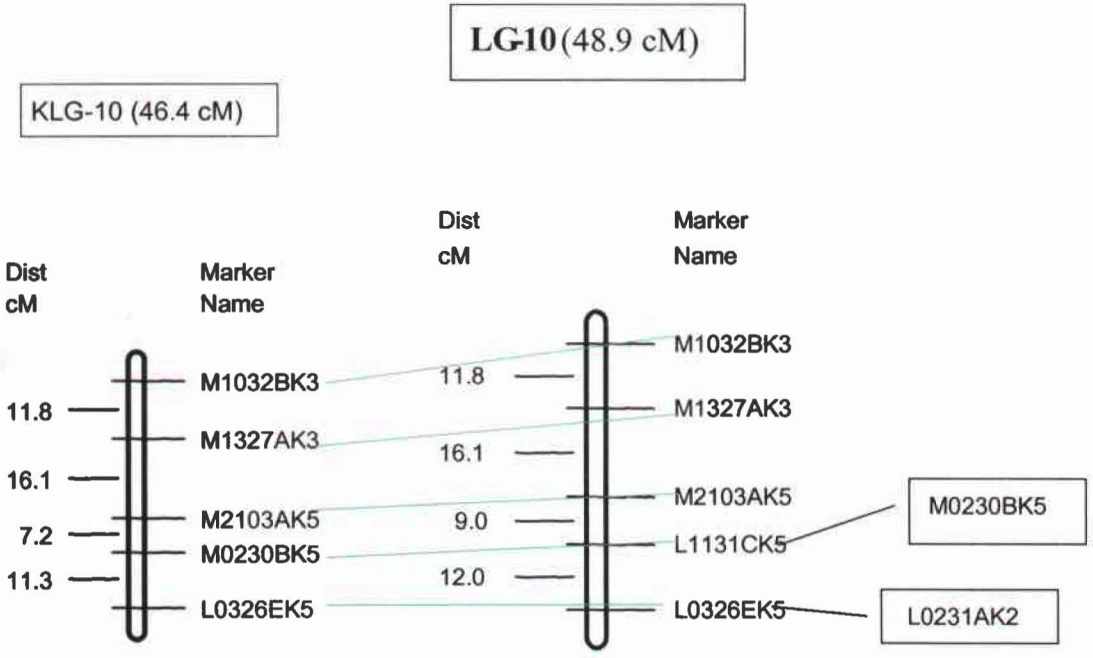

LG-5 (221.6 cM)
KLG-5a (37.2 cM)
Dist cM
Marker Name
L2008EK3
12.6
S1411AK3
7.0
M0512CK0
4.4
M0230AK0
4.7
M0732AK0
2.2
L0532DK0
6.3
M0330BK0
KLG-5b (16.3 cM)
Dist cM
Marker Name
3.0
M1509BCC
13.3
M0509CK3
M2104DCC
Dist cM
Marker Name
M1110BC0
14.4
L2008EK3
11.7
S1411AK3
23.1
M1130DS2
10.9
M1130CK0
11.0
PGM
8.3
L0331CS2
1.0
M0531BS2
1.9
M0409BS2
1.7
M2004AK0
1.3
M1410CS2
1.9
L0532DK0
2.5
M1704AS2
0.6
L1412KK0
0.8
M1028DS2
2.2
M0110HSF
3.0
L0232HS2
1.0
M0110FS2
4.5
M0532CS2
5.9
M1030CS2
5.1
M0431AS2
5.0
L0932DKF
14.3
M1210CS2
20.0
M1608DS5
10.4
M0330BK0
23.5
M2005AS2
18.7
M1509BCC
2.8
M0509CK3
14.0
M2104DCC
L2103BC0
SLG-5 (117.4 cM)
Dist cM
Marker Name
M1130DS2
15.8
PGM
12.1
M1410CS2
3.6
M1704AS2
1.2
L0232HS2
1.0
M1028DS2
3.9
M0431AS2
17.3
M1608DS5
28.8
M2005AS2
18.3
M1509BCC
15.4
M2104DCC

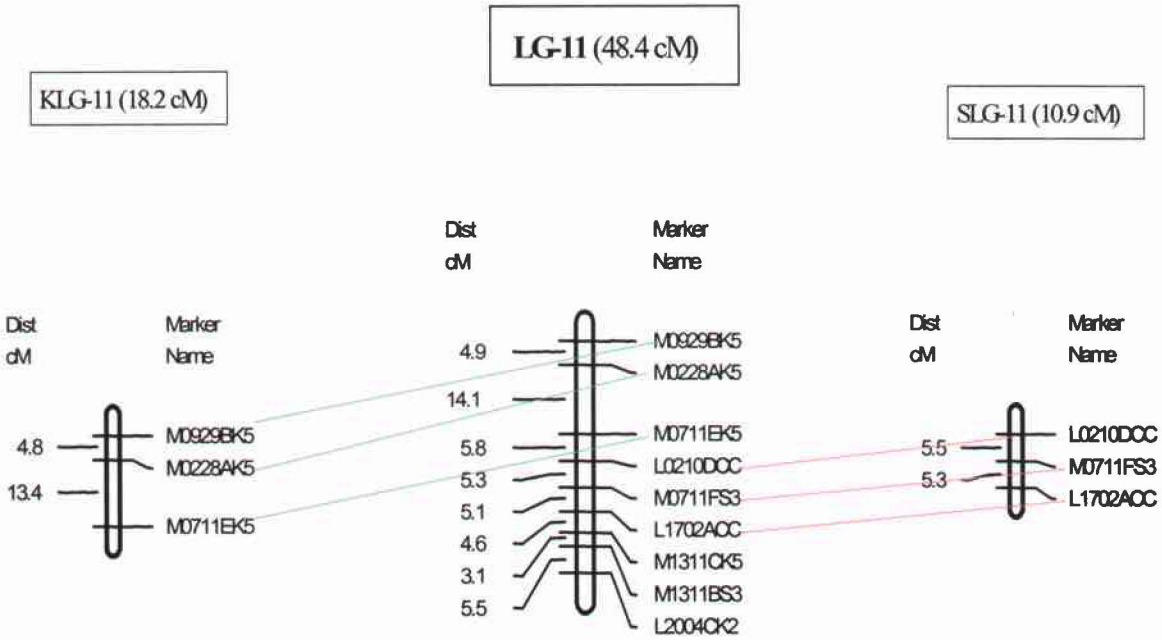

LG-6(198.5 cM)
KLG-6 (47.4 cM)
Dist cM
Marker Name
6.2 M1608AK5
4.1 L0430CCC
6.3 M0811AK5
6.0 M0410CK5
4.9 M0111AK3
1.5 M3012BK5
7.3 M1608FK5
11.1 M0930BK5
M0210AK3
M0711CK5
M0632CK5
M1907BK5
M0811CK5
Dist cM
Marker Name
M0509AS3
12.6
L2106GS5
12.8
M1512DS5
3.3 M1608AK5
3.8 M0609GS5
2.1 L0211AS5
3.2 M1331AK5
4.1 M0411BS5
3.1 M1131BK5
6.3 L0430CCC
5.4 M1412BK5
6.5 M0811AK5
1.3 M1011AK5
2.2 M0711AK5
4.5 L1606BK5
1.1 M1907AK5
1.1 L0729CK5
3.4 S0609AK5
5.8 M0111AK3
6.2 M0410CK5
11.2 M2108AK5
4.3 S0330AK5
1.5 M1608FK5
1.2 M0711CK5
2.3 M1807AK5
2.5 L0931BK5
2.5 M0630AK5
3.8 S1130AK5
1.2 M0232EK5
3.9 M0126BK5
6.6 M0632BK5
1.3 M0532BK5
8.5 M0210AK3
11.0 M0930BK5
10.1 M1010AS5
10.8 M1031ES5
5.4 M1130ES5
6.8 M1211CS2
9.2 M0329ES5
3.5 M1310AS5
2.4 M0410DS5
M1011DK5
M0811CK5 M3012BK5
M1907BK5 M0632CK5
SLG-6 (45.7 cM)
Dist cM
Marker Name
M0509AS3
13.9
5.9 M0329ES5
4.7 L0431CS5
4.7 M0829CS5
8.3 L0211AS5
8.2 M1010AS5
M0311AS5


LG-7 (167.3 cM)
KLG-7 (66.3 cM)
SLG-7 (80.0 cM)
Dist
cM
Marker
Name
6.4
8.9
9.9
5.8
22.9
6.7
5.7
L0409DK5
M0432AK5
L0325BK3
L0630BK5
L1509DCC
M1529EK5
L3511CCC
M0410BCC
M1329BK3
18.6
6.7
1.9
5.7
15.7
3.0
7.1
5.2
6.5
8.9
1.0
2.0
2.1
8.4
8.1
3.9
6.9
1.1
6.1
17.0
4.3
4.1
4.1
8.5
10.4
M0230DS2
L0409DK5
M0432AK5
M0611AK5
M1011CK5
M2008CK5
L0325BK3
M0429CS5
L0630BK5
L1509DCC
M0432BS2
L0432CS2
L0329IS5
L0430BS5
M2006FS5
M0529BK5
M1529EK5
M1412GS3
L3511CCC
M0410BCC
M0610AS3
L0325CS0
M0528BSF
M0526BS3
M1131AS3
M2104BS3
M1329BK3
10.4
8.8
17.4
13.5
11.5
8.3
10.1
L1509DCC
L0430BS5
M2006FS5
L3511CCC
M0610AS3
M0526BS3
M1131AS3
M2104BS3

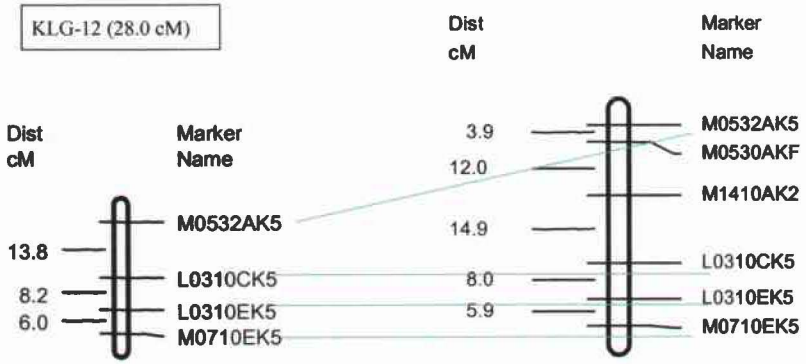

LG-8 (61.3 cM)
Dist
cM
Marker
Name
M0110DK5
21.9
L1029DK5
6.1
M0110CS3
8.0
M0910AS3
2.7
M0429BKF
3.1
M0528CS5
19.5
M1509CS5
SLG-8 (10.6 cM)
Dist
cM
Marker
Name
M0110CS3
6.4
M0910AS3
4.2
M0528CS5
M0528DS3


KLG-9 (47.7 cM)
LG-9 (52.8 cM)
Dist cM
Marker Name
L2101EK5
21.6
M2007AK5
17.7
S0430AK5
8.4
M1526AK5
Dist cM
Marker Name
L2101EK5
21.4
M2007AK5
17.7
S0430AK5
2.6
L0330GK5
11.1
L0528EK5
M1526AK5

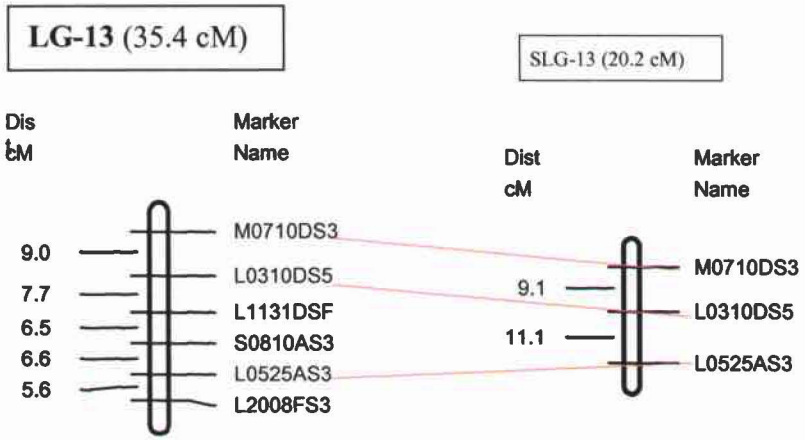

LG-10 (48.9 cM)
KLG-10 (46.4 cM)
Dist cM
Marker Name
11.8
16.1
7.2
11.3
M1032BK3
M1327AK3
M2103AK5
M0230BK5
L0326EK5
Dist cM
Marker Name
11.8
16.1
9.0
12.0
M1032BK3
M1327AK3
M2103AK5
L1131CK5
L0326EK5
M0230BK5
L0231AK2


LG-11 (48.4 cM)
KLG-11 (18.2 cM)
SLG-11 (10.9 cM)
Dist
cM
Marker
Name
4.8
13.4
M0929BK5
M0228AK5
M0711EK5
4.9
14.1
5.8
5.3
5.1
4.6
3.1
5.5
M0929BK5
M0228AK5
M0711EK5
L0210DCC
M0711FS3
L1702ACC
M1311CK5
M1311BS3
L2004CK2
5.5
5.3
L0210DCC
M0711FS3
L1702ACC

LG-12 (44.7 cM)

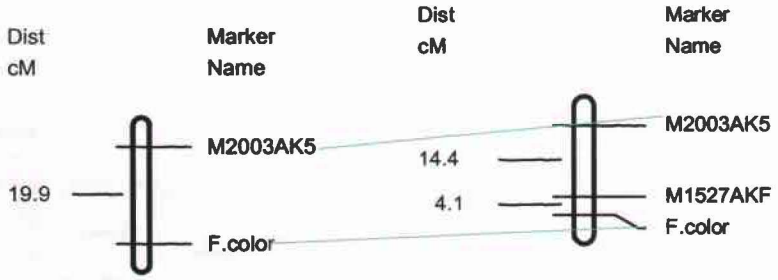


LG-13 (35.4 cM)



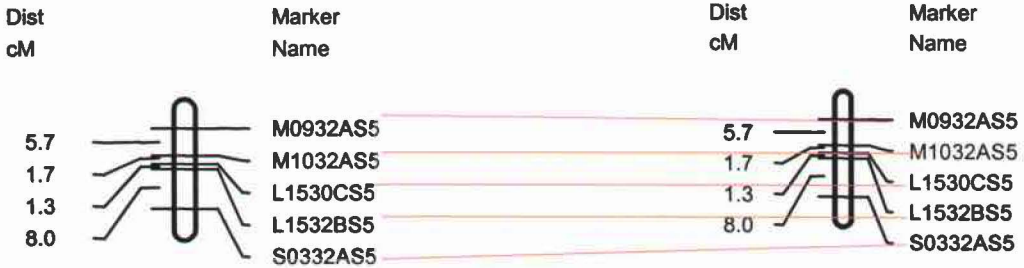
LG-14 (18.5 cM)

KLG- 14 (19.9 cM)



LG-15 (16.7 cM)

SLG-15 (16.7 cM)



Further-more, in LG-5, the distortion rate was highest at 60-96 cM from the tp of the linkage group (Fig. 2.12), and the markers at the ends of LG-5 followed Mendelian segregation ratio. The pattern of distortion rate in LG-5 is given in Fig. 2.13.

2.4.3.1.2 Marker distribution

Some of the linkage groups in which markers were ordered without LOD criteria as described in Materials and methods had non-random distributions of dominant marker alleles. There were regions in the linkage groups where consecutive dominant markers arise from the same parent. Further, some of the small linkage groups had dominant markers from only one parent. Assuming that the distribution of dominant alleles follows a random binomial model, the probability of obtaining such uniparental blocks by chance was very small for three Kapoho blocks and one Saipan Red block (Table 2.13). The number of uniparental markers ranged from 15-29 in those four blocks.

When the LOD score for ordering of markers was kept at 2.0, in mixed-phase maps, dominant markers were ordered in alternating as mixture of blocks of coupling-phase markers. Co-dominant markers appeared to be ordered as bridges between blocks of coupling-phase markers from different parents. Repulsion phase markers were added only at lower LOD score for ordering. The effect of higher LOD criteria on marker order in LG-5 and LG-6 is presented in Fig 2.14 and Fig 2.15. The ordering of markers was repeated five times to confirm the pattern of markers. Different attempts ended in different uniparental blocks (either Kapoho or Saipan Red) and this was related to, which randomly chosen markers were used as the starting framework.

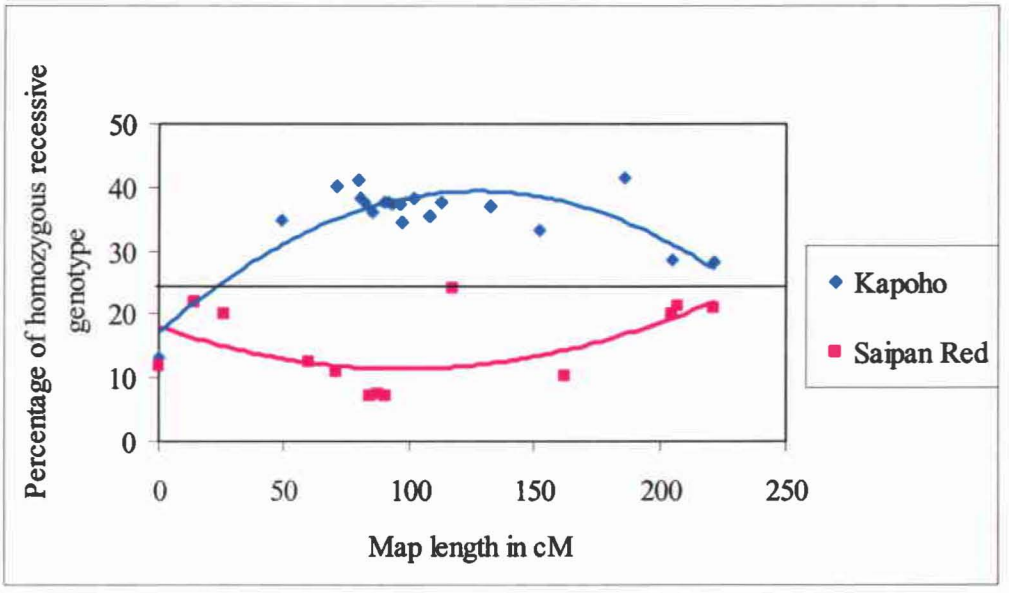


Fig. 2.13 Trend of distorted markers in LG-5.

——— Expected percentage of homozygous recessive genotypes in the population.

Regression curve for Kapoho homozygous recessive genotypes.

Regression curve for Saipan Red homozygous recessive genotypes.

Table 2.13 Details of blocks in mixed-phase map, where adjacent dominant markers arise from the same parent.

Linkage group	Total number of markers in the linkage group	No. of markers in the block	Block distance in cM	Probability of observing the block by chance
LG-3	104	19 (K)	55.9	0.0001
LG-4	32	16 (K)	93.7	0.0003
LG-6	46	29 (K)	93.9	< 0.0001
LG-2	71	15(S)	51.5	0.0011

K-dominant markers from Kapoho
S- dominant markers from Saipan Red

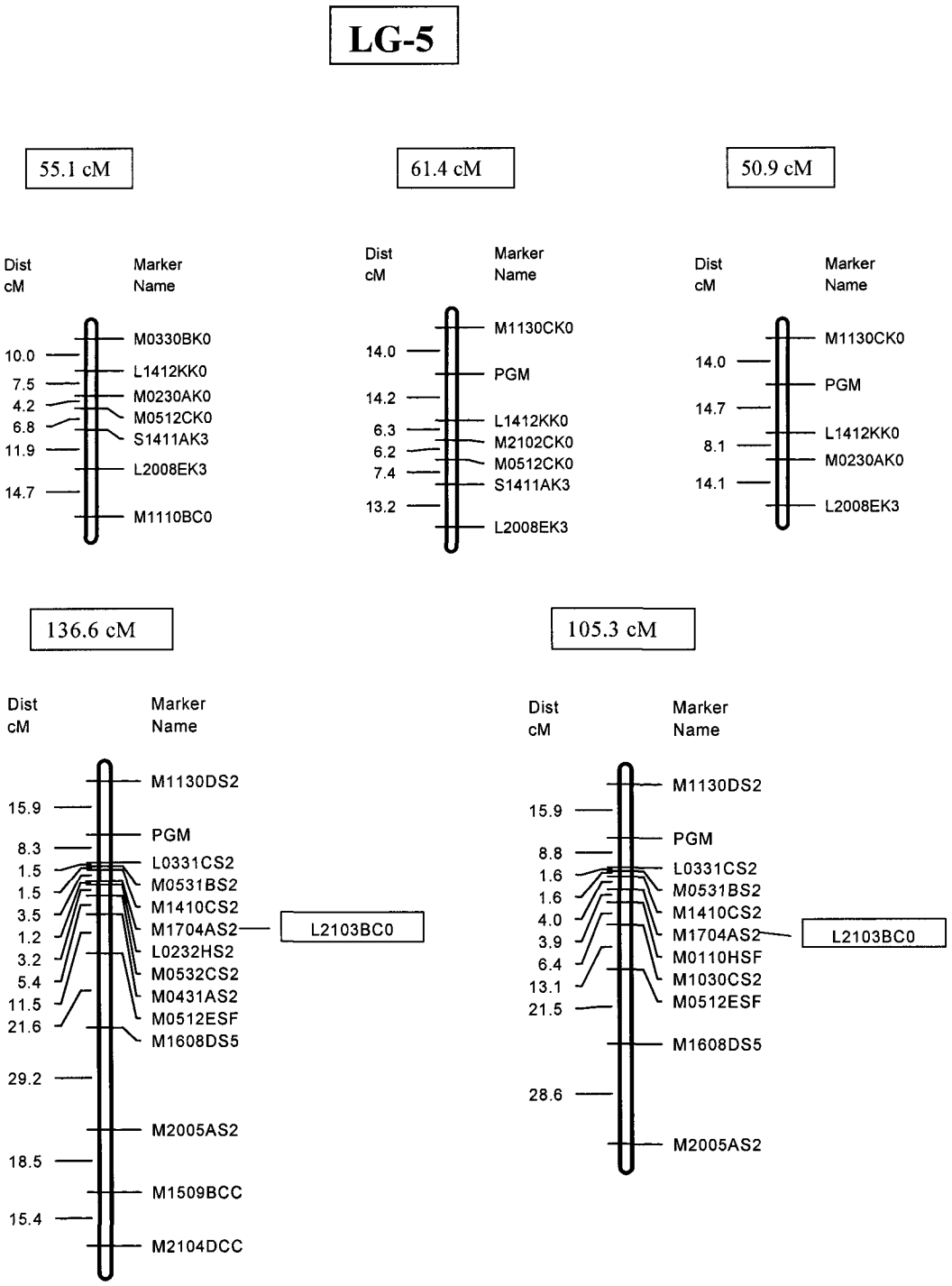


Fig.. 2.14 Five mixed-phase marker maps for LG-5 created from different initial markers chosen at a threshold of LOD 2.0 for ordering of markers.

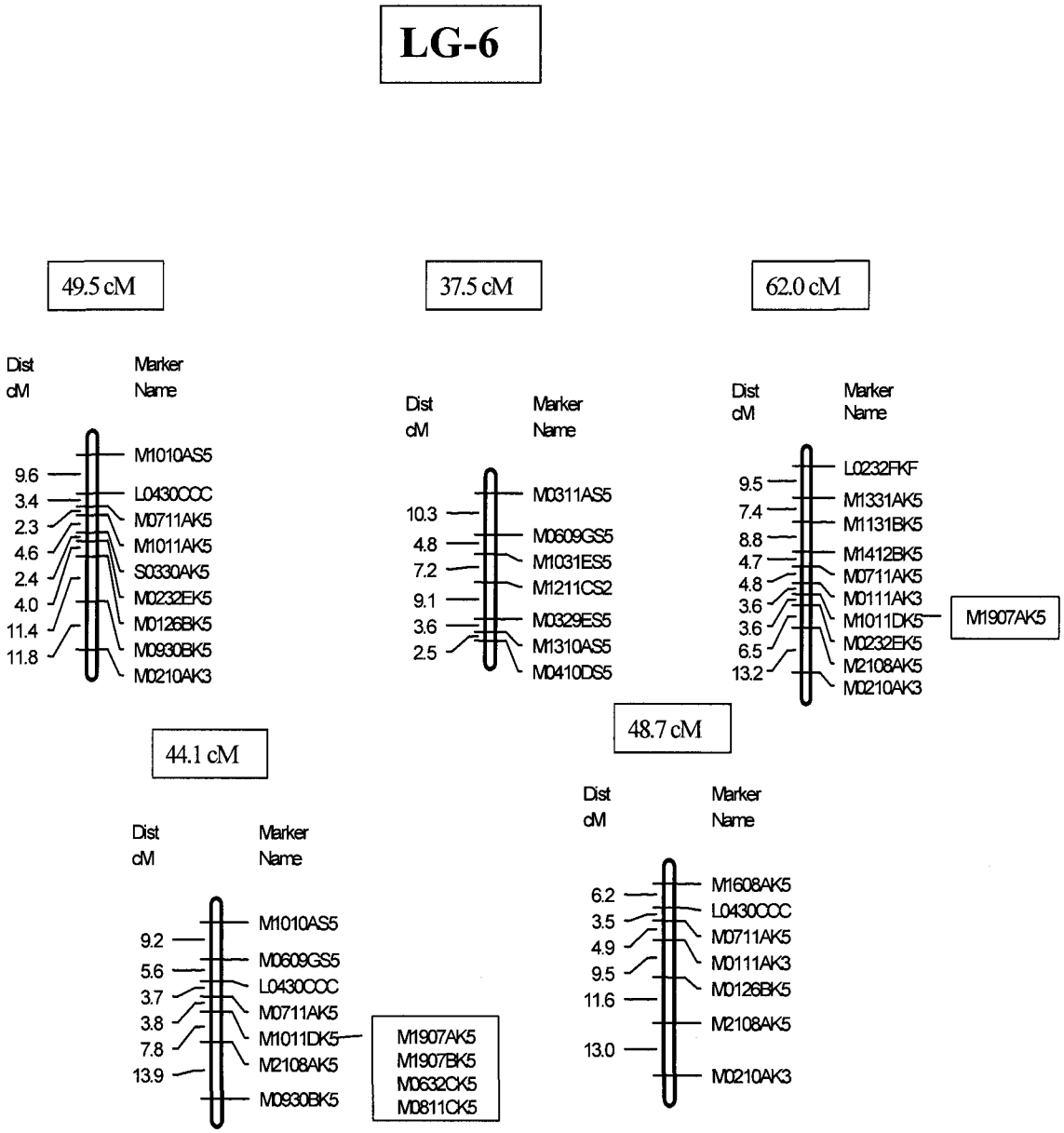


Fig.. 2.15 Five mixed-phase marker maps for LG-6 created from different initial markers chosen at a threshold of LOD 2.0 for ordering of markers.

2.4.3.1.3 Markers in common with Ma *et al.*'s map (20004)

Based on the primer sets used and the approximate size of the amplified fragments, 25 common markers were found between the present linkage map and a papaya linkage map created by Ma *et al.* (2004). LG-3 and LG-8 in the present map corresponded to LG-2 and LG-4 of the map created by Ma *et al.* (2004) respectively.

2.4.3.2 Coupling-phase maps

Although it is important to create a linkage map with maximum coverage of the genome, and to saturate the map with more markers, it is also important to create a reliable map for QTL analysis. It is usual to restrict QTL analysis to linkage maps created with framework markers, whose order meets strict statistical thresholds (Keats *et al.,* 1991). Because of the stringent statistical thresholds, only 222 markers were able to be mapped in the coupling-phase maps, as contrasted with 413 markers in the mixed-phase map. The Kapoho coupling-phase map included 121 markers, covering 778.8 cM of the genome with average interval of 6.4 cM in between the markers. Saipan Red coupling-phase map covered 693.0 cM of the genome with 101 markers with an average marker interval of 6.9 cM between two markers. The number of markers in each linkage group, map distances and average marker interval of each linkage group were given in the Table 2.14.

2.4.4 Mapping QTLs

2.4.4.1 Node number at floral conversion

Single marker analysis by linear regression produced 112 markers that were significantly (P=0.005) associated with node number at floral conversion (Table 2.14). Interval mapping using Mapmaker/QTL on coupling-phase maps produced two QTL peak on KLG-4 explained 64.2% of the phenotypic variability, and QTL peak on SLG-4b explained 23.8% of the phenotypic variability (Fig. 2.16 and Fig. 2.17, and Table 2.15) The above two QTL peaks are located adjacently in the mixed-phase map and, suggested that both QTL peaks were essentially a single QTL peak.

2.4.4.2 Carpellody

Single marker analysis, done by the nonparametric Kruskal-Wallis rank sum test (MAPQTL V 5.0), produced 23 markers that were significantly (P=0.005) associated with carpellody in Feb '02 and 36 markers that were significantly (P=0.005) associated with carpellody in Aug '02. Eleven markers showed significant association with carpellody in both of the seasons (Table 2.16). In linkage group LG-7, a series of markers was significantly associated with carpellody in both Feb '02 and Aug '02. The linkage groups LG-2 and LG-4 had multiple markers that were significantly associated with carpellody, in Aug '02 only. LG-3 had a region, where couple of markers showed significant association with the trait in both of the seasons and another couple of markers that showed significant association with the trait only in Feb '02. Even though phenotypic data of carpellody did not follow normal distribution, parametric interval

Table 2.14 Marker loci that showed association with node number at floral conversion at $P \leq 0.005$ in single marker analysis by linear regression method.

Marker	LG	P (Significance)	Marker	LG	P (Significance)	Marker	LG	P (Significance)
M2104EK5	4	<0.0001	M0329AS5	3	0.0002	M0832CK5	3	0.0012
M2106ES5	3	<0.0001	M1608BCC	3	0.0002	S1706AK5	2	0.0012
M0811BCC	4	<0.0001	L2007GK5	3	0.0002	S1230AK5	*	0.0013
M1411BS3	4	<0.0001	S0410AS3	3	0.0002	L2004BCC	4	0.0014
M0230DS2	7	<0.0001	L1029ES3	3	0.0002	M0609FS3	3	0.0014
M0729BS3	4	<0.0001	M0810BS3	4	0.0003	L0232GS3	3	0.0014
M0832AK2	4	<0.0001	M1329ACC	3	0.0003	L0432CS2	7	0.0014
M0832BS5	3	<0.0001	M1130DS2	5	0.0003	L3112BK5	2	0.0014
M0332CS5	3	<0.0001	L0210CS5	3	0.0003	M1210CS2	5	0.0015
M1604AS5	3	<0.0001	L0410ES5	3	0.0003	M2006HS5	3	0.0016
M1426CS5	3	<0.0001	L0329IS5	7	0.0003	M0110AK5	*	0.0017
M1426DS5	3	<0.0001	L0432DK5	2	0.0003	M0110BK5	2	0.0017
M1526AK5	9	<0.0001	M1130CK0	*	0.0003	L1010ES5	2	0.0020
M1526BS5	3	<0.0001	M1605BC0	*	0.0003	L1509DCC	7	0.0021
S0710AS5	*	<0.0001	M0411AS5	3	0.0004	L0631AS5	*	0.0021
L0409ES3	3	<0.0001	M0412CCC	2	0.0004	M1528AK5	2	0.0022
L1428BS5	3	<0.0001	M0330ECC	3	0.0004	L1209AK5	1	0.0023
L1412KK0	5	<0.0001	M2102BS5	3	0.0005	L1031HK3	3	0.0023
M0732AK0	*	<0.0001	M1109AS5	2	0.0005	M2104AS5	*	0.0024
M1031CS0	*	<0.0001	M1110AK3	*	0.0005	M1512CK5	2	0.0024
L0230IKF	*	<0.0001	M0810CS5	3	0.0006	L2104HK5	2	0.0024
M0611BS5	3	<0.0001	M0232BK5	2	0.0007	L1031IS5	3	0.0027
M2101CK3	2	0.0001	M1029AS5	3	0.0007	L1310CS5	2	0.0028
M1529FS3	3	0.0001	M0629CK5	2	0.0007	M1129BS3	4	0.0030
L0710IK3	*	0.0001	L0230GS5	3	0.0007	M1529BS5	3	0.0031
L0910DS5	3	0.0001	M0711BS5	3	0.0008	M1330DKF	*	0.0031
S0412AS5	3	0.0001	M1030AS5	3	0.0008	M1211ACC	2	0.0032
L0331BS5	3	0.0001	M0429CS5	7	0.0008	M1529CK5	*	0.0032
M1110BC0	5	0.0001	L0429FS5	3	0.0008	L0429GK5	2	0.0032
M0330BK0	5	0.0001	L1129CCC	3	0.0009	M0332DKF	2	0.0032
L0532DK0	5	0.0001	M2006AS5	3	0.0010	M0230AK0	*	0.0035
L0230JSF	*	0.0001	M1510AK5	*	0.0010	M1807BS5	*	0.0036
M2101BS5	2	0.0002	S1605AS3	4	0.0010	PGM	5	0.0038
M2106BS5	3	0.0002	M1030BK2	4	0.0011	L0710GS5	*	0.0038
M1510BS5	3	0.0002	L2010IK5	3	0.0011	M0326CK5	4	0.0039
M1412FK3	*	0.0002	L0228BK5	4	0.0011	M1031BK5	3	0.0044
M1110CS5	3	0.0002	M1111BK3	3	0.0012	M0512CK0	*	0.0049
						M1526FK5	2	0.0050

*-marker unlinked to any linkage group

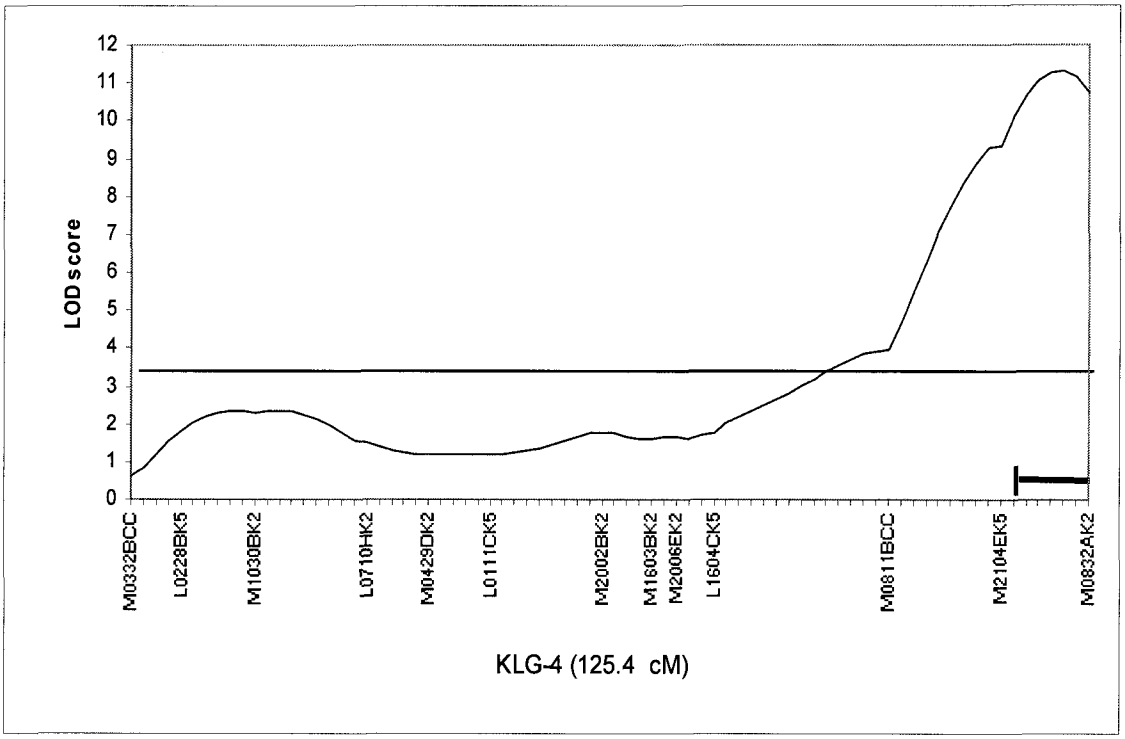


Fig. 2.16 Interval mapping for node number at floral conversion on linkage group KLG-4. On X-axis, 1 unit = 2 cM. ▬ Support interval within a LOD difference of 1.0 from the peak.

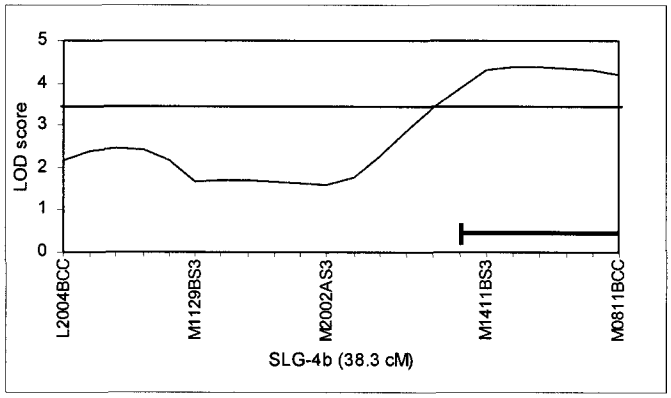


Fig. 2.17 Interval mapping for number of nodes at floral conversion on linkage group SLG-4b. On X-axis, 1 unit = 2 cM. ▬ Support interval within a LOD difference of 1.0 from the peak.

Table 2.15 Characteristics of QTLs for node number at floral conversion as identified by interval mapping.

Linkage group	Interval	LOD score	Phenotypic variation explained	Additive	Dominance	Gene action
KLG-4	M2104EK5-M0832AK2	11.31	64.2%	-5.6169	-0.3483	Additive
SLG-4b	M0811BCC-M1411BS3	4.38	23.8%	-3.6467	-0.2376	Additive

Table 2.16 Marker loci that showed association with carpellody at $P \leq 0.005$ in single marker analysis by Kruskal-Wallis rank test.

Feb '02				Aug '02			
Marker name	LG	K	P (significance)	Marker name	LG	K	P (significance)
M0232BK5	2	7.900	0.005	M1412ES5	*	7.886	0.005
M1526EK2	2	8.064	0.005	M1409AS3	*	7.902	0.005
M1011BS2	2	8.141	0.005	M1029BSF	*	7.981	0.005
M0326DK5	2	8.231	0.005	L1010ES5	2	8.104	0.005
M1528AK5	2	8.492	0.005	M1530AS5	1	8.124	0.005
M0332DKF	2	8.566	0.005	M0609CK2	2	8.250	0.005
L0232FKF	*	8.598	0.005	M1028CS5	2	8.361	0.005
M0410DS5	6	8.670	0.005	L0232GS3	3	8.428	0.005
L0232GS3	3	8.734	0.005	M0611AK5	7	8.486	0.005
M1329BK3	7	8.862	0.005	M0326DK5	2	8.502	0.005
M2001BSF	3	8.919	0.005	M2104EK5	4	8.622	0.005
M2003BK5	2	9.179	0.005	M2105BS5	3	8.639	0.005
M1030AS5	3	9.301	0.005	L0128BS5	1	8.672	0.005
M1526FK5	2	9.424	0.005	L1409ES5	*	8.785	0.005
M0911CS5	2	9.464	0.005	M1526FK5	2	8.878	0.005
L0325BK3	7	9.627	0.005	M1230BS5	*	8.913	0.005
M2007AK5	9	9.978	0.005	M0326AS3	*	8.924	0.005
M1011CK5	7	10.194	0.005	M0911CS5	2	9.125	0.005
M2105BS5	3	10.403	0.005	M2003BK5	2	9.159	0.005
M0609CK2	2	10.672	0.005	M0110AK5	*	9.295	0.005
M0611AK5	7	13.622	0.001	M1512AS5	2	9.363	0.005
M0432AK5	7	15.156	0.0005	S1028AK5	2	9.548	0.005
L0409DK5	7	18.303	0.0005	M0526AS5	*	9.627	0.005
				M0110BK5	2	9.703	0.005
				M0729BS3	4	10.218	0.005
				L0230HSF	2	10.249	0.005
				M0432AK5	7	10.685	0.005
				M0811BCC	4	10.708	0.005
				L1010DK5	*	11.589	0.001
				S1605AS3	4	11.684	0.001
				L0230IKF	*	11.836	0.001
				M0130AK5	*	11.920	0.001
				M2007AK5	9	12.182	0.0005
				M1129BS3	4	12.507	0.0005
				L2004BCC	4	16.589	0.0005
				L0409DK5	7	15.427	0.0001

* - marker unlinked to any linkage group

mapping was tried. Many LOD score peaks with very high LOD score in the range of 20 to 30 were produced. The markers under most of these LOD score peaks did not show significant association with carpellody in single marker analysis. However, two QTL peaks were linked with markers shown to have significant association with carpellody by single marker analysis. For these two QTL peaks (Fig. 2.18, Fig. 2.19), at least, the location of the QTL might be real (Table 2.15). The QTL peak on KLG-7 was not influenced by the environment, as it had been expressed in both Feb '02 and Aug '02. At this QTL peak, Saipan Red allele had a positive effect on the trait in both of the seasons. At KLG-2a, carpellody had shown two QTL peaks in Aug '02.

2.4.4.3 Fruit weight

Single marker analysis by linear regression method (QTL Cartographer) showed five markers as significantly associated (P = 0.005) with fruit weight (Table 2.16). Out of the five markers, two fell on LG-3 with 13.8 cM distance between them. LG-2 and LG-7 had one significant marker each, and one more marker was unlinked to any group. Though the broad sense heritability (H^2) was 74.0%, in interval mapping, no QTL peak was produced above the LOD score 3.3, the threshold determined by 1000 permutation.

2.4.4.4 Relationship between quantitative and qualitative traits

The phenotypic observation of association between PGM locus and number of nodes to flowering, and the association between carpellody and flesh color were not confirmed in QTL mapping.

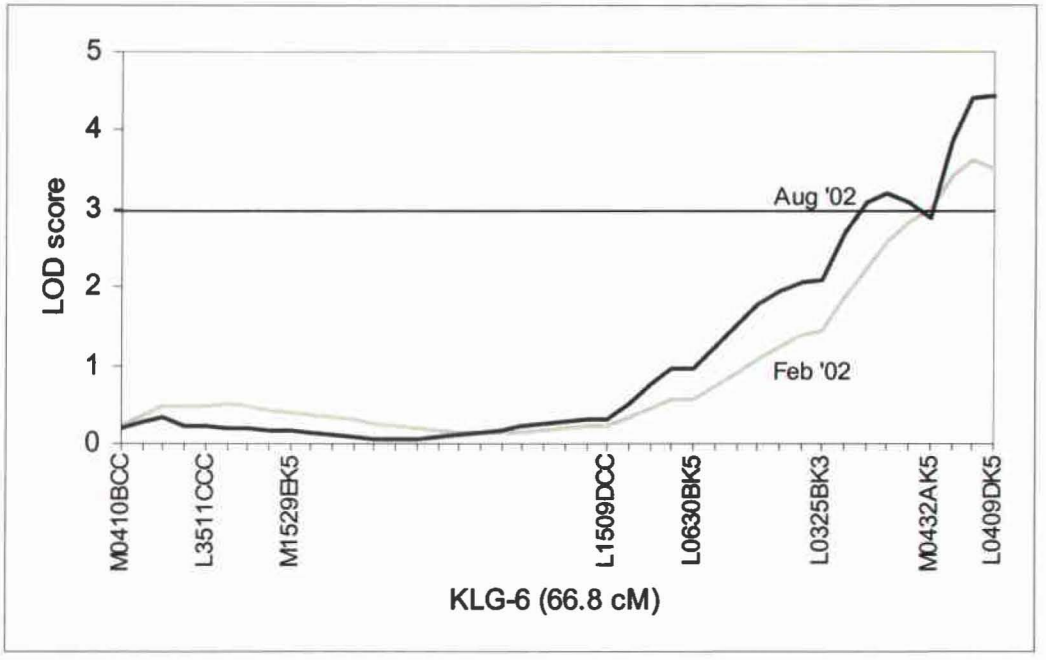


Fig. 2.18 Interval mapping for carpellody on linkage group KLG-7 during Feb '02 and Aug '02. On X-axis, 1 unit = 2 cM.

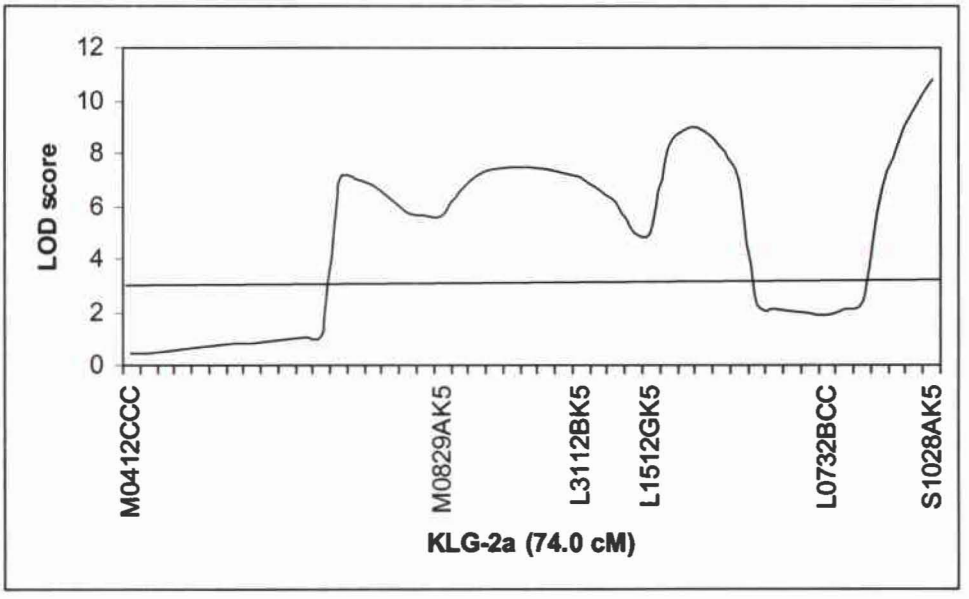


Fig. 2.19 Interval mapping for carpellody on linkage group KLG-2a during Aug '02. On X-axis, 1 unit = 2 cM.

Table 2.17 Characteristics of QTLs for carpellody as identified by interval mapping.

Season	Linkage group	Interval	LOD score	Phenotypic variation explained	Additive	Dominance	Gene action
Feb '02	KLG-7	M0432AK5-M0409DK5	3.60	13.2%	0.213	-0.339	Recessive
Aug '02	KLG-7	M0432AK5-M0439DK5	4.43	29.3%	0.395	-0.873	Recessive
Aug '02	KLG-2a	L0732BCC-S1028AK5	11.57	68.1%	-0.916	-0.900	Dominant

Table 2.18 Marker loci that show association with fruit weight at $P \leq 0.005$ in single marker analysis by linear regression method.

Marker	LG	P(Significance)
M0711BS5	3	0.0013
L0409DK5	2	0.0016
S0412BS5	*	0.0024
S1028AK5	7	0.0040
M0332CS5	3	0.0044

* - Unlinked to any linkage group

2.5 Discussion

2.5.1 Genetic linkage map

A linkage map gives the genetic architecture of the plant by arranging markers and linked genes in a linear order. Linkage maps assist in locating genes controlling qualitative traits through proximity to co-segregating markers, and also form the framework for QTL analysis. We developed a linkage map for an F_2 population (N=281) of papaya using 513 AFLP markers, an isozyme (PGM) marker and a morphological marker (flesh color). The map consisted of 15 linkage groups covering 2066 cM of the genome with an average interval of 5.0 cM between the markers. Since dominant markers in F_2 population may cause problems in ordering of markers (Knapp, 1995; Mester *et al.*, 2003a), using co-dominant markers as common markers, we created two coupling-phase maps, one for AFLP markers dominant in Kapoho and another for AFLP markers dominant in Saipan Red.

2.5.1.1 AFLP polymorphism

The parents used in the investigation, Kapoho and Saipan Red, have contrasting characteristics for the traits that we studied. Kapoho is a commercial variety from Hawaii, and Saipan Red is a breeding line obtained from Saipan in the Northern Mariana Islands and maintained at the University of Hawaii. The parents originate from different geographical areas, and are fairly polymorphic for AFLP markers, with an average of 3.3 polymorphisms per primer pair, higher than the 1.8 polymorphisms per primer pair reported previously in papaya by Ma *et al.* (2004). This heterozygosity may result

because the parents used in the present study are genetically more diverse than the parents used in Ma *et al.*'s study. In a study of genetic similarity based on pair-wise simple matching coefficients (Kim *et al.,* 2002), Kapoho and Saipan Red, the parents used in the present investigation, showed genetic similarity of 0.88, while the parents used in Ma *et al.*'s study showed genetic similarity of 0.90. Though AFLP usually generates dominant markers, allelic differences can occasionally be identified as co-dominant markers. The percentage of co-dominant markers in the current study (7.4%) was similar to the previous report (7.9%) in papaya (Ma *et al.*, 2004).

2.5.1.2 Genome coverage

In the present study, the mixed-phase map of papaya (genome size is 372 Mbp, Arumuganathan and Earle, 1991) covers 2066.6 cM, while in crops with bigger genome sizes like tomato (907-1000 Mbp), peas (3947-4397 Mbp) and *Brassica oleraceae* (603-662 Mbp), shorter map lengths have been reported (Zhang *et al.*, 2002; Prioul *et al.*, 2004; Rocherieux *et al.*, 2004). This suggests that the papaya map length was inflated. Similar map length inflation had been previously reported in the papaya linkage map produced using AFLP markers (Ma *et al.,* 2004).

One of the reasons for inflation of genetic distances in maps developed from linkage analysis of dominant markers in F_2 populations is that dominant markers in repulsion phase lead to biased recombination estimation (Knapp *et al.*, 1995; Mester *et al.*, 2003a). When dominant markers are used in F_2 populations, the precision of marker

order within the linkage group is less reliable (Knapp *et al.*, 1995, Korol *et al.*, 2003). Incorrect marker order leads to spurious recombination and inflation of map lengths (Sybenga, 1996). Though the map length covered by mixed-phase map was 2066 cM, when the markers were divided into coupling-phase maps, the total map length covered by Kapoho and Saipan Red maps individually were 778.8 cM and 693.0 cM, respectively. No specific threshold was used to limit the inclusion of markers in the linkage group while ordering the markers in mixed-phase maps. However, in the coupling-phase maps, unless a marker had a well-supported position (LOD 2.0) in the map, it was not included. The difference in the threshold level employed for ordering markers in the mixed-phase versus the coupling-phase maps was probably responsible for most of the map length inflation in the mixed-phase map. In Fig. 2.12, the position of coupling-phase framework markers had been indicated in the mixed-phase map. The figures show that the order of markers in the mixed-phase map was not co-linear with the order in the coupling-phase map. The figure also showed that the distance between terminal markers in mixed-phase maps has been inflated relative to the coupling-phase map. Adding to the inflation was the Kosambi mapping function used by the Mapmaker software in the present study, which had been reported to increase the recombination fractions compared to what would be explained based on meiotic chiasma frequencies (Sybenga, 1996).

Though papaya has only nine haploid chromosomes, the present map has 15 linkage groups. This was because the present map was incomplete. The presence of long (> 50 cM) regions in chromosomes that were low in polymorphisms would result in

depicting single chromosome as more than one linkage group. In our coupling-phase maps, LG-4 of Saipan Red linkage group was separated into smaller groups because of the presence of a long stretch of chromosome consisting entirely of dominant markers contributed in coupling-phase by the Kapoho parent.

2.5.1.3 Conservation with previous map

Out of 158 primer pairs used in the present study, 138 were in common with the primer pairs used by Ma *et al.* (2004) in papaya. These 138 primer pairs produced 429 markers in the present map, of which only 25 markers (5.8%) showed a common fragment size and co-segregation pattern with Ma *et al.*'s map (2004). LG-3 in the present map and LG-2 in Ma *et al.*'s (2004) map shared 22 common markers. Similarly, LG-8 in the present map and LG-4 in Ma *et al.*'s map shared 3 common markers. Twenty- nine additional markers from the present study seemed to appear in Ma *et al.*'s work (2004), based on fragment size of the marker and the primer sets used, but the linkage relationships did not correspond to those in Ma *et al.*'s map (2004). As shown in Appendix B, the markers that co-segregate into the same linkage group in our map, segregated into different linkage groups in the map of Ma' *et al* (2004). Hence, these 29 markers may not really be common to both maps. Since AFLP markers are generated without prior knowledge about the sequence, the same fragment size generated by the same primer set may not necessarily indicate the same nucleotide sequence in different F_2 populations. Thus, AFLP markers may not be a robust method to study different populations of the same species.

In the present map, LG-3 showed a cluster of 15 co-segregating markers at locus M2106BS5. LG-2 in Ma *et al.*'s map (2004), which corresponded to LG-3 in the present map, also has two large clusters of markers consisting of 56 and 77 co-segregating markers. The clusters in Ma *et al.*'s map are separated by 18.9 cM. This region showed many co-segregating markers and could be the centromeric region of the chromosome. Since recombination in the centromeric region is considerably suppressed (Paterson, 1996), it is difficult to resolve loci in our small population (n=84) and the markers appear to co-segregate at the same locus.

2.5.1.4 Non-random distribution of dominant marker alleles

A notable feature of the mixed-phase linkage maps was the occurrence of regions, where contiguous dominant markers arose from the same parent rather than randomly from both parents. Also in the present study, smaller linkage groups consisted only of dominant markers originating from one parent. The assumption that either parent was equally likely to contribute the dominant allele at any particular marker locus lead to the conclusion that it would be highly improbable to find many consecutive markers with dominant alleles from a single parent. In the present map, four blocks with 15-29 consecutive markers with dominant alleles from the same parent were observed. Assuming a random binomial model (Burnside, 1959), the probability of observing such large blocks by chance ranged from <0.0001 to 0.0011. The papaya linkage map, created by Ma *et al.*(2004) also has two similar regions. In that map, LG-5 has a stretch of 25 contiguous markers from the parent Kapoho. In LG-9, another stretch of 13 contiguous

dominant markers arising from Kapoho was observed. The probability of observing such blocks on LG-5 and LG-9 by chance are <0.0001 and 0.0022 respectively.

A specific reason for the occurrence of these blocks was unknown. The fact that such low-probability blocks existed in multiple linkage groups in two different maps strongly suggested that chance alone was not the cause. These uniparental blocks might reflect same underlying biological mechanism, or (more probably) they may be artifacts of the algorithm employed by Mapmaker. A parent lacking any dominant form of AFLP marker over a long stretch of linkage group could result only under two circumstances: a) inability of restriction enzymes to cleave DNA at restriction sites in that particular stretch of chromosome, or b) inability of PCR primers to amplify AFLP fragments in that particular stretch of chromosome. Though specific methylation at restriction sites had been reported to prevent restriction in bacteria (Rocha *et al.*, 2001), no such protective mechanism had been reported in plants. Failure of amplification of restricted fragments arises when there is mutation in the binding region of the extension nucleotides in the primer, for example, ACC in the primer E-ACC. But mutation in the extension nucleotides binding site of all primer sets in a stretch of chromosome was unexpected.

When Mapmaker was used to map dominant markers in a F_2 population, a stringent LOD criterion for ordering (≥2.0), would restrict the software's algorithm to the selection of markers in coupling-phase only (Fig. 2.13 and Fig. 2.14). If there were some co-dominant markers in the group, blocks of coupling phase markers from different parents were connected by co-dominant markers. When we ordered markers in the

mixed-phase linkage maps, we reduced the LOD score for ordering gradually from LOD 5 to LOD 0.01. In this process, coupling phase markers were ordered first, and then at lower LOD scores, repulsion phase markers were added. This might have led to the stretches of consecutive markers with dominant alleles from the same parent.

2.5.1.5 Distorted markers

Another interesting observation in the present study was the pattern caused by the distribution of markers with distorted segregation ratios. In the present study, 21.6% of the markers showed segregation distortion. Almost half of the distorted markers (46.7%) fell into two linkage groups - LG-4 and LG-5. In LG-4, 13 markers showed segregation distortion, and all the distorted markers were deficient in Kapoho alleles. In LG-5, 23 out of a total of 30 markers were distorted and the distortion followed a trend within the linkage group. All the distorted markers in LG–5 were deficient in Saipan Red alleles. Similar unidirectional distorted segregation has been reported in interspecific crosses of many crops including *Brassica* (Landry *et al.*, 1992), tomato (de Vicente and Tanksley, 1991), *Mimulus* species (Fishman *et al.*, 2001), *Populus* species (Yin *et al.*, 2004) and an intraspecific cross in *Medicago* (Jenczewski, 1997; Thoquat *et al.*, 2002). Segregation distortions can be caused by both pre-zygotic and post-zygotic factors. While many possible reasons for segregation distortion have been reported including pollen-pistil incompatibility (Diaz and MacNair, 1999), meiotic drive (Lyttle, 1991), inbreeding depression (Remington and O'Malley, 2000), epistatic interaction between gametophytes (Fishman *et al.*, 1951) and selection for specific alleles (Launey and Hedgecock, 2001;

Yin *et al.*, 2004), we felt that the segregation distortion in LG-5 could have arisen from post-zygotic, positive selection for Kapoho alleles. In a linkage map created from a back cross of *Populus* species, Yin *et al.* (2004) noticed that most of the markers in a chromosome were distorted and favored the donor parent allele. Yin *et al.* claimed that the segregation distortion resulted from natural selection for leaf rust resistant alleles contributed by the donor parent. When the *Populus* plants were young, they were affected by leaf rust. The donor parent was resistant to leaf rust and the recurrent parent was susceptible to the disease. Two genes conferring leaf rust resistance were located on the chromosome which later showed distorted segregation. We speculate that a similar situation might exist in the present papaya study. When the F_2 population was in the seedling stage, some unknown disease killed a significant number of the seedlings. It was possible that a locus conferring resistance to that disease was located in LG-5, and Kapoho carried the resistance alleles and Saipan Red carried the susceptible alleles. Since segregation distortion rate was high in the middle of the chromosome and decreased gradually towards the ends, segregation distortion factors might be located in between 60 and 96 cM from the top end of LG-5 (mixed-phase map). Since a single differential fitness locus cannot affect loci farther than 50 cM, we think, there were at least a couple of differential fitness loci situated in between 60 and 96 cM from the top end of LG-5 (mixed-phase map). These loci affected nearly the whole linkage group, but the farther the marker was from these loci, the less distortion was observed. At both ends of LG-5, marked by S1411AK3 and M1509BCC, segregation started to follow normal Mendelian ratios.

2.5.2 QTL map

QTL maps are useful for finding the number of important genetic loci involved in the expression of quantitative traits, chromosomal locations of these loci, and the relative contribution of each locus to the trait expression (Stuber, 1992). In this study, we have identified a single QTL affecting the node number at which flowering occured and 4 linkage groups significantly associated with the expression of carpellody.

2.5.2.1 Node number at floral conversion

Through interval mapping, two significant QTL peaks were found for node number at floral conversion, one on KLG-4 and another on linkage group SLG-4b. Since markers under these peaks showed significant association with the trait in single marker analysis also, QTL peaks on KLG-4 and SLG-4b were highly reliable. The QTL on KLG-4 explained 64.2% of the phenotypic variability. Since the plants showed extremes of this trait were deliberately included in the mapping population, the LOD score for the QTL peak and the percentage of phenotypic variability explained were inflated. The QTLs on both KLG-4 and SLG-4b showed negative effects on the trait. Substitution of a Kapoho allele by a Saipan Red allele decreased the node number at which conversion of the axillary meristems occurs, which was in accordance with the low-bearing nature of Saipan Red. Both of the QTLs showed additive gene action, which supported the earlier report by Subadrabandhu and Nontaswatsri (1997). Their study involving three parents (Khaek Dam, Tainung #5, and Eksotika #20) in a diallel analysis revealed that the gene action for height to first flower is additive.

Since the QTL peaks from the coupling-phase maps were located in essentially the same region of the mixed-phase map, it was likely that there was only a single QTL on LG-4. Though a single QTL peak was unexpected for quantitative traits, Blum *et al.* (2003) have reported single QTL for capsaicinoid content in peppers. The QTL on LG-4 should have had a significant effect on the expression of node number at floral conversion. Support interval for the QTL on LG-4 fell between the markers M2104EK5 and M0832AK2, which were separated by 13.2 cM. Since these two markers were linked to the Saipan Red allele in repulsion phase, these markers can be used to select seedlings that were homozygous for Saipan Red allele, which reduced the node number at floral conversion. Since the above two markers were not present in the papaya map created by Ma *et al.* (2004), the markers may be useful only in this population.

2.5.2.2 Carpellody

The nonparametric Kruskal-Wallis test has been previously used to identify QTLs in similar situation of ordinal phenotypic data in tomato for bacterial canker resistance (Van Ooijen, 1993) and in Asiatic lily for *Fusarium* resistance (Van Heusden *et al.*, 2002). Since carpellody was measured on an ordinal scale, and the phenotype was not normally distributed, Kruskal-Wallis test was used for single marker QTL analysis. In the Kruskal-Wallis test, 48 markers showed significant association with carpellody, of which 36 markers were located in linkage groups. The other 12 markers (25%) were unlinked to any linkage group. This was in accordance with the expectation, because 20% of all the markers were unlinked. All but 4 of the 36 mapable markers fell into non-random clusters

on four linkage groups (LG- 2, 3, 4 and 7). The QTLs on LG-2, 3 and 7 were associated with carpellody in both seasons (Feb'02 and Aug'02), whereas QTLs on LG-4 and part of LG-2 influenced the trait only in Aug'02. The QTLs on LG-4 and part of LG-2 were largely responsible for the more carpellodic phenotypes seen in Aug'02. For fruit that we scored for carpellody in Aug'02, the flower would have opened in Mar'02. Stamen differentiation in papaya begins eight weeks before anthesis, and ovary differentiation occurs seven weeks before anthesis (Arkle and Nakasone, 1984). Thus, transmutation of stamens to carpels would have occurred seven months before the fruit was scored in the field. Carpellodic transmutation of stamens would have occurred in July 2001 and January 2002 for fruit that was scored in the field during Feb'02 and Aug'02 respectively. The temperature during January 2002 (Mean maximum 76.68° F and mean minimum 64.52° F), was cooler compared to July 2001 (Mean maximum 64.52° F and mean minimum 70.61° F) and low temperature is reported to increase the incidence of carpellody (Awada, 1958). Further, the total rainfall in January 2002 (6.11 inches) was higher than that in July 2001 (0.44 inches), which would have lead to higher soil moisture content in January 2002. Higher soil moisture content has been reported to cause carpellodic expression (Awada and Ikeda (1957). Hence, low temperature and high soil moisture might be necessary for the expression of QTLs on LG-1, LG-4 and part of LG-2. Two clusters of markers showed significant association with carpellody on LG-2 and LG-7 were located on blocks of Kapoho dominant markers in coupling-phase. Whether coupling-phase block influences carpellodic expression or the observation was just a coincidence was unknown.

When the phenotypic data had a significant number of individuals in the zero category (in the ordinal scale), maximum likelihood estimation has a tendency to produce spurious LOD peaks in regions between widely spaced markers (Broman *et al.*, 2003). Zou *et al.* (2003) showed that the power of detection of QTL for non-normally distributed data is higher for rank-based tests compared to maximum likelihood or regression methods. Nevertheless, interval mapping was attempted to check the power of parametric testing in finding real QTL for carpellody, apart from spurious QTL peaks. Interval mapping of carpellody data using maximum-likelihood method resulted in many QTL peaks with LOD scores of most in the range of 20-30, but the markers under the peaks did not show significant association with the trait by the nonparametric test. Besides these doubtful QTLs, QTL peaks were also identified on two of the five blocks of markers identified as significant by the nonparametric test. The LOD peak on KLG-7 was significantly expressed in both of the seasonal evaluations, while the QTL peak on KLG-2a was present only in Aug '02. The markers under these QTL peaks showed a similar significant pattern in the nonparametric test. At the QTL on KLG-7, alleles from Saipan Red increased the expression of carpellody whereas at the QTL peak on KLG-2a, Saipan Red alleles reduced carpellody, which contradicts its carpellodic nature. However, alleles affecting the progeny in a manner contrary to the phenotype of the contributing parent have been reported for this trait by Sondur (1994). Sondur reported three QTLs for carpellody and at one of these, the allele effects were opposite to those expected on the basis of the parental phenotypes. Allelic effect of line-356, the parent, which had the factors for carpellody, decreased the expression of carpellody at QTLX17-7 on LG-7.

The phenotypic distribution of carpellody in Sondur's work (1994) was not normal and the log transformation did not alter the distribution appreciably. Sondur used only parametric methods, which assumes that the distribution is normal. There was no difference in the results whether or not he included the "no carpellodic fruit category", and he got very high percentage of phenotype variation explained by the 3 QTLs (96%). His results, while perhaps not totally incorrect, have to be treated with caution.

2.5.2.3 Fruit weight

Contrary to the results of Sondur (1994), who identified three QTLs for fruit weight, the current study did not produce any QTL peak for fruit weight by interval mapping. In single marker analysis, only five markers were significantly associated with the trait. The broad sense heritability (H^2) of 74% implies that the genetic component of fruit weight inheritance was significant. The reason for not seeing any QTLs for fruit weight in Kapoho X Saipan Red F_2 population could be that the trait was controlled by many polygenes with minor effects, which act in additive fashion. The population size of our analysis may also be too small for resolving minor QTLs.

2.5.2.4 Pleiotropism

In linkage group LG-4, the markers M0811BCC and M2104EK5 under the QTL peak for node number at floral conversion were also in a stretch of markers, that showed significant association with carpellody in Aug'02. Though no pleiotropism controlling these two traits had been reported earlier, in the present study, both F_2 population and

genotyped population showed significant correlation (P=0.05) between node number at floral conversion and carpellody during Aug'02. Neither F_2 nor genotyped population showed significant correlation between these traits in Feb'02. The QTLs that affect carpellody were spread over multiple chromosomes confirmed that carpellody is a polygenic trait. The carpellody factor located on LG-4 may also show that pleiotropism affected the node number at floral conversion.

2.5.2.5 Relationship between quantitative and qualitative traits

Though the phenotypic observation of association between PGM and node number at floral conversion was supported by single marker analysis, interval mapping did not show any QTL for node number at floral conversion on LG-5. It was not clear whether segregation distortion of markers affects the identification of QTL on LG-5 or whether there was no QTL for node number at floral conversion on LG-5. The association between flesh color and carpellody was not significant by nonparametric statistical testing, which implied that the association at the phenotypic level was possibly accidental.

CHAPTER 3

FINDING MARKERS LINKED TO *DIMINUTIVE* MUTANT OF PAPAYA

3.1 Introduction

Papaya shows a number of mutations that produce changes in the phenotype of leaves, flowers and fruits (Storey, 1953). Mutations that show morphological changes are usually recessive in nature and often have deleterious effects on fitness (Tanksley, 1980). While mutations like changes in fruit-flesh color are desirable in breeding, some of the mutations are undesirable. One of the undesirable mutants that has been described in papaya is *diminutive*, a recessive phenotype, that is characterized by small stature of the tree with small fruits (Storey, 1953). A mutantion similar to diminutive occurred in the variety Sunset in the early 1990's and was accidentally carried into the genetically engineered, papaya ringspot virus-resistant SunUp variety, (heterozygous for *diminutive* locus) was used as one of the parents in a recurrent selection breeding program at the University of Hawaii and *diminutive* was carried into those populations. Since this mutant is completely recessive, it is difficult to detect in heterozygous condition when making selections, unless an extra generation of inbreeding is resorted to. Identification of a marker linked in coupling-phase to *diminutive* allele could help the papaya breeders to identify the mutant in both homozygous and heterozygous condition. The following study used bulked segregant analysis (BSA) to identify an AFLP marker linked to the normal phenotype in coupling-phase.

3.2 Literature Review

3.2.1 Mutants of papaya

The mutants related to the structure of the tree and leaf phenotypes of papaya as described by Storey (1953), are:

d-Dwarf plant: This recessive trait is first reported by Hofmeyr (1949) and is characterized by excessive branching of the tree.

dp-Diminutive plant: These plants are recessive to normal trees. The trees are characterized by short and thin trunk, small leaves with short, slender petioles, small flower and fruit.

cp-Crippled leaf: These are recessive to normal leaf. Leaves appear as if they are infected by papaya mosaic virus.

rp- Rugose leaf: This is another recessive mutant characterized by upward puckering of blade areas between veinlets and by various degrees of curling of the margins.

w-Wavy leaf: These are recessive to normal leaf.

The mutant observed in Sunset and SunUp population showed loss of anthocyanin in the petioles and showed corky bark at the soil level region, in addition to the symptoms of diminutive mutant.

3.2.2 Bulked segregant analysis (BSA)

Unless many traits are investigated, finding markers through linkage mapping is unnecessarily laborious and time consuming. Bulked segregant analysis (BSA),

developed by Michelmore *et al.* in 1991 helps to find polymorphisms linked to specific regions of the genome, where the trait of interest resides.

In developing markers linked to a specific Mendelian trait, DNA is bulked from plants at the extremes of the phenotypic distribution of the trait in a segregating F_2 or backcross population. Since each bulk is uniform for a different genotype of the trait under study, yet segregates for all other unlinked loci, the bulks will show a clear difference (presence or absence of the band) only for primers that amplify polymorphic region linked to the trait of interest. On all other sites, both of the bulks will show the band. In a bulk containing different genotypes at a single locus, when the rarer allele constitutes 20-40% of the total DNA, it is detectable (Michelmore *et al.*, 1991). If two plants in a bulk of 10 plants carry the allele in coupling phase with the dominant form of the marker, the band will be detectable in the gel. AFLP combined with BSA had been previously used to identify an AFLP marker linked to a recessive reverse thermo-sensitive genic male sterile gene in rice (Jia *et al.,* 2001). BSA together with AFLP is not only used in finding markers linked to simple traits, but in some cases, may also be useful to find markers linked to QTLs. Lehmensiek *et al.*(2001) identified markers linked to QTLs for gray leaf spot resistance genes in maize. The markers identified with the QTLs were converted to RFLP and SSR markers and were found to segregate in different populations in different seasons. Hence, these markers can be used to identify gray leaf spot resistant plants in all seasons.

3.3 Materials and methods

3.3.1 Populations used in the study

Two different populations were used in this investigation. One population consisted of SunUp spontaneous *diminutive* mutants and normal plants. The other population was a recurrent selection population segregating for *Diminutive* locus, which used a SunUp parent that was heterozygous for this mutant locus.

3.3.2 DNA extraction and quantification

DNA was extracted from 80 mg of fresh leaf with DNeasy plant mini kit from Qiagen Inc (Valencia, USA) as per the instructions of the manufacturer. The DNA was separated on 1.2% agarose electrophoretic gel, along with uncut lambda DNA in concentrations of 10, 20, 40, 60, 80, and100 ng/µl. The gel was stained with ethidium bromide and visualized with a UV trans-illuminator. DNA was quantified by comparing the intensity of the band with that of the lambda standards.

3.3.3 DNA bulks and AFLP

Equal amounts of DNA (143 ng from each plant) from seven homozygous *diminutive* plants of cv. SunUp were pooled as a single bulk. Pooling equal amounts of DNA (125ng from each plant) from eight SunUp plants that were homozygous for normal phenotype made another bulk. Using 125 ng of DNA from each bulk, AFLP was performed as per the protocol in Chapter 2.3 of this dissertation. Pre-amplification was performed with E-G + M-C primer pair. Selective amplification was performed using

different AFLP primer sets amd the primer set E-GA and M-CAG revealed a polymorphism between the two bulks studied. A 175 bp band was present in the normal bulk and was absent in the *diminutive* bulk. Another pair of DNA bulks was made from 10 homozygous *diminutive* plants (100 ng DNA from each plant) or 10 normal (homozygous or heterozygous for the normal allele) plants, from the segregating recurrent selection population involving SunUp as one of the parents. This bulk was used to confirm the co-segregation of marker and phenotype detected in the SunUp bulks. Later, individual diminutive plants were tested for marker co-segregation.

3.3.4 Cloning the marker

The band showing polymorphism was cut from the acrylamide gel using Odyssey, infrared detection system from Li-Cor Inc (Lincoln, NE). The cut band was frozen at –80°C for 15 minutes and spun at 12,000 rpm for 10 minutes. The supernatant was used as the template and the band was re-amplified using non-labeled E-GA and M-CAG as primers. The PCR product was purified by precipitating with two parts (V/V) of 95% ice-cold ethanol and 1/10 parts (V/V) of 3.5 M sodium acetate (pH 5.2) and washed with 70% ethanol. Purified PCR product was cloned into *Topo® TA* cloning vector from Invitrogen Inc. (Carlsbad, CA) and transformed into *One shot® E.coli* competent cells from Invitrogen Inc. (Carlsbad, CA) by electroporation, as per the manufacturer's instruction. The competent cells were plated on LB agar containing 40 µg/µl of kanamycin and 20 µg/µl of X-gal. The white colonies were selected and cultured on LB agar with ampicillin (100 µg/ml) at 37°C in a shaking incubator.

3.3.5 Plasmid DNA extraction

The culture tubes containing the transformed *E.coli* were spun at 3000 rpm at 4°C for 15 minutes. The supernatant was poured off, and the pellet was resuspended in 200 µl of cold solution-I, made up of 50 mM glucose, 25 mM tris-HCl (pH 8.0) and 10 mM EDTA. The mixture was transferred to a 1.5-ml microfuge tube and incubated for 5 minutes on ice. Freshly prepared solution-II (400 µl of 1% (W/V) SDS, 0.2 N sodium hydroxide) was added, and the tube was inverted gently 8-10 times until the solution turned translucent. The solution was then incubated on ice for 5 minutes and ice cold solution–III (300 µl of 3 M potassium acetate (pH 4.5) and 28.5 % of glacial acetic acid) was added, and the tube was inverted gently 8-10 times until a white precipitate was formed. The tube was frozen at –80°C for 20 minutes and spun for 12 minutes at 12,000 rpm at 4°C. The supernatant was transferred to a 1.5-ml micro centrifuge tube and RNAse (3.5 µl) added, and incubated at 37°C for 30 minutes, then, incubated on ice for 5 minutes. Ice-cold ethanol (600 µl) was added, mixed gently, and frozen at –80°C for 20 minutes. The DNA was pelleted by spinning the tube at 12,000 rpm at 4°C for 15 min and the supernatant was removed and the DNA pellet was washed with 1 ml of 70% ice-cold ethanol, spun for 10 minutes at 12,000 rpm and the ethanol was decanted. The DNA pellet was dried in the hood and resuspended in 150 µl of TE.

3.3.6 Sequencing

To check the insertion of the PCR product, the plasmid DNA was digested with EcoRI. The reaction mixture consisted of 10 µl volume containing 100 ng of plasmid

DNA, 1X restriction enzyme buffer, 2 mg acetylated BSA (bovine serum albumin), and 6 units of EcoRI enzyme. The reaction mixture was incubated at 37°C for 2 hours and the fragments were separated on 1.2% agarose gel containing ethidium bromide. The DNA was visualized with a UV trans-illuminator, and the presence of the insert (175 bp) was confirmed. The concentration of plasmid DNA was measured by subjecting the sample to electrophoresis in agarose gel, along with the standards.

Sequencing was done following the procedures of the Thermo Sequenace Cycle Sequencing kit from USB Corporation (Cleveland, Ohio), except that the primers labeled with infrared-dye were from Li-Cor (Lincoln, NE). The sample was run on a Li-Cor IR^2 automated sequencer. The sequence was compared at the National Center for Biotechnology Information (NCBI) sequence database using Basic Local Alignment Search Tool (BLAST). Two sets of primers were designed, one set to check a point mutation in the restriction site and another set to check for a point mutation in the extension base pairs or for a deletion mutation of the whole region. The primers were designed by taking the end sequences of the cloned marker fragment. One set of primers included the restriction site and another set excluded the restriction site (Table 3.1 and Table 3.2).

Table 3.1 Primer pair used to test the mutation at the restriction site

Sequence	% of GC	Melting temperature
Forward 5' GAA TTC GAG AGT TTT TGA CGG 3'	42.9	55.4°C
Reverse 5' TTA ACA GGA TAC CAA AGA GTT AC 3'	34.8	51.5°C

Table 3.2 Primer pair used to test the mutation at the selective extension nucleotide binding site

Sequence	% of GC	Melting temperature
Forward 5' GAG AGT TTT TGA CGG GAA AG 3'	45.0	52.3°C
Reverse 5' CAG GAT ACC AAA GAG TTA CAG TAT 3'	37.5	54.2°C

3.4 Result

Primer set E-GA + M-CAG generated a presence/absence polymorphism band of 178 bp. When the band was used to amplify normal and *diminutive* DNA bulks from SunUp, as well as DNA bulks from the recurrent selection population. This polymorphism was confirmed by testing individual plants in the bulks. All normal plants showed the presence of the 178 bp band, and all the *diminutive* plants lacked the band. Hence an AFLP marker that consistently co-segregates with the gene that control the *diminutive* trait was found (Fig. 3.1).

The marker was cloned, and the sequence found was:

1 GACTGCGTAC CAATTCGAGA GTTTTTGACG GGAAAGTTAG TAGTCCTAAT

51 AGTATTAGAG ATAGATGGAT GGATTCTTTT GGGACCATCT TATTAGATCA

101 TAACATTATC AGTGATCTGC ATTTTGGGAA TGTCATTGAT ACTGTAACTC

151 TTTGGTATCC TGTTACTCAG GACTCATC

The sequence did not show homology with any other available sequences in the NCBI database. Both of the simple PCR primer pairs designed specifically for the marker sequence amplified normal as well as diminutive DNA.

3.5 Discussion

The 178-bp AFLP marker that completely co-segregates with the diminutive mutant locus was identified, cloned and sequenced. Since the dominant form of the marker was linked to normal phenotype, AFLP markers cannot be used to identify normal phenotypes carrying *diminutive* allele in heterozygous form. The polymorphism produced was between the SunUp-normal and the SunUp-*diminutive* plants. The marker

RECEIVED AS FOLLOWS

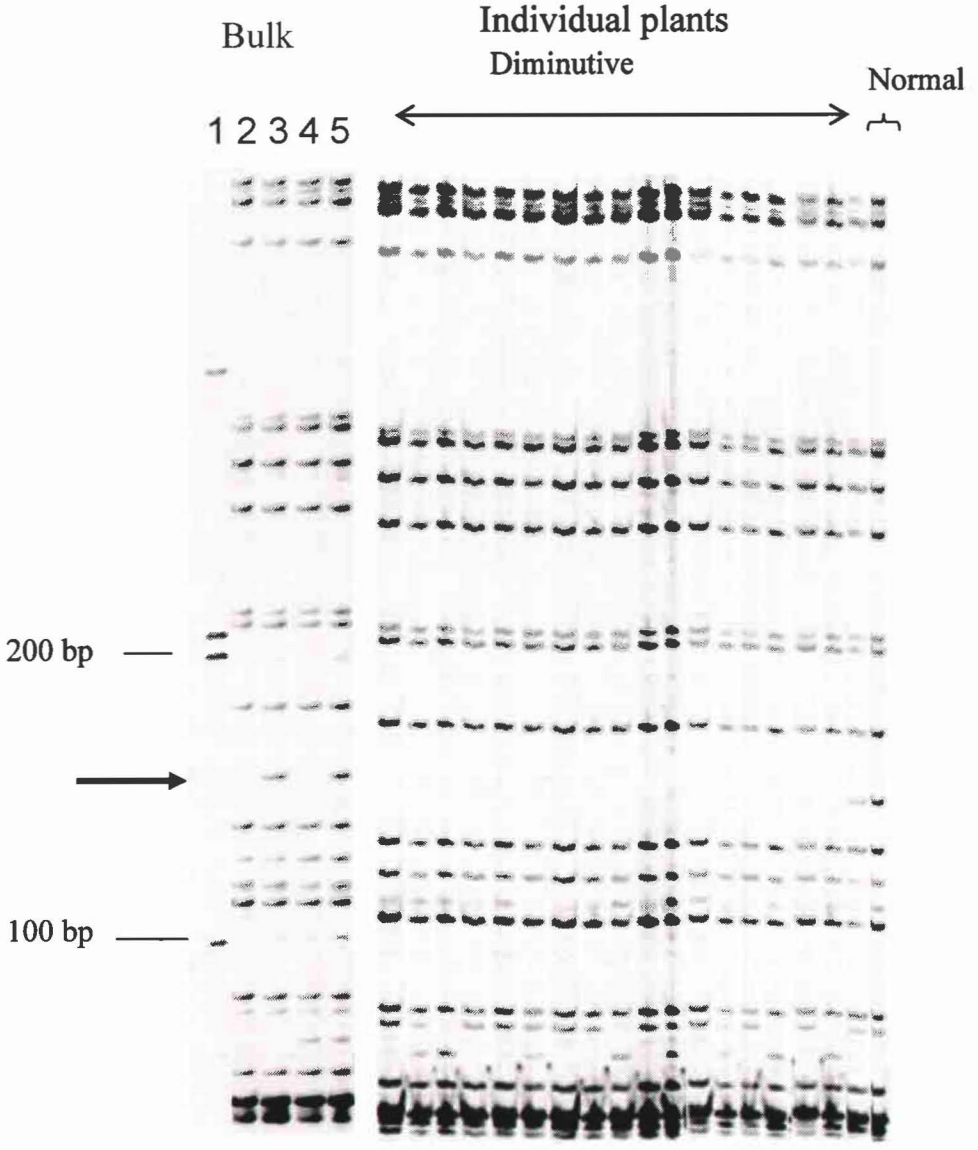


Fig. 3.1 AFLP gel image showing the marker linked to *diminutive* mutation in papaya. 1-Molecular weight, 2-*Diminutive* bulk from SunUp, 3-Normal bulk from SunUp, 4-*Diminutive* bulk from the recurrent selection population, and 5-Normal bulk from the recurrent selection population. The arrow indicates the marker.

was expected to be in the coding region that causes the *diminutive* phenotype, but the cloned sequence did not align with any other available plant sequences. Both sets of AFLP marker-sequence-specific primers produced monomorphic bands in *diminutive* and normal plants. Hence, the mutation that gives rise to the polymorphism between *diminutive* and normal plants was probably a point mutation in one of the following three possible regions: a) restriction site, b) binding site of 3' end of pre-amplification primer, or, c) binding site of 3' end of selective-amplification primer.

The flanking regions sequences of the AFLP marker, from a papaya BAC library would help to design PCR primers outside the marker segment. Cloning and comparing the sequences of the marker between *diminutive* and normal phenotype plants would help to identify the point mutation that gives rise to the marker. Designing a single nucleotide polymorphism (SNP) marker at the mutation site would allow us to distinguish phenotypes carrying homozygous and heterozygous normal alleles.

CHAPTER 4

ISOLATION AND PARTIAL CHARACTERIZATION OF LYCOPENE β-CYCLASE FROM PAPAYA

4.1 Introduction

When breeding fruit crops like papaya, breeders need to wait atleast one year to study fruit traits. This problem can be reduced by marker assisted selection, where DNA markers are used to screen progeny at the seedling stage. Many of the commonly employed marker procedures, RAPD, and AFLP are only useful in the particular population in which they were created (Staub *et al.*, 1996). These markers would be more robust and applicable to all populations, if the molecular basis of the trait difference was understood, and the marker incorporates the genetic basis of the alleleic polmorphism.

The molecular basis of color difference in tomato, bell pepper, and onion has been investigated (Ronen *et al.*, 2000; Lefebvre *et al.,* 1998; Kim *et al.*, 2004). In tomato, a mutation in chromoplast-specific lycopene β-cyclase is responsible for up-regulation of β-carotene production, making tomatoes orange instead of red (Ronen *et al.*, 2000). To investigate the possibility of polymorphism in lycopene β-cyclase as the molecular basis for flesh color difference in papaya, we attempted to isolate a lycopene β-cyclase gene fragment.

4.2 Literature review

Papaya fruit flesh is either red or yellow. All wild papayas are yellow in color. From the segregation patern, Hofmeyr (1938) reported that flesh color is controlled by a single gene with yellow flesh color being dominant over red flesh color. Storey (1969) stated that although red and yellow are basic fruit-flesh colors, each color has a range of shades due to modifying genes acting upon their expression.

4.2.1 Carotenoids in papaya

Karrer and Jucker (1950) found that the major pigment in papaya yellow flesh is cryptoxanthin. Later, in 1964, Yamamoto documented the carotenoids in both yellow- and red-fleshed papaya fruits. Yellow-fleshed papayas had a higher amount of cryptoxanthin, cryptoxanthin epoxide and ζ-carotene compared to red-fleshed fruit. Red-fleshed papaya had a high amount of lycopene, which was absent in yellow-fleshed fruit. Recently, Chandrika (2003) confirmed that red-fleshed fruit have lycopene, and that yellow-fleshed fruit lack it.

4.2.2 Carotenoid biosynthetic pathway

Carotenoid pigments are synthesized in the isoprenoid biosynthetic pathway (Fig. 4.1) and are controlled by enzymes that are encoded in the nuclear genome (Cunningham and Gnatt, 1998; Ronen *et al.*, 2000). Carotenoid biosynthesis takes place in the plastids (Hirschberg, 2001). In the carotenoid pathway, red-colored lycopene is converted into yellow colored cryptoxanthin in a two step process. First, lycopene is converted into

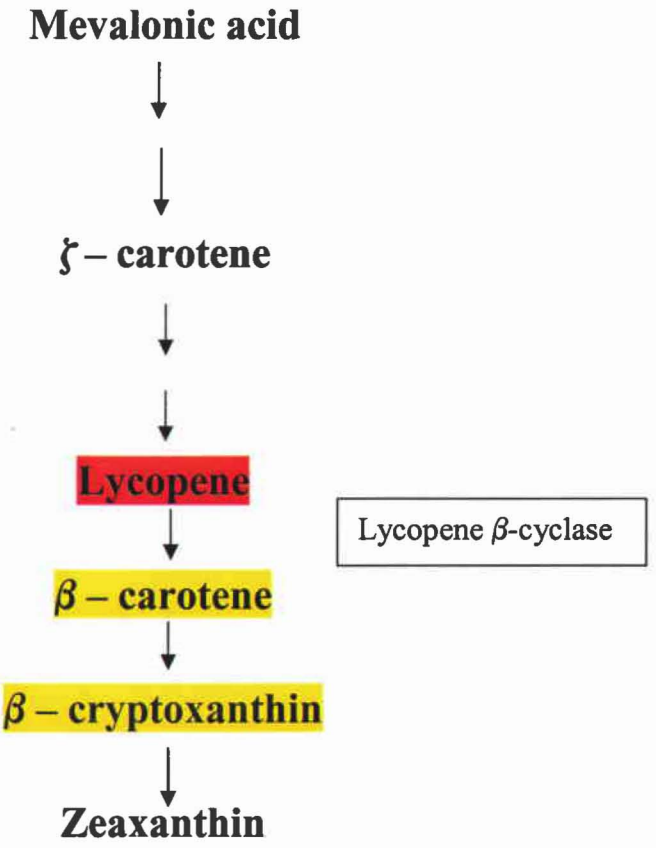


Fig. 4.1 The simplified carotenoid biosynthetic pathway in plants.

yellow-colored β-carotene by lycopene β-cyclase, and then, β-carotene is converted into β-cryptoxanthin by β-ring hydroxylase.

4.2.3 Cyclases in carotenoid biosynthetic pathway

A set of cyclases, lycopene β-cyclase (LCY-B) and lycopene ε-cyclase (LCY-E), convert lycopene into β-carotene and δ-carotene respectively. LCY-B adds two β rings, one to each end of linear, red-colored lycopene, in two steps leading to the production of cyclic, yellow-colored β-carotene. Alternatively, lycopene ε-cyclase (LCY-E) adds one ε ring to the lycopene creating δ-carotene. Lycopene β-cyclase converts δ-carotene to α-carotene by adding one β ring. Alpha-carotene is the precursor for lutein, which plays a major role in photochemical quenching in the chlorophyll. Both lycopene ε-cyclase and lycopene β-cyclase are encoded by single genes in *Arabidopsis* (Cunningham *et al.*, 1996).

While all plants have lycopene ε- and β-cyclases, some plants have another cyclase. In pepper, the enzyme capsanthin-capsorubin synthase (CCS) converts violoxanthin and antheraxanthin to capsanthin and capsorubin, which give red color to the fruits. In addition, CCS has some LCY-B activity also (Huguency *et al.*, 1995; Cunningham *et al.*, 1996). In tomato, a second lycopene β-cyclase specific to the chromoplast has been identified (Ronen *et al.*, 2000) and named CYC-B (Hirschberg,

2001). CYC-B is homologous to pepper CCS (86.1 %), and produces β-carotene in tomato fruit. Another enzyme in the carotenoid pathway, neoxanthin synthase (NXS), shows 98% sequence homology to tomato CYC-B. Function, gene sequences and map position suggests that CCS and CYC-B are orthologs arising from LCY-B by gene duplication (Hirschberg, 2001).

4.2.4 Molecular basis of color differences in other crops

Mutations in genes coding for enzymes in biosynthetic pathways have been identified as the cause of color differences in fruit. In tomato, the major pigment responsible for red color of the fruit is lycopene. Beta (B) is a partially dominant, single locus mutant that produces orange-colored fruit due to the accumulation of β-carotene. Old-gold (og) and old-gold crimson (og^c) are mutants that do not produce any β-carotene and accumulate increased amount of lycopene compared to the wild type. Ronen *et al.* (2000) studied the molecular basis for red and orange color differences in tomato fruit. A chromoplast-specific lycopene β-cyclase (CYC-B) was cloned and sequenced by map-based cloning. The tomato CYC-B sequence showed 86.1% homology to pepper CCS and 53% homology to tomato LCY-B. By comparing the alleles of CYC-B in normal and B mutant, it was found that the B mutant allele has six sequence elements in the 2520 bp upstream of the coding region that are not present in the wild type allele. Wild type allele has a single sequence element that is not present in the B mutant allele. This sequence difference in CYC-B gene between wild type and B mutant alleles has led to higher amount of β-carotene accumulation in the B mutant fruit (45-50% of total fruit

carotenoids) compared to that of wild types (5-10% of total fruit carotenoids). When CYC-B gene sequence of wild type, and old-gold (og) and old-gold crimson (og^c) mutants were compared, a frameshift mutation in the coding region was detected in the mutants. This mutation has caused the production of non-functional CYC-B enzyme in the old-golds. Compared to the wild type, old-gold (og) and old-gold crimson (og^c) accumulate increased amount of lycopene at the cost of β-carotene. In bell pepper, a deletion mutation in the gene for the enzyme capsanthin-capsorubin synthase leads to the production of yellow peppers instead of red peppers (Lefebvre *et al.,* 1998). In onion, a single nucleotide polymorphism (SNP) in the promotor region of dihydroflavonol 4-reductase (DFR) gene stops anthocyanin production in the bulbs, making them appear yellow instead of red. A marker specific for the SNP co-segregates completely with the color difference in the onion bulb (Kim *et al.*, 2004).

4.3 Materials and methods

4.3.1 Designing degenerate primers

Basic Local Alignment Search Tool (BLAST) at National Center for Biotechnology Information (NCBI), identified cDNA sequences of lycopene β-cyclase (LCY-B) from navel orange, grape fruit, *Capsicum*, tomato, tobacco, *Adonis*, marigold and *Arabidopsis*. All LCY-B sequences were aligned (Seqweb, Accelrys Inc, San Diego, CA) and a consensus sequence was obtained. A 29-bp forward-degenerate primer and 28-bp reverse degenerate primer were designed (Table 4.1). The primers were checked for dimer and hairpin formation (Oligoanalyzer 3.0, Integrated DNA Technologies, Inc.,

Coralville, Iowa). Using genomic DNA from papaya var. SunUp, consisting of 35 PCR cycles of denaturation at 90°C, annealing at 50°C and elongation at 72°C was carried out. The reaction was carried out in 50-µl volume consisting of 100 ng of the template DNA, 0.2 mM of dNTPs, 1X PCR buffer, 1.5 mM $MgCl_2$ and 2 IU of Taq polymerase. Since the LCY-B gene from *Arabidopsis* has no intron (Cunningham *et al.*, 1996), genomic DNA could be used for amplification. The genomic DNA was extracted by CTAB method and cleaned with CsCl (Doyle and Doyle, 1987), which is a modification from Saghai-Maroof *et al.* (1984). The detailed extraction protocol has been reported in Chapter 2.3 of this dissertation. PCR samples were run on 1.2% agarose gel containing ethidium bromide (8 µg /100 ml of agarose gel), and visualized with a UV trans-illuminator. A DNA fragment of 332 bp, which was the expected size, was obtained.

Degenerate primers designed for a consensus sequence obtained from cDNA sequences of CYC-B gene from tomato, CCS gene from *Capsicum*, and navel orange, and NXS from tomato and potato, were used to clone CYC-B gene (Table 4.1). This CYC-B gene from tomato also does not have intron (Ronen *et al.*, 2000), genomic DNA from papaya var. SunUp was used. Amplification was done with tomato DNA as a control. Annealing temperatures ranging from 50-54°C were tried. Primer concentration was varied from 50-100 µM per 50 µl reaction mixture, and DNA concentration was varied from 100-200 ng per reaction.

Table 4.1 Degenrate primers designed for LCY-B and CYC-B gene in papaya

LCY-B and CYC-B Forward	5' AAY TAT GGW GTT TGG GTD GAT GAR TTT GA 3'
LCY-B Reverse	5' TAD GGY TTR TCA TAY TGV ACH A 3'
CYC-B Reverse	5' GGC TCA TTM CCY AAA TGR G 3'

where, Y=C,T; W=A,T; D=A,G,T; R=A,G; V=A,C,G; H=A,C, and M=A,C.

4.3.2 Excision of DNA fragment from agarose gel and cloning

A spin column was prepared by placing a 500-µl centrifuge tube inside a 1500-µl centrifuge tube. The 500-µl centrifuge tube had a hole in the bottom, over which glass wool was placed. The PCR fragment was cut from the agarose gel, placed on the glass wool, and spun at maximum speed for five minutes. The eluate was frozen at -20°C. The PCR fragment was cloned into pGEM-T Easy vector (size 3 kb) from Promega (Madison, WI). A reaction mixture consisting of 3 µl of PCR product, 1µl (50 ng/µl) of vector, 5 µl of 2X ligation buffer and 1µl of T4 DNA ligase (100 units/µl) was incubated at 15°C overnight. The vector has an insertion site in the middle of *lacZ* gene sequence coding for β-galactosidase protein, without interrupting the production of the functional enzyme.

4.3.3 Transformation in *E.coli*

Competent cells of *E. coli* strain DH5α were transformed with vectors by heat-shock transformation. Five µl of ligation mixture was added to the competent cells and the mixture was kept on ice for 60 minutes, then incubated at 42°C for 45sec. Four hundred microlitre of Super Optimal Catabolite (SOC) media was added to the competent cells, incubated at 37°C for 45 minutes, plated (200µl/plate) into Mc Conkey agar (pH 7.0) containing ampicillin (100µg/ml) and phenol-red indicator and were incubated overnight at 37°C. White colonies were selected against red colonies. When the vector had no insert, β-galctosidase was produced that ferments lactose in the media and produces lactic acid. The acid reduced the pH of the media around the colonies and the

colonies turn red. If the vector had an insert, β-galactosidase was not produced, and the *E. coli* colonies appeared white.

4.3.4 Plasmid DNA extraction

The transformed *E. coli* colonies were picked and placed individually into separate tubes with 5 ml of Luria-Bertani (LB) agar with (100μg/ml) ampicillin. The tubes were incubated at 37°C overnight at 300 rpm on a shaker. The culture (1.5 ml) was taken in a microcentrifuge tube and spun at 12,000 g for 30 seconds. The medium was removed and the bacterial pellet was allowed to dry. Ice cold solution I (100 μl of 50 mM glucose, 25 mM tris-HCL, pH 8.0 and 10 mM EDTA) was added and vortexed vigorously. Freshly prepared solution II (200 μl of 1% SDS, 0.2N sodium hydroxide) was added. The tube was inverted once and placed on ice,150 μl of ice-cold solution III (3M potassium acetate, pH 4.5, 28.5% glacial acetic acid) was added and vortexed gently upside down for 10 seconds. The tube was stored on ice for 5 minutes. The tube was centrifuged at 12,000 g for 5 min. The supernatant was placed in fresh tube. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added, vortexed and left at the room temperature for two minutes and then spun at 12,000 g for five minutes. The supernatant was removed, and 600 μl of ice-cold ethanol was added. And then spun at 12,000g for 5 minutes, and the DNA was precipitated as a pellet. The pellet was rinsed with 1ml of 70% ethanol and the supernatant removed, and the pellet was dried. The pellet was redissolved in 50 μl of TE (pH 8.0) that contained DNAse-free RNAse (20 μg/ml).

Plasmid DNA was cut with *EcoRI* to check for the presence of the insert. The reaction mixture consisted of 10 µl volume that contained 100 ng of plasmid DNA, 1X restriction enzyme buffer, 2 mg acetylated BSA (bovine serum albumin), and 6 units of *EcoRI* enzyme. The reaction mixture was incubated at 37°C for 6 hours and run on 1.2% agarose gel containing ethidium bromide. When the DNA was visualized with UV trans-illuminator, two bands were observed, one ~3000 bp (vector) and the other ~350 bp (insert). The concentration of plasmid DNA was measured by running the sample in agarose gel along with the standards.

4.3.5 Sequencing

Sequencing of the sample was done at Greenwood Molecular Biology Facility at the University of Hawaii, at Manoa. The sequence obtained was compared with other sequences available at NCBI, using BLAST, and it confirmed that the cloned fragment was LCY-B.

4.4 Results

4.4.1 LCY-B gene

The 332 bp sequence cloned from papaya was identified as lycopene β-cyclase. The sequence was,

5' AATTATGGAG TTTGGGTGGA TGAGTTTGAG GCCATGGACT TGCTAGACTG

CCTTGACACC ACCTGGTCCG GTGCTGTTGT GTACATAGAC GATAAAAGCA

AGAAAGATTT GGATAGACCT TATAGAAGGG TTAACAGGAA GCAGCTGAAG

TCAAAGATGT TGCAGAAATG TATAGCAAAT GGGGTGAAGT TTCACCAAGC

TAAAGTTATT AAAGCCATTC ACGAGGAATC AAAGTCTTTG TTGATCTGCA

ATGATGGTGT TACGATTCAG GCCACTGTAG TTCTCGATGC AACTGGCTTT

TCTAGATGCC TGGTTCAATA TGACAAGCCC TA 3′

with the primer sequence underlined. When the primer sequences were excluded, the fragment length was 281bp. The sequence was submitted to Genbank (Accession number is AY753202). The results of BLASTN search at NCBI with databases non-redundant coding regions without EST (nr) and genomic survey sequences (gss) were summarized (Table 4.2). During the NCBI sequence homology search, a papaya genomic clone (GI=33899150) of 818 bp from cv. SunUp submitted by TIGR showed homology with papaya LCY-B. The fragment that we cloned and the genomic clone submitted by TIGR overlapped by 244 bp. We speculate that papaya has a single copy of LCY-B and the genomic clone represents part of coding region of papaya LCY-B. Although both the sequences were from cv. SunUp, a single bp nucleotide difference was observed between the two sequences. In the TIGR clone, the 58th nucleotide was guanine, while in our sequence,the nucleotide was adenine. The LCY-B gene fragment that we cloned along with the genomic clone submitted by TIGR constituted 855 bp of the coding region of papaya LCY-B. The papaya LCY-B showed higher sequence homology to citrus LCY-B than to that of any other crop.

Table 4.2 Results of BLASTN search for papaya LCY-B at NCBI.

Source	GenBank ID	Data-base	Identity	Score (bits)	Probability
Carica papaya	GI=33899150	gss	243/244 (99%)	476	e-132
Citrus sinensis	GI=44887637	nr	201/243 (82 %)	149	4e-33
Citrus limon	GI=40809762	nr	200/243 (82 %)	141	9e-31
Citrus x paradisi	GI=13959730	nr	200/243 (82 %)	141	9e-31
Citrus unshiu	GI=40809730	nr	200/243 (82 %)	141	9e-31
Citrus maxima	GI=40756517	nr	200/243 (82 %)	141	9e-31
Lycopersicon esculentum	GI=1006689	nr	163/195(83%)	133	2e-28
Tagetus erecta	GI=21360356	nr	132/156 (84%)	119	3e-24
Capsicum annuum	GI=999440	nr	160/195 (82%)	109	3e-21
Nicotiana tabacum	GI=1006689	nr	166/206 (80%)	91.7	8e-16
Bixa orellana	GI=45237490	nr	95/113 (84%)	41	8e-13
Daucus carota	GI=6601473	nr	155/194 (79%)	75.8	5e-11
Arabidopsis thaliana	GI=20259238	nr	41/47 (87%)	46.1	0.042

4.4.2 CYC-B gene

Papaya DNA did not show any amplification for the degenerate primers designed for CYC-B gene. These primers amplified only tomato DNA at all the combinations of DNA concentrations, primer concentrations and annealing temperature tested.

4.5 Discussion

A 332 bp fragment of papaya LCY-B gene was cloned and sequenced. The cloned papaya sequence together with a papaya genomic clone submitted by TIGR (GI=33899150) constituted 855 bp of LCY-B gene in papaya. Though papaya is phylogenetically closely linked to *Arabidopsis,* the papaya LCY-B showed high sequence homology to the citrus LCY-B. This homology may be because the papaya LCY-B fragment used for the analysis was a partial coding region (855 bp). When the complete gene is sequenced, the papaya LCY-B may show closer sequence homology to *Arabidopsis* LCY-B.

Primers for CYC-B did not amplify any product from papaya. Southern blotting of restriction digests of papaya genomic DNA using the LCY-B fragment as a probe would have told us how many copies of LCY-B gene were present in papaya, and the possibility of a papaya chromoplast specific LCY-B gene.

The molecular basis of color difference in tomato, bell pepper and onion have been revealed as mutations in the genes coding enzymes in carotenoid and flavonoid pathways. To find whether a similar mutation in lycopene β-cyclase was responsible for fruit-flesh color difference in papaya, future work would involve mRNA extraction from red- and yellow-fleshed papaya fruits at different ripening stages. Northern blotting using the 855 bp papaya LCY-B fragment as the probe would determine the expression at different stages of fruit ripening. If differential expression of LCY-B was observed between yellow- and red-fleshed fruits, we could conclude that the expression of LCY-B controls the flesh color difference in papaya. Reduced expression of LCY-B in red-fleshed papaya could be caused by a mutation in the *cis* elements, the promotor, 5' untranslated region (UTR) or 3' untranslated region (UTR) or in *trans* binding proteins. To check for the presence of mutation in the *cis* elements, the papaya BAC library at Hawaii Agriculture Research Center (Ming *et al.*, 2001) would be probed with the 855 bp LCY-B fragment to locate the genomic sequence, and complete LCY-B gene sequence would be obtained. Primers could be designed specifically for the LCY-B gene, and the gene sequence from yellow–fleshed fruit. Sequences in the coding region, promotor region, 5'UTR and 3'UTR would be compared between red- and yellow-fleshed fruit. If there is any difference in the sequences between red and yellow-fleshed fruit, primers will be designed to detect the mutation, and the co-segregation pattern of red and yellow flesh color with the mutation would be studied. If consistent co-segregation was observed, the site of mutation would be used as a marker to screen papaya seedlings for flesh color.

CHAPTER 5.0

SUMMARY

A linkage map was created for papaya using 281 F_2 plants derived from the parents Kapoho and Saipan Red. From the linkage map, an AFLP marker associated with a simple trait, fruit-flesh color was identified. QTL analysis was performed for three quantitative traits - node number at floral conversion, carpellody, and fruit weight.

The phenotypically contrasting parents Kapoho and Saipan Red showed high polymorphism (3.3/primer pair) for AFLP markers. Using 158 primer pairs, 38 co-dominant markers and 475 dominant markers were identified. A mixed-phase linkage map was created (Mapmaker/Exp V 3.0) at LOD 5.0 and recombination frequency 0.25. The map consisted of 15 linkage groups with 413 markers, covering 2066 cM of the papaya genome with average marker interval of 5.0 cM.

An association was observed between a) PGM locus and node number at floral conversion, and, b) flesh color and carpellody at field level, but these associations could not be confirmed at the molecular level.

One of the qualitative traits, PGM, was located on linkage group LG-5 with flanking markers 11.0 and 8.3 cM away. The other simple trait, fruit-flesh color, was located on one end of LG-14 with a linked marker at 4.1 cM distance.

An interesting feature noticed in the mixed-phase map was the appearance of blocks of consecutive dominant markers, which arose from the same parent. The probability of observing such blocks by chance was statistically extremely low for three large Kapoho blocks and one large Saipan Red block. These blocks may have some unknown biological origin, or, might be artifacts of ordering dominant markers in F_2 population.

Another feature observed in the map was the pattern of distorted markers. Of the distorted markers, 49.4% were located on LG-4 and LG-5. Nearly the entire linkage group of LG-5 showed distorted segregation. All the distorted markers in LG-5 had selective advantage for Kapoho alleles. The distortion rate was the highest at 60-96 cM from the top of the linkage group and decreased farther away from this locus. At both ends of the linkage group, the markers started to show Mendelian segregation. Segregation distortion might have originated at the seedling stage by positive selection for Kapoho alleles.

After removing distorted and faint markers, two coupling-phase maps, one for each parent were created. A minimum threshold of LOD 2.0 was used for ordering of markers. The Kapoho map covered 778.8 cM with 121 markers, and the Saipan Red map covered 693.0 cM with 101 markers. QTL analysis was done on coupling-phase maps.

A single major QTL was found for node number at floral conversion on LG-4 by interval mapping (Mapmaker/QTL). In coupling-phase maps, this QTL appeared as two

QTL peaks, one on KLG-4 and another on SLG-4b. Both QTL peaks showed additive genetic action and had a negative effect on the trait. The QTL peaks on KLG-4 and SLG-4b explained 64.3% and 23.8% of the phenotypic variation respectively. The markers under the QTL peaks showed significant association with the trait in single marker analysis by linear regression and increased the reliability of the QTL.

The nonparametric Kruskal-Wallis rank sum test, single marker analysis, was done on carpellody (MAPQTL V 5.0). More than 50 % of the markers that showed significant association with carpellody aligned in blocks on four linkage groups, LG-2, LG-3, LG-4 and LG-7. On LG-7, a block of markers showed significant association with carpellody in both seasons – 2002 February and 2002 August.

Five markers were identified to be associated with fruit weight in single marker analysis by linear regression. No QTL peak was produced in interval mapping, in spite of relatively high heritability (H^2=74.0%). This implies that fruit weight was controlled by many QTLs with small individual effects.

An AFLP marker that completely co-segregates in coupling-phase with normal phenotype at the *diminutive* mutant locus in papaya was found by bulked segregant analysis. The marker was cloned and sequenced. SCAR primers for this marker amplified both diminutive and normal phenotypic DNA.

A 281 bp fragment of lycopene β cyclase (LCY-B) gene was isolated from papaya, sequenced and submitted to the GenBank (Accession No.=AY753202). The sequence homology was closer to that of citrus LCY-B than to *Arabidopsis*. The papaya LCY-B fragment can be used as a probe in investigating the possibility of polymorphism in LCY-B as the molecular basis for fruit-flesh color difference in papaya.

Appendix A. Primers used in the present investigation and the number of markers obtained.

Primer	M-CAA	M-CAC	M-CAG	M-CAT	M-CTA	M-CTT	M-CTC	M-CTG	M-CC	M-CG	M-CA	M-CT	M-GC	M-GG	M-GA	M-GT	M-AC	M-AG	M-AA	M-AT	Total
E-AA	5	5	3	3	4	6	8	7													41
E-AG	5	3	2	8	3	3	8	2													34
E-AAG									4	3			10	1	0	9				2	29
E-AAC									9	4			1	1	1		2	1	1	2	22
E-AGG									3	6			1		3	4					17
E-AGC									9	8			3		3	2					25
E-AAT									5	3			3	2	3	4					20
E-AGT									5	4			3	9	6	4	1	2		5	39
E-AGA									3	3			4	1	2	2		1			16
E-ATA									4	3	2	11					4	0		2	26
E-ATG									3	1	5	9	3	1	6	2	6	1		2	39
E-ACA									5	1	0		7	4	9	4	6	1	3		40
E-ATC									3	2	1		5	4	3	0					18
E-ACC									6	2	5	3	4	3	7	4					34

Appendix A (continued). Primers used in the present investigation and the number of markers obtained.

Primer	M-CAA	M-CAC	M-CAG	M-CAT	M-CTA	M-CTT	M-CTC	M-CTG	M-CC	M-CG	M-CA	M-CT	M-GC	M-GG	M-GA	M-GT	M-AC	M-AG	M-AA	M-AT	Total
E-ATT									3	2	1	0	3	0	2	0					11
E-ACG									0	0	4	5	1	2	2	4	2		1	7	28
E-ACT									3	3	8	2	2	1	4	4					27
E-GG		0			0			2													2
E-GT		1						2													3
E-GC		1	0	1		1	1			0											4
E-GA		3	3	5	4	6	2			2											25
E-GTA										1											1
E-GTC										3											3
E-GTG									2												2
E-GTT									1												1
E-GGA												4									4
E-GGC												2									2
Total	10	13	8	17	11	16	19	13	68	51	26	36	50	29	51	43	21	6	5	20	513

Appendix B. Markers that appear as common between the present map and Ma *et al.*'s map (2004).

Kapoho X Saipan Red F_2			Map of Ma *et al.*(2004)		
Marker name	**Size (bp)**	**LG**	**Marker name**	**Size (bp)**	**LG**
L0409ES3	695	3	L0409BS5	700	2
L0410ES5	480	3	L0410AS5	485	2
M0329AS5	102	3	M0329AS5	105	2
M0330CK5	185	3	M0330AK5	190	2
M0411AS5	195	3	M0411AS5	195	2
M0429AK5	120	3	M0429AK3	122	2
M0512AK3	125	3	M0512AK5	123	2
M0610BK3	160	3	M0610AK3	170	2
M0832BS5	190	3	M0832AS5	198	2
M0910BS3	250	3	M0910DS3	252	2
M1330AS3	105	3	M1330AS5	105	2
M1412HK5	255	3	M1412DK5	255	2
M1426CS5	210	3	M1426GS5	214	2
M1426DS5	295	3	M1426HS5	295	2
M1526BS5	210	3	M1526ES5	217	2
M1529BS5	200	3	M1529CS2	198	2
M1608BCC	120	3	M1608BCC	110	2
M3112AK5	305	3	M3112BK5	305	2
S0310AK5	80	3	M0310GK5	80	2
L2108CK5	520	3	L1308DK5	520	2
M2102BS5	155	3	M1302AS5	150	2
M0110CS3	135	8	M0110CS5	135	4
M0110DK5	140	8	M0110DK5	140	4
M0528CS5	200	8	M0528AS5	198	4
M0609ECC	275	2	M0609BCC	276	4
M1028CS5	230	2	M1028BS3	226	2
M1526FK5	317	2	M1526FK3	312	Unlinked
M2102AS3	115	2	M0802AS5	105	3
M0932CK5	210	3	M0932DK5	204	4
M2001CK3	200	3	M0901CK5	195	6
L3710AS5	380	3	L3710CS5	388	5
M2006HS5	345	3	M0206BS2	340	1
M2105AK5	110	3	M0705AK0	118	Unlinked
M1329ACC	~125	3	M1329AS3	124	Unlinked
			M1329BK3	126	2

Appendix B (Continued) Markers that appear as common between the present map and Ma *et al.*'s map (2004).

Kapoho X Saipan Red F_2			Map of Ma *et al.* (2004)		
Marker name	**Size (bp)**	**LG**	**Marker name**	**Size (bp)**	**LG**
M0911AK5	110	4	M0911AK5	110	2
M0310BK3	285	4	M0310BK3	275	6
M0326CK5	210	4	M0326CK3	220	3
M2003CS2	130	4	M0903AS3	125	4
M2106DS2	265	4	M1306BS2	272	unlinked
M0431AS2	190	5	M0431BS3	200	unlinked
M1410CS2	290	5	M1410DS5	287	1
M2104DCC	225	5	M1004DCC	224	10
M0411BS5	335	6	M0411BS0	340	7
M1608AK5	105	6	M1608AK3	100	3
M1412BK5	130	6	M1412BK5	125	11
M0429CS5	155	7	M0429BS3	155	3
M0529BK5	310	7	M0529EK3	305	3
L0528EK5	375	9	L0528BK3	381	3
M1526AK5	110	9	M1526AK3	112	unlinked
M0929BK5	240	10	M0929CS3	250	9
M1410BS5	170	unlinked	M1410BS5	160	1
M1409AS3	200	unlinked	M1409CS5	198	Unlinked
M1412ES5	170	unlinked	M1412CS3	175	Unlinked
M0730BK3	143	unlinked	M0730BK5	140	2

Linkage group is given as identified during grouping of markers. Some of the markers are not shown in the linkage map as we could not find unique place in the map.

REFERENCES

Andreoli, C.and A. Khan. 1993. Improving papaya seedling emergence by matriconditioning and gibberellin treatment. Hort Science 28: 708-709.

Arkle, T.D.J.and H.Y. Nakasone. 1984. Floral differentiation in the hermaphroditic papaya. Hort Science 19: 832-834.

Arumuganathan, K.and E.D. Earle. 1991. Nuclear DNA content of some important plant species. Plant molecular biology reporter 9: 208-218.

Awada, M. 1958. Relationships of minimum temperature and growth rate with sex expression of papaya plants. Hawaii Agriculture Experiment Station Technical Bulletin 38: 1-16.

Awada, M.and W. Ikeda. 1957. Effects of water and nitrogen application on composition growth, sugars in fruits, yield and sex expression of the papaya plants. Hawaii Agr. Exp. Stn. Tech. Bull. 33: 16.

Badillo, V.M. 1971. Monographia de la familia Caricaceae, Maracay Venezuela.

Badillo, V.M. 2000. Carica Vs Vasconcellea St.Hil.(Caricaceae) Con la rehabiitacion de este vltimo 10(2): 74-79. Ernstia 10: 74-79.

Badillo, V.M. 2001. Nota correctiva Vasconcellea St. Hill. Y no Vasconcella. Ernstia 11: 75-76.

Basten, C., S. Wang, B. Weir, and Z. Zeng, 2001. QTL Cartographer version 1.21 for Windows Release 1.21., Department of Statistics, North Carolina State University, Raleigh, North Carolina.

Blum, E., M. Mazourek, M. O'Connell, J. Curry, T. Thorup, K.D. Liu, M. Jahn, and I. Paran. 2003. Molecular mapping of capsaicinoid biosynthesis genes and quantitative trait loci analysis for capsaicinoid content in *Capsicum.* Theoretical and Applied Genetics 108: 79-86.

Bousquet, J., W.M. Cheliak, and M. Lalonde. 1987. Allozyme variability in natural-populations of green Alder (*Alnus Crispa*) in Quebec. Genome. 29: 345-352.

Botstein, D., R.L. White, M. Skolnicks, and R.W.Davis. 1980. Construction of genetic linkage map in man using RFLP. American Journal of Human Genetics 32: 314-331.

Broman, K.W. 2003. Mapping quantitative trait loci in the case of a spike in the phenotype distribution. Genetics 163: 1169-1175.

Burnside, W. 1959.Theory of probability. Dover publications, New York.

Buerstmayr, H., M. Lemmens, L. Hartl, L. Deldi, B. Steiner, M. Stierschneider, and P.Ruckenbauer. 2002. Molecular mapping of QTLs for Fusarium head blight resistance in spring wheat. I.Resistance to fungal spread (Type II reistance). Theoretical and Applied Genetics 104: 84-91.

Burr, B., F.A. Burr, K.H. Thompson, M.C. Albertson, and C.W. Stuber. 1988. Gene-Mapping with Recombinant Inbreds in Maize. Genetics 118: 519-526.

Chandrika, U.G., E.R. Jansz, S.N. Wickramasinghe, and N.D. Warnasuriya. 2003. Carotenoids in yellow- and red-fleshed papaya (*Carica papaya* L). Journal of the Science of Food and Agriculture 83: 1279-1282.

Churchill, G.A.and R.W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. Genetics 138: 963-971.

Cunningham, F.X.and E. Gantt. 1998. Genes and enzymes of carotenoid biosynthesis in plants. Annual Review of Plant Physiology and Plant Molecular Biology 49: 557-583.

Cunningham, F.X., B. Pogson, Z. sun, K. McDonald, D. Dellapenna, and E.Gantt. 1996. Functional analysis of the β and ε lycopene cyclase enzymes of *Arabidopsis* reveals a mechanism for control of cyclic carotenoid formation. The Plant Cell 8: 1613-1626.

Devicente, M.C.and S.D. Tanksley. 1991. Genome-wide reduction in recombination of backcross progeny derived from male versus female gametes in an interspecific cross of tomato. Theoretical and Applied Genetics 83: 173-178.

Diaz, A.and M.R. Macnair. 1999. Pollen tube competition as a mechanism of prezygotic reproductive isolation between *Mimulus nasutus* and its presumed progenitor *M. guttatus*. New Phytologist 144: 471-478.

Doerge, R.W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. Nature Review Genetics 3: 43-52.

Doerge, R.W.and G.A. Churchill. 1996. Permutation tests for multiple loci affecting a quantitative character. Genetics 142: 285-294.

Don, R.H., P.T. Cox, B.J. Wainwright, K. Baker, and J.S. Mattick. 1991. Touchdown PCR to circumvent spurious priming during gene amplification. Nucleic Acids Research 19: 4008-4008.

Doyle, J.J.and J.L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11-15.

Edwards, A., A. Civitello, H.A. Hammond, and C.T. Caskey. 1991. DNA typing and genetic-mapping with trimeric and tetrameric tandem repeats. American Journal of Human Genetics 49: 746-756.

Fazio, G., J.E. Staub, and M.R. Stevens. 2003. Genetic mapping and QTL analysis of horticultural traits in cucumber (*Cucumis sativus* L.) using recombinant inbred lines. Theoretical and Applied Genetics 107: 864-874.

Fishman, L., A.J. Kelly, E. Morgan, and J.H. Willis. 2001. A genetic map in the *Mimulus guttatus* species complex reveals transmission ratio distortion due to heterospecific interactions. Genetics 159: 1701-1716.

Flandez-Galvez, H., P.K. Ades, R. Ford, E.C.K. Pang, and P.W.J. Taylor. 2003. QTL analysis for ascochyta blight resistance in an intraspecific population of chickpea (*Cicer arietinum* L.). Theoretical and Applied Genetics 107: 1257-1265.

Frary, A., S. Doganlar, M.C. Daunay, and S.D. Tanksley. 2003. QTL analysis of morphological traits in eggplant and implications for conservation of gene function during evolution of solanaceous species. Theoretical and Applied Genetics 107: 359-370.

Grodzicker, T., J. Williams, P. Sharp, and J. Sambrook. 1974. Physical mapping of temperature sensitive mutations of adenovirus. Cold Spring Harbour symposium of Quantitative Biology p. 439-446.

Hanotte, O., T. Burke, J.A.L. Armour, and A.J. Jeffreys. 1991. Cloning, characterization, and evolution of Indian peafowl *Pavo cristatus* minisatellite loci., p. 193-216. In: T. Burke, G. Dolf, A.J. Jeffreys, and R. Wolf (eds.). DNA fingerprinting: approaches and applications. BirkhauserVerlag AG, Basal, Switzerland.

Haldane, J.B.S. 1919. The combination of linkage values and the calculation of distances between the loci of linked factors. Journal of Genetics 8: 299-309.

Heusden, A.W.v., M.C. Jongerius, J.M.v. Tuyl, T.P. Straathof, and J.J. Mes. 2002. Molecular assisted breeding for disease resistance in lily.Proceedings of 20 EUCARPIA symposium on new ornamentals II.Eds. J. Van Huylenbroeck *et al.* Acta Horticulturae 572: 131-138.

Higgins, J.E.and V.S. Holt. 1914. The papaya in Hawaii. Hawaii Agriculture Experiment Station Bulletin 32: 44.

Hirschberg, J. 2001. Carotenoid biosynthesis in flowering plants. Current Opinion in Plant Biology 4: 210-218.

Hofmeyr, J.D.J. 1938. Genetical studies of *Carica papaya*. L. South African Journal Of Science 35: 300-304.

Hofmeyr, J.D.J. 1939. Sex-linked inheritance in *Carica papaya* L. South African Journal Of Science 36: 283-285.

Hofmeyr, J.D.J. 1949. Inheritance of dwarfness in *Carica papaya* L. South African Journal Of Science 45: 98.

Hugueney, P., A. Badillo, H.C. Chen, A. Klein, J. Hirschberg, B. Camara, and M. Kuntz. 1995. Metabolism of cyclic carotenoids - a model for the alteration of this biosynthetic-pathway in *Capsicum annuum* Chromoplasts. The Plant Journal 8: 417-424.

Jacob, H.J., K. Lindpaintner, S.E. Lincoln, K. Kusumi, R.K. Bunker, Y.P. Mao, D. Ganten, V.J. Dzau, and E.S. Lander. 1991. Genetic mapping of a gene causing hypertension in the stroke-prone spontaneously hypertensive rat. Cell. 67: 213-224.

Jansen, R.C. 1993. Interval mapping of multiple quantitative trait loci. Genetics 135: 205-211.

Jeffreys, A.J., V. Wilson, and S.L. Thein. 1985. Hypervariable minisatellite regions in human DNA. Nature. 314: 67-73.

Jenczewski, E., M. Gherardi, I. Bonnin, J.M. Prosperi, I. Olivieri, and T. Huguet. 1997. Insight on segregation distortions in two intraspecific crosses between annual species of *Medicago* (Leguminosae). Theoretical and Applied Genetics 94: 682-691.

Jia, J.H., D.S. Zhang, C.Y. Li, X.P. Qu, S.W. Wang, V. Chamarerk, H.T. Nguyen, and B. Wang. 2001. Molecular mapping of the reverse thermo-sensitive genic male-sterile gene (rtms1) in rice. Theoretical and Applied Genetics 103: 607-612.

Jones, C.J., E. K.J, S. Castiglione, M.O. Winfield, F. Sala, C.V.d. Wiel, B.L. Vosman, M. Matthes, A. Daly, R. Brettschneider, P. Bettini, M.Buiatti, E.Maestri, N. Marmiroli, R.L. Aert, G.Volckaert, J.Rueda, A.Vazquez, and A.Karp. 1998. Reproducibility testing of AFLP by a network of European laboratories, p. 176-211. In: A. Karp, P.G. Issac, and D.S. Ingram (eds.). Molecular tools for screening biodiversity. Chapman and Hall publishers, London.

Kang, B.C., S.H. Nahm, J.H. Huh, H.S. Yoo, J.W. Yu, M.H. Lee, and B.D. Kim. 2001. An interspecific (*Capsicum annuum x C. chinese*) F_2 linkage map in pepper using RFLP and AFLP markers. Theoretical and Applied Genetics 102: 531-539.

Karrer, P. and E. Jucker. 1950. Carotenoids of known structure containing hydroxyl groups. P. 171-214. In: Braude, E.A (Translated). Carotenoids. Elsevier publishing company, New York.

Keats, B.J.B., S.L. Sherman, N.E. Morton, E.B. Robson, K.H. Buetow, P.E. Cartwright, A. Chakravarti, U. Francke, P.P. Green, and J. Ott. 1991. Guidelines for human linkage maps - an international system for human linkage maps. Annals of Human Genetics 55: 1-6.

Kim, M.S., P.H. Moore, F. Zee, M.M.M. Fitch, D.L. Steiger, R.M. Manshardt, R.E. Paull, R.A. Drew, T. Sekioka, and R. Ming. 2002. Genetic diversity of Carica papaya as revealed by AFLP markers. Genome 45: 503-512.

Kim, S., M. Binzel, S. Park, K. Yoo, and L. Pike. 2004. Inactivation of DFR(Dihydroflavonol 4-reductase) gene transcription results in blockage of anthocyanin production in yellow onions. The American Society for Horticultural Science-2004 annual conference, Austin, Texas, July 17-20. p. 889.

Klein, P.E., R.R. Klein, S.W. Cartinhour, P.E. Ulanch, J.M. Dong, J.A. Obert, D.T. Morishige, S.D. Schlueter, K.L. Childs, M. Ale, and J.E. Mullet. 2000. A high-throughput AFLP-based method for constructing integrated genetic and physical maps: Progress toward a sorghum genome map. Genome Research 10: 789-807.

Knaap, E.V.D.and S.D. Tanksley. 2003. The making of a bell pepper-shaped tomato fruit: identification of loci controlling fruit morphology in yellow stuffer tomato. Theoretical and Applied Genetics 107: 139 - 147.

Knapp, S.J., J.L. Holloway, W.C. Bridges, and B.H. Liu. 1995. Mapping dominant markers using F_2 matings. Theoretical and Applied Genetics 91: 74-81.

Kosambi, D.D. 1944. The estimation of map distances from recombination values. Annals of Eugenetics 12: 172-175.

Korol, A.B., Y.I. Ronin, A.M. Itskovich, J.H. Peng, and E. Nevo. 2001. Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits. Genetics 157: 1789-1803.

Kruglyak, L.and E.S. Lander. 1995. A nonparametric approach for mapping quantitative trait loci. Genetics 139: 1421-1428.

Lander, E., P. Green, J. Abrahamson, A. Barlow, M. Daly, S. Lincoln, and L. Newburg. 1987. Mapamker: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1: 174-181.

Lander, E.and L. Kruglyak. 1995. Genetic dissection of complex traits - guidelines for interpreting and reporting linkage results. Nature Genetics 11: 241-247.

Lander, E.S.and D. Botstein. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-199.

Landry, B.S., R.V. Kesseli, B. Farrara, and R.W. Michelmore. 1987. A genetic-map of Lettuce (*Lactuca Sativa* L) with restriction-fragment-length-polymorphism, isozyme, disease resistance and morphological markers. Genetics 116: 331-337.

Launey, S.and D. Hedgecock. 2001. High genetic load in the pacific oyster *Crassostrea gigas*. Genetics 159: 255-265.

Lavi, U., P. Cregon, T. Schaap, and J. Hillel. 1995. DNA markers for identification and breeding. Plant Breeding Reviews 12: 195-226.

Lebot, V., M.K. Aradhya, and R.M. Manshardt. 1991. Geographic survey of genetic variation in Kava (*Piper methysticum* Forst. F. and *P. wichmannii* c.Ds.). Pacific science 45: 169-185.

Lee, M. 1995. DNA markers and plant breeding programs. Advances in Agronomy 5: 265-344.

Lefebvre, V., M. Kuntz, B. Camara, and A. Palloix. 1998. The capsanthin-capsorubin synthase gene: a candidate gene for the y locus controlling the red fruit color in pepper. Plant Molecular Biology 36: 785-789.

Lehmann, E.L. 1975. Nonparametrics. McGraw-Hill, NewYork.

Lehmensiek, A., A.M. Esterhuizen, D. Van Staden, S.W. Nelson, and A.E. Retief. 2001. Genetic mapping of gray leaf spot (GLS) resistance genes in maize. Theoretical and Applied Genetics 103: 797-803.

Li, Z.K., S.B. Yu, H.R. Lafitte, N. Huang, B. Courtois, S. Hittalmani, C.H.M. Vijayakumar, G.F. Liu, G.C. Wang, H.E. Shashidhar, J.Y. Zhuang, K.L. Zheng, V.P. Singh, J.S. Sidhu, S. Srivantaneeyakul, and G.S. Khush. 2003. QTL x environment interactions in rice. I. Heading date and plant height. Theoretical and Applied Genetics 108: 141-153.

Lincoln, S., M. Daly, and E. Lander, 1992a. Constructing genetic maps with MAPMAKER/EXP 3.0, Whitehead Institute, Cambridge,MA.

Lincoln, S., M. Daly, and E. Lander, 1992b. Mapping genes controlling quantitative traits with MAPMAKER/QTL 1.1, Whitehead Institute, Cambridge, MA.

Litt, M.and J.A. Luty. 1989. A Hypervariable microsatellite revealed by invitro amplification of a dinucleotide repeat within the cardiac-muscle actin gene. American Journal of Human Genetics 44: 397-401.

Liu, B.H. 1998. Statistical genomics: linkage, mapping, and QTL analysis. CRC press, Boca Raton, Florida.

Liu, Z.Y., P.H. Moore, H. Ma, C.M. Ackerman, M. Ragiba, Q.Y. Yu, H.M. Pearl, M.S. Kim, J.W. Charlton, J.I. Stiles, F.T. Zee, A.H. Paterson, and R. Ming. 2004. A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature 427: 348-352.

Lyttle, T.W. 1991. Segregation distorters. Annual Review of Genetics 25: 511-557.

Ma, H., P.H. Moore, Z.Y. Liu, M.S. Kim, Q.Y. Yu, M.M.M. Fitch, T. Sekioka, A.H. Paterson, and R. Ming. 2004. High-density linkage mapping revealed suppression of recombination at the sex determination locus in papaya. Genetics 166: 419-436.

Manshardt, R. 1999. 'UH Rainbow' papaya, a high-quality hybrid with genetically engineered disease resistance. University of Hawaii cooperative extension service, New Plants for Hawaii NPH-1.

Mauricio, R. 2001. Mapping quantitative trait loci in plants: Uses and caveats for evolutionary biology. Nature Review Genetics 2: 370-381.

Mester, D., Y. Ronin, D. Minkov, E. Nevo, and A. Korol. 2003. Constructing large-scale genetic maps using an evolutionary strategy algorithm. Genetics 165: 2269-2282.

Mester, D.I., Y.I. Ronin, Y. Hu, J. Peng, E. Nevo, and A.B. Korol. 2003. Efficient multipoint mapping: making use of dominant repulsion-phase markers. Theoretical and Applied Genetics 107: 1102-1112.

Michelmore, R.W., I. Paran, and R.V. Kesseli. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis - a rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of National Academy of Sciences of the United States of America 88: 9828-9832.

Ming, R., P.H. Moore, F. Zee, C.A. Abbey, H. Ma, and A.H. Paterson. 2001. Construction and characterization of a papaya BAC library as a foundation for molecular dissection of a tree-fruit genome. Theoretical and Applied Genetics 102: 892-899.

Morshidi, M. 1998. Genetic control of isozymes in *Carica papaya* L. Euphytica 103: 89-94.

Muranty, H. 1996. Power of tests for quantitative trait loci detection using full-sib families in different schemes. Heredity 76: 156-165.

Myburg, A.A., D.L. Remington, D.M. O'Malley, R.R. Sederoff, and R.W. Whetten. 2001. High-throughput AFLP analysis using infrared dye-labeled primers and an automated DNA sequencer. Biotechniques 30: 348-357.

Nakamura, Y., M. Leppert, P. Oconnell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin, and R. White. 1987. Variable number of tandem repeat (VNTR) markers for human-gene mapping. Science 235: 1616-122.

Nakasone, H.V.and Lamoureaux. 1982. Transitional forms of hermaphrodite papaya flowers leading to complete maleness. Jounal of American Society for Horticultural Sciences 107: 589-592.

Nakasone, H.V.and W.B. Storey. 1955. Studies on the inheritance of fruiting height of *Carica papaya* L. Jounal of American Society for Horticultural Sciences 66: 168-182.

Ott, J. 1991. Analysis of human genetic linkage. The Johns Hopkins University press, Baltimore, Maryland.

Paterson, A.H. 1996. Genome mapping in plants. Academic press. San Diego, CA.

Pejic, I., P. Ajmone-Marsan, M. Morgante, V. Kozumplick, P. Castiglioni, G. Taramino, and M. Motto. 1998. Comparative analysis of genetic similarity among maize inbred lines detected by RFLPs, RAPDs, SSRs, and AFLPs. Theoretical and Applied Genetics 97: 1248-1255.

Peng, J., A.B. Korol, T. Fahima, M.S. Roder, Y.I. Ronin, Y.C. Li, and E. Nevo. 2000. Molecular genetic maps in wild emmer wheat, *Triticum dicoccoides*: Genome-wide coverage, massive negative interference, and putative quasi-linkage. Genome Research 10: 1509-1531.

Prakash, S., R.C. Lewontin, and J.L. Hubby. 1969. Molecular approach to the study of genetic heterozygosity in natural populations. IV patterns of geneic variation in central marginal and isolated populations of *Drosophila pseudoobscura*. Genetics 61:841-858.

Prioul, S., A. Frankewitz, G. Deniot, G. Morin, and A. Baranger. 2004. Mapping of quantitative trait loci for partial resistance to *Mycosphaerella pinodes* in pea (*Pisum sativum* L.), at the seedling and adult plant stages. Theoretical and Applied Genetics 108: 1322-1334.

Raina, S.K. 1997. Doubled haploid breeding in cereals. Plant Breeding Reviews 15: 141-186.

Remington, D.L.and D.M. O'Malley. 2000. Whole-genome characterization of embryonic stage inbreeding depression in a selfed loblolly pine family Genetics 155: 337-348.

Roberts, P. 1965. Difference in the behaviour of eu- and hetero-chromatin. Nature 205: 725-726.

Rocha, E.P.C., A. Danchin, and A. Viari. 2001. Evolutionary role of restriction modification systems as revealed by comparative genome analysis. Genome Research 11: 946-958.

Rocherieux, J., P. Glory, A. Giboulot, S. Boury, G. Barbeyron, G. Thomas, and M.J. Manzanares-Dauleux. 2004. Isolate-specific and broad-spectrum QTLs are involved in the control of clubroot in *Brassica oleracea*. Theoretical and Applied Genetics 108: 1555-1563.

Ronen, G., L. Carmel-Goren, D. Zamir, and J. Hirschberg. 2000. An alternative pathway to beta-carotene formation in plant chromoplasts discovered by map-based cloning of Beta and old-gold color mutations in tomato. Proceedings of National Academy of Sciences of the United States of America 97: 11102-11107.

Ronen, G., M. Cohen, D. Zamir, and J. Hirschberg. 1999. Regulation of carotenoid biosynthesis during tomato fruit development: Expression of the gene for lycopene epsilon-cyclase is down-regulated during ripening and is elevated in the mutant Delta. The Plant Journal 17: 341-351.

Russell, J.R., J.D. Fuller, M. Macaulay, B.G. Hatz, A. Jahoor, W. Powell, and R. Waugh. 1997. Direct comparison of levels of genetic variation among barley accessions detected by RFLPs, AFLPs, SSRs and RAPDs. Theoretical and Applied Genetics 95: 714-722.

Saghai-Maroof, M.A., K.M. Soliman, R.A. Jorgensen, and R.W. Allard. 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. Proceedings of National Academy of Sciences of the United States of America 81: 8014-8018.

SAS Institute. 1995. *SAS/QC Software: Usage and Reference I and II.* Cary, NC: SAS Institute.

SAS Institute. 1999. *SAS Procedures Gudide, Version 8.* Cary, NC: SAS Institute.

Sax, K. 1923. The association of size differences with seed coat pattern and pigmentation in *Phaseolus vulgaris*. Genetics 8: 552-560.

Soltis, D.E., C.H. Haufler, D.C. Darrow, and G.J. Gastony. 1983. Starch gel electrophoresis of ferns: A compilation of grinding buffers, gel and electrode buffers, and staining schedules. American Fern Journal 73: 9-27.

Sondur, S.N. 1994. Construction of a genetic linkage map of papaya and mapping traits of horticultural importance. Ph.D. Dissertation, University of Hawaii at Manoa, Honolulu.

Sondur, S.N., R.M. Manshardt, and J.I. Stiles. 1996. A genetic linkage map of papaya based on randomly amplified polymorphic DNA markers. Theoretical and Applied Genetics 93: 547-553.

Staub, J.E., F. Serquen, and M. Gupta. 1996. Genetic markers, map construction, and their application in plant breeding. Hort. Science 31: 729-741.

Stiles, J.I., C. Lemme, S. Sondur, M.B. Morshidi, and R. Manshardt. 1993. Using randomly amplified polymorphic DNA for evaluating genetic-relationships among papaya cultivars. Theoretical and Applied Genetics 85: 697-701.

Storey, W.B. 1938. Segregations of sex types in solo papaya and their application to the selection of seed. Proceedings of American Society for Horticultural Science 35: 83-85.

Storey, W.B. 1941. The botany and sex relations of the papaya. Hawaii Agriculture Experiment Station Bulletin 87:5-22.

Storey, W.B. 1953. Genetics of papaya. Journal of Heredity 44: 70-78.

Storey, W.B. 1958. Modification of sex expressions in papaya. Horticultural Advance 2: 49-60.

Storey, W.B. 1967. Theory of derivations of the unisexual flowers of Caricaceae. Agronomia Tropical (Maracay) 17: 273-321.

Storey, W.B. 1969. Papaya. In: F.P. Ferwerda (ed.). Outlines of perennial crop breeding in the tropics. H. Veenman and N.V. Zonen, Wageningen, The Netherlands.

Stuber, C.W. 1992. Biochemical and molecular markers in plant breeding. Plant Breeding Reviews 9: 37-60.

Subadrabandhu, S.and C. Nontaswatsri. 1997. Combining ability analysis of some characters of introduced and local papaya cultivars. Scientia Horticulturae 71: 203-212.

Sybenga, J. 1996. Recombination and chiasmata: few but intriguing discrepancies. Genome 39: 473-484.

Tanksley, S.D. 1983. Molecular markers in plant breeding. Plant Molecular Biology Reporter 1: 3-8.

Tanksley, S.D. 1993. Mapping Polygenes. Annual Review of Genetics 27: 205-233.

Tanksley, S.D. and C.M. Rick. 1980. Isozymic gene linkage map of the tomato - applications in genetics and breeding. Theoretical and Applied Genetics 57: 161-170.

Tanksley, S.D.and C.M. Rick. 1982. Isozyme linkage map of the tomato: applications in genetics and breeding. Theoretical and Applied Genetics 57: 161-170.

Tar'an, B., T. Warkentin, D.J. Somers, D. Miranda, A. Vandenberg, S. Blade, S. Woods, D. Bing, A. Xue, D. DeKoeyer, and G. Penner. 2003. Quantitative trait loci for lodging resistance, plant height and partial resistance to mycosphaerella blight in field pea (Pisum sativum L.). Theoretical and Applied Genetics 107: 1482-1491.

Thoday, J.M. 1961. Location of polygenes. Nature 193:168-370

Thoquet, P., M. Ghérardi, E. Journet, A. Kereszt, J. Ané, J. Prosperi, and T. Huguet, 2002. The molecular genetic linkage map of the model legume Medicago truncatula: an essential tool for comparative legume genomics and the isolation of agronomically important genes, BMC Plant Biology. (http://www.biomedcentral.com/1471-2229/2/1)

Van Ooijen, J.W., 2004. MapQTL Version 5.0, Software for the mapping of quantitative trait loci in experimental populations, Plant Research International B.V. and Kyazma B.V., Wageningen, The Netherlands.

Van Ooijen, J.W., H. Sandbrink, C. Purimahua, R. Vrielink, R. Verkerk, P. Zabel, and P. Lindhout. 1993. Mapping quantitative genes involved in a trait assessed on an ordinal scale: A case study with bacterial canker in *Lycopersicon peruvianum.*, p. 59-74. In: J.I. Yoder (ed.). Molecular biology of tomato. Technomic Publishing Co. Inc., Lancaster, PA.

Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Vandelee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. Aflp - a New Technique for DNA-Fingerprinting. Nucleic Acids Research 23: 4407-4414.

Weber, J.L.and P.E. May. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain-reaction. American Journal of Human Genetics 44: 388-396.

Welsh, J.and M. McClelland. 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Research 18: 7213-7218.

Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey. 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic-markers. Nucleic Acids Research 18: 6531-6535.

Xu, M.L.and S.S. Korban. 2000. Saturation mapping of the apple scab resistance gene Vf using AFLP markers. Theoretical and Applied Genetics 101: 844-851.

Yamamoto, H.Y. 1964. Comparison of the carotenoids in yellow and red *Carica papaya.* Nature 201: 1049-1050.

Yin, T.M., S.P. DiFazio, L.E. Gunter, D. Riemenschneider, and G.A. Tuskan. 2004. Large-scale heterospecific segregation distortion in populus revealed by a dense genetic map. Theoretical and Applied Genetics 109: 451-463.

Zeng, Z.B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proceedings of National Academy of Sciences of the United States of America 90: 10972-10976.

Zeng, Z.B. 1994. Precision mapping of quantitative trait loci. Genetics 136: 1457-1468.

Zhang, L.P., A. Khan, D. Nino-Liu, and M.R. Foolad. 2002. A molecular linkage map of tomato displaying chromosomal locations of resistance gene analogs based on a *Lycopersicon esculentum x Lycopersicon hirsutum* cross. Genome 45: 133-146.

Zou, F., B.S. Yandell, and J.P. Fine. 2003. Rank-based statistical methodologies for quantitative trait locus mapping. Genetics 165: 1599-1605.