UNIVERSITY OF HAWAII LIBRARY

MAPPING PHYSIOLOGICAL TRAITS IN *CARICA PAPAYA* USING MICROSATELLITE MARKERS

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

MASTER OF SCIENCE

IN

MOLECULAR BIOSCIENCES AND BIOENGINEERING

DECEMBER 2006

By
Olivia J. Veatch

Thesis Committee:

Ray Ming, Chairperson
Qingyi Yu
Dulal Borthakur
Paul H. Moore
Robert Paull

We certify that we have read this thesis and that, in our opinion, it is satisfactory in scope and quality as a thesis for the degree of Master of Science in Molecular Biosciences and Bioengineering.

THESIS COMMITTEE

Chairperson

ACKNOWLEDGEMENTS

I am very grateful to Dr. Ray Ming for giving me this opportunity, without whom, I would never have been able to accomplish the following work. I am also incredibly thankful to Dr. Qingyi Yu for supervising me during the majority of my graduate work. I would like to thank Dr. Yu for all of her sound advice and for being such an understanding and enjoyable person to work with in the lab. Thank you to Dr. Paul H. Moore for always being more than willing to lend his vast knowledge, Dr. Robert Paull for all of his invaluable input on the project and Dr. Dulal Borthakur for being so implemental in the entire process, I could not have finished without him.

I would like to thank USDA T-STAR program at the University of Hawai'i (UH) and the Hawai'i Agriculture Research Center (HARC) for their financial support during my graduate study. I would also like to thank the entire community at HARC and the Department of Molecular Biosciences and Bioengineering at UH for every bit of valuable guidance and advice given.

I am very lucky to have had lab colleagues of such caliber. Thank you especially to Cuixia Chen, Rachel Ostroff, Moriah Eustice, Andrea Blas, Annie Byers, Eric Tong, Taesik Uhm, Daniel Adamski, Meghan Jones and Jan Murray for everything they have done for me over the past two years.

Personally, I would like to express gratitude to my mother, Shirley Veatch, my entire family, James Coyle and Wendy Ullmer for their incredible support and belief in me.

ABSTRACT

Different varieties of papaya (*Carica papaya* L.) vary in the phenotypic expression of agronomically important traits. Genetic loci responsible for these differences can be mapped using DNA markers to genotype a segregating progeny population derived from a controlled cross between parents having different phenotypes.

Two unique varieties of papaya, Khak dam and 2H94, were chosen as parental materials based on their widely varied phenotypic expression of fruit size, shape, and flesh color. These varieties were used to produce an F2 population that was surveyed for DNA polymorphisms using microsatellite markers. The 100 female and hermaphrodite F2 individuals segregating for flesh color were scored for this trait. Flesh color segregated in a 3:1 ratio with yellow flesh color being dominant suggesting that flesh color is a single gene controlled trait. All of the 137 hermaphrodite individuals in the F2 population were scored for fruit weight, length, and width. Fruit phenotype data was analyzed by constructing histograms to assess the distribution of these traits. Normal distribution was detected for length and width suggesting quantitative inheritance for these traits. Fruit weight may segregate as a mix of normal and bimodal distribution suggesting involvement of a major gene modified by quantitative trait loci (QTL).

A total of 1497 microsatellite markers screened for polymorphisms between AU9 female and Sunup hermaphrodite parents showed 607 of them to be polymorphic. Of these, 199 SSR markers were used to genotype the F2 progeny population in order to construct a preliminary genetic linkage map. From this map, a total of 38 markers

spaced, on average, 25cM apart across 16 linkage groups were found to also be polymorphic between Khak dam and 2H94 and chosen for genotype screening.

To identify major gene loci and QTLs controlling these traits, analysis of variance and MapQTL® 5.0 were used. The map location of markers correlating to QTLs involved in expression of fruit size, or weight, fruit shape, or length and width, and the major gene locus for flesh color were estimated. Significant QTLs affecting weight mapped to linkage group 6, QTLs affecting length to linkage group 7, QTLs affecting width to linkage group 6 and QTLs affecting shape to linkage group 7. The location for the gene affecting flesh color was determined to be located close to marker P3K132 on linkage group 1. These sequence specific DNA markers linked to target traits will be fine mapped for exact location and used for marker-assisted breeding system in the future.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:

INTRODUCTION

Papaya is a principal fruit crop of tropical and subtropical regions worldwide. In 2003, papaya sales made up 56% of all tropical fruit sales for Hawai'i. They are the second largest fruit crop for the state after pineapple. The value of utilized production of papaya for the United States over the past three years ranged from 11 million to 13 million dollars, with Hawai'i contributing the majority (www.nass.usda.gov/hi/stats/stat-24.htm). The successful release of Papaya ringspot virus-resistant papayas has increased the acceptance of selective breeding and genetic modifications of papayas for commercial production (Fitch et al. 1992).

Molecular markers can be useful in the construction of genetic linkage maps to help in identifying and tagging selected genes in organisms. Amplified fragment length polymorphisms (AFLPs) and simple sequence repeats (SSRs) are two of the most commonly used markers for these applications. AFLP markers are mostly dominant markers that are generated without sequence data using polymerase chain reaction (PCR) by amplifying fragments of DNA created by restriction enzymes. SSRs are the sequences of DNA where one to a few bases are tandemly repeated for few to hundreds of times. Co-dominant variations in these repeated regions are detected among F2 individuals by designing oligonucleotides to amplify the repeated sequences (Mohan et al. 1997). Using both AFLPs and SSRs, and the maps created from them, it is possible to locate genes involved in qualitative as well as quantitative agronomically important traits.

To determine the locus or loci responsible for traits of interest, a cross must be made between F1 progeny that will produce an F2 population segregating for the traits of interest or an F1 must be crossed back to one of the original parents to produce a backcross population segregating for the traits of interest (Boluarte-Medina and Veilleux, 2002). A linkage map must then be constructed by genotyping either the F2 or backcross population using polymorphic DNA markers. Linkage maps can be easily generated by inputting data into a computer program designed for map creation. When a linkage map is available, marker-based trait analysis can be accomplished. Phenotype data collection should be done by assessing each trait of interest and determining whether the trait is qualitative or quantitative. Qualitative traits are typically monogenic and are expressed with the phenotype as discrete categories. Quantitative traits are multigenic and show continuous variation (Sargent et al. 2004). After collecting phenotypic data to assess the traits of interest, the average phenotypic variable for each tree in a population is plotted against the frequency of trait expression among the entire population. If a trait is quantitative the histogram will show normal distribution. If a trait is qualitative the histogram will show a bimodal or skewed distribution.

Once phenotype data is analyzed and the linkage map is created, a correlation between the genotype and trait phenotype for each studied marker can be determined using statistical analysis of variance. The marker(s) that show significant correlation between genotype and phenotype are considered the putative locus or loci that affect the trait of interest (Broman, 2001). Other methods commonly used to locate significant locus or loci are interval mapping methods. Interval mapping correlates genotype and

phenotype information for each centimorgan (cM) location on a linkage group including known markers. The location on the map with the highest significant logarithm of odds (LOD) score is considered the location of the gene(s) affecting the trait being studied (Van Ooijen, 1999; 2004).

For marker-assisted selection, fine mapping of each trait must be done in order to verify the purported results (Paterson et al. 1990; Xie et al. 2006). Eventually, it will be possible to selectively breed papaya to produce the most marketable fruit for a certain region of interest, thereby increasing the value of papaya production in the United States and in all tropical and subtropical regions world-wide.

CHAPTER 2:

LITERATURE REVIEW

2.1 PAPAYA

2.1.1 Uses and Importance

Papaya trees are grown for their flowers, fruit and leaves as well as for the enzyme papain. Papaya is a healthy choice for consumption being high in potassium content, vitamins A and C and low in fat (Chan and Paull, 2007). These vitamins are necessary for healthy skin and proper bone development (Morton, 1987). Papaya fruit are also a source of antioxidants shown to be involved in the prevention of various cancers, age-related macular degeneration and multiple sclerosis (Edge et al. 1997). Other parts of papaya are also consumed, including the young leaves that can be cooked and eaten like spinach and the seeds that are used for their peppery flavor (Morton, 1987). In traditional medicines, different parts of papaya have been found useful for their antibiotic, anti-inflammatory and immuno-stimulatory activity (Mojica-Henshaw et al. 2003).

The enzyme papain is a milky white latex secreted by green papaya. The papaya fruit uses papain possibly as a natural defense mechanism against predators. Papain is a cysteine proteolytic enzyme and is used commercially in beer brewing to prevent cloudiness caused by precipitation of proteins during chilling (Cobley, 1976).
It is also used industrially as a meat tenderizer, in leather tanning and in shrink resistant processes for wool and silk (Dunne and Horan, 1992).

In addition to the importance of papaya as a major fruit crop, it may be ideal to use as a genomic model system for all tropical fruit crops due to its very unique characteristics. These characteristics include the ease of crossing due to hand pollination techniques as well as seed propagation. Trees flower and fruit continuously throughout the year allowing for crosses to be made often and each cross has the ability to produce large numbers of seed for study (Ming et al. 2001). It has a small genome of only 372 Mbp with diploid inheritance and a total of nine chromosomes (Arumuganathan and Earle, 1991).

2.1.2 Taxonomy

Papaya is a dicotyledonous diploid species belonging to the plant family *Caricaceae*. It is thought to be native to Central America and was brought to Southeast Asia during the 16th century by the Spanish. From Southeast Asia, it spread throughout the tropical regions of the world (Chin, 2003). The *Caricaceae* family is small with seven genera and about 32 species with *Carica papaya* being the only species in the genus *Carica* (Decraene and Smets, 1998; Chan and Paull, 2007). Other genera are *Vasconcella* and *Jacaratia* from South America, *Jarilla* from Mexico and Guatemala, *Horovitzia* from Mexico and *Cylicomorpha* from Africa (Van Droogenbroeck et al. 2002). An AFLP analysis of genetic diversity among selected papaya species showed that the genetic diversity was limited. A finding that does not support the observed high levels of phenotypic diversity found in the field (Kim et al. 2002). The family *Caricaceae* is closely related *Brassicaceae* that includes Arabidopsis (Bremer et al. 1998) whose genome is completely sequenced, another reason papaya may be ideal for genomic analysis.

2.1.3 Botany and Physiology

Papaya trees are propagated by seed and are either sown directly in the field or first sown in a nursery and then planted in the open field (Chan and Paull 2007). On average, each papaya fruit produces approximately 800-1200 seeds and each fruit bearing tree produces hundreds of fruits in its lifetime making it incredibly easy to propagate (Ming et al. 2001). Each tree typically grows from 6 to 9 meters in height with a hollow trunk that ranges from 30 to 40 centimeters in thickness at the base. The leaves emerge from the upper stem on a long hollow petiole and have a deeply divided palmate blade consisting of 5 to 9 lobes.

Papaya is polygamous having three sex types. The genetic mechanism for sex determination in papaya is still under review but the most current understanding is that there are two slightly different Y chromosomes and that at least two genes differentiate these two Y chromosomes (Ming et al. unpublished data). The two Y chromosomes are Y_1 controlling dioecious papaya and Y_2 controlling gynodioecious papaya. Dioecious papaya produce male and female trees and gynodioecious papaya produce hermaphrodite and female trees. This would make the tentative genotypes male (XY_1), female (XX) and hermaphrodite (XY_2) (Storey, 1953; Horovitz and Jimenez, 1967). The sex of a tree can be typed early in development according to flower morphology and then eventually by fruit shape. A strong dimorphism exists between staminate, or male, and pistillate, or female, flowers. Staminate flowers are recognized by the existence of small sepal lobes, a long petal tube and a central reduced ovary. The female flowers have no trace of stamens but instead grow five carpel primordia opposite the sepals. Basal growth

eventually leads to the development of a large ovary. Female flowers can be distinguished from staminate flowers in that they are much larger, have petal lobes fused only at the base, and lack stamens (Decraene and Smets, 1998). Bisexual flowers, otherwise known as hermaphrodites, also arise and resemble male flowers but with a swollen base and unlike male flowers are capable of producing fruit. The bisexual flowers are primarily self-pollinated, often before the flower opens. Pistillate flowers can be pollinated by wind and insects considered pollinators like beetles, moths, flies and mosquitoes. Hand pollination can be easily done by removing the stamens before anthesis and placing a paper bag over the flower and then transferring the pollen to another flower's pistil. In order to make a controlled cross or to produce the maximum number of fruit-bearing trees possible and accurately predict the sex of the progeny, hand pollination is necessary (Malo and Campbell, 1994).

Once a flower is pollinated, the content of the hormone auxin in the ovary increases to initiate fruit growth and inhibit flower stalk abscission (Loveless, 1983). It generally takes four to five months for fruit to mature in the tropics and in the subtropics (Nakasone and Paull, 1998). Papaya fruit begins a dark green color. Once the fruit has finished growing to capacity color break begins at the blossom end of the fruit and proceed toward the stem turning the fruit from green to yellow.

The fruit expresses a broad range of phenotypes. Among the different papaya varieties, fruit range in shape from oval and pear-shaped to an elongated club-shape that can be 7 to 50 centimeters long and 4 to 20 centimeters thick. The fruit ranges in weight from as small as 30 grams to as large as 9 kilograms.

Papaya, as the only species in the genus, is genetically and reproductively isolated from other members of the *Caricaceae* family making it difficult to use traditional breeding methods as a way of incorporating agronomically important traits from those family members into adapted cultivars. The offspring of inter-generic crosses are often sterile with drastic reductions in fitness levels, and without acceptable productivity and quality of fruit (Samson, 1980). Inter-generic crosses are not a viable means for obtaining disease resistance or improving agronomic traits such as flesh color and fruit size and shape. This makes it necessary to seek out other means for improving these traits, including identifying the genes involved in desirable traits using molecular mapping.

2.1.4 Cultivars

The cultivars used in this experiment include the Israeli variety AU9, the Hawai`ian varieties Sunup and Kapoho, the mutant variety 2H94, and the Thai variety Khak dam. Sunup, Kapoho, Khak dam and 2H94 are gynodioecious and AU9 is dioecious. Sunup produces fruit with red flesh and is a transgenic, Papaya ringspot virus-resistant, cultivar derived from the Hawai`ian cultivar Sunset (Fitch et al. 1992). Kapoho, like Sunup, is a 'Solo' type Hawai`ian cultivar producing fruit with yellow flesh color. 2H94 was a mutant found among the F2 population of a Sunup and Kapoho cross. The fruit produced by 2H94 are very small with weights ranging from 40g to 200g, lengths from 7cm to 10cm and widths from 3cm to 8cm. The flesh is yellow. 2H94 has been selfed for four generations and has maintained a stable phenotype. AU9 is Israeli variety producing round fruit with yellow flesh color. Khak dam is a cultivated Thai

variety producing large fruit with weights ranging from 900g to 1500g, lengths from 17cm to 30cm, widths from 8cm to 14cm and red flesh color.

2.1.5 Genomics Advancements

An early genetic linkage map was constructed using a total of 62 random-amplified polymorphic DNA (RAPD) markers mapped to 11 linkage groups (Sondur et al. 1996). A later genetic map of 1775 AFLP loci and a 13.7 X bacterial artificial chromosome (BAC) library were constructed and can be used as a foundation to characterize the papaya genome (Ma et al. 2004; Ming et al. 2001). The construction of the BAC library enabled the sequencing of BAC ends. A total of 35,472 BAC ends, accounting for 4.7% of the papaya genome, were sequenced (BES) (Lai et al. 2006). Various genetic markers, including SSRs and sequence characterized amplified region (SCAR) markers, have been developed to test the sex type of papaya seedlings (Deputy et al. 2002; Parasnis et al. 1999). Five cDNA libraries were constructed from staminate, pistillate and bisexual flower buds at the different developmental stages. These libraries were used to sequence expressed sequence tags (ESTs) that can be utilized as anchors for a linkage map (Barkley et al. 2005). Whole genome sequencing is currently being accomplished by the Hawai'i Papaya Genome Consortium using the whole genome shotgun sequencing approach.

2.2 GENETIC MAPPING

2.2.1 Molecular Markers

Mapping and sequencing of plant genomes can be very beneficial when elucidating gene function, regulation and expression. Molecular markers are used to

develop these maps and to identify and tag desired genes. They have many advantages over traditional phenotypic markers, since they are not environmentally regulated allowing them to be unaffected by the conditions in which a plant is grown. They are also detectable in all stages of plant growth. They offer a great opportunity for improving the efficiency of conventional plant breeding methods by enabling the phenotypic selection of important traits by means of indirect selection based on markers linked to the gene(s) of interest instead of directly on the trait of interest (Gallais, 1996 as cited in Billotte et al. 2005). The establishment of a linkage map for a certain crop allows the identification of the chromosomal portion(s) containing the genes responsible for a quantitative trait by relating the molecular polymorphism of the markers with the phenotypic variation of the trait (Charcosset, 1996). Two major types of markers are used for genetic mapping: amplified fragment length polymorphisms (AFLP) and microsatellites or simple sequence repeats (SSRs).

AFLP markers are generated from PCR amplification of restriction fragments. These fragments are the result of a digestion of template DNA using specific restriction enzymes whose cut sequence is well known. After digestion, the fragments are ligated to oligonucleotide adapters containing the sequences of the restriction sites. These fragments are amplified with oligonucleotides homologous to the adapters with an addition of one to three base pairs on the 3' end. This process accounts for the selective nature of AFLP analysis. Since the 3' extensions are not entirely homologous to the adapters, only DNA fragments complementary to the extensions are amplified (Liscum and Oeller, 2004). The main advantage of using AFLP to determine markers is its

capacity for the simultaneous screening of many unique DNA regions distributed randomly throughout a genome. AFLP methods rapidly generate hundreds of markers that are easily replicated. These polymorphic markers are genome-wide and can be generated without prior sequence knowledge. AFLP markers have been used extensively for constructing linkage maps for QTL analyses of agronomic plant traits such as disease resistance and salt tolerance (Mueller and Wolfenbarger, 1999).

AFLP is not without its drawbacks. First, the markers are somewhat expensive to generate since the bands are detected by silver staining, fluorescent dye or radioactivity (Mohan et al. 1996). Also, AFLP markers generally capture only the dominant loci which may not be present in different mapping populations (Miao et al. 2005). The frequency of co-dominant AFLP markers is very low, only about 4% to 15% of all polymorphic markers (Mueller and Wolfenbarger, 1999). This means that multilocus markers are scored as either present or absent making artificial or failed amplification of a fragment a large threat to the reliability and accuracy of the resulting maps. This makes AFLP more useful for saturation mapping rather than for linkage map creation (Billotte et al. 2003).

SSRs are repeats of one to six base pair (bp) DNA motifs that are arranged serially and frequently exhibit variation at a specific locus (Temnykh et al. 2001). The hypervariability of SSRs is a distinguishing feature resulting from the expansion and contraction of the repeated motif (Chen et al. 2002). This feature enables SSRs to provide a useful source of polymorphic DNA markers for connecting genetic maps with genomic sequences and ultimately with phenotypic variation (Temnykh et al. 2001).

SSR markers are identified by either screening SSR-enriched genomic libraries, by hybridizing oligonucleotides and then sequencing, or searching genome sequences in DNA sequence databases (Sharapova, 2002). SSR-based mapping is more desirable than AFLP-based mapping because of the co-dominant nature and abundance of SSRs, allow for the creation of denser genetic maps. This is attributed to the fact that polymorphisms in these repeated regions can occur due to strand slippage. Strand slippage is a mutation event that occurs at a much higher rate than does single nucleotide mutations, insertions, or deletion events, which are two major ways to generate AFLP markers (Belaj et al. 2003). Determining polymorphic markers using SSRs can be a very inexpensive method. One major disadvantage to using SSR markers for linkage mapping as well as genetic analysis in crop plant improvement is the requirement for available sequencing data. When sequence data is available, SSR markers appear to be the best method for linkage map development as well as marker-assisted selection for traits of interest. SSR-based maps have been developed for many species with available sequence data including *Arabidopsis thaliana* (Bell and Ecker, 1994), maize (Sharopova et al. 2002), oil palm (Billotte et al. 2005), and rice (Temnykh et al. 2000). SSR marker-based map construction is done by simple PCR reactions where oligonucleotides are designed to amplify the determined variable repeated region from its flanking sequences in each closely related individual. Genetic differences among these individuals are detected by comparison of SSR allele sizes measured as migration distances on agarose gels (Mohan et al. 1997; Chen et al. 2002). Another advantage of SSR markers is they are generally co-dominant, solving the issue of inaccurate genotype scoring that could occur when

using only AFLP markers. SSR markers, like AFLP, are rapidly detected and highly reproducible. SSR markers have also been used previously to map and characterize genes involved in agronomically important traits for crops such as barley (Dahleen et al. 2005) and bamboo (Barkley et al. 2005).

2.2.2 Linkage Map Construction

After molecular markers and genotypes have been determined for a population, but before genes can be tagged for traits of interest, a linkage map must be constructed as a template for each marker's specific location in a genome. Polymorphic markers are linked to each other using a logarithm of odds score (LOD). The LOD score for each marker is determined by using the following mathematical calculation:

$$\log \left[\frac{(1 - \text{recombination frequency}/\#\text{ parents})^{\text{parental type}} (1 - \text{recombination frequency}/\#\text{ parents})^{\text{recombination type}}}{(1 - 0.5 / \#\text{ parents})^{\text{total progeny}}} \right]$$

Where the divisor is representative of the probability for a gene to be linked and the dividend is representative of the probability for a gene to be unlinked. The recombination frequency is used to determine the map distance in cM for each marker. Recombination frequency is calculated by dividing the number of recombinant individuals by that of the total population and the map distance is determined by multiplying the quotient by 100. Markers that are determined as linked according to their LOD scores are separated into groups called linkage groups. An accurate linkage map should contain the same number of linkage groups as there are chromosomes in the

organism being mapped (van Ooijen, 2000). Linkage maps can be easily constructed using computer software programs. Some programs that have been previously utilized for linkage map creation in crops are JoinMap® 3.0, used to create a map for oil palm (Billotte et al. 2005) and Mapmaker® 3.0, used to create a map for papaya (Ma et al. 2004).

2.2.3 Marker-based Trait Analysis

Simple Mendelian traits are traits that segregate in accordance with the laws of genetics set forth by Gregor Mendel who derived these laws by studying single-gene traits in peas. Single gene traits are those where the relationship between genotype and phenotype can be observed directly. If a trait is controlled by a single gene and a parent expressing the trait is crossed with a parent not expressing the trait then a dominant ratio would be observed for expression of this trait in the phenotype of the offspring and in the marker genotypes. However, most gene expression is not the result of a single gene functioning alone. Traits can be influenced by environmental factors, referred to as multifactorial traits and/or by other genes, referred to as quantitative or complex traits (Lewis, 2001). A marker-based analysis of phenotypic traits consists of measuring the correlation between the phenotype and genotype score of marker loci or nearby unmarked loci affecting the quantitative trait.

QTL analysis is usually done by statistical analysis of variance (ANOVA) and/or interval mapping. ANOVA is done by performing statistical tests based solely on single DNA marker information. The calculations are based on phenotypic means and variances within each of the genotypic classes (Coffman et al. 2003). In interval

mapping, a population segregating for the trait of interest is created and a linkage map is calculated. Along the linkage map, at each cM, a LOD score is calculated along with the genetic effects and residual variance. The regions of the genome that contain QTLs are identified when they show significant LOD score values above a determined LOD significance threshold. The LOD significance threshold is determined by doing permutation tests that calculate the frequency distribution of the maximum LOD score over a set of iterations where the QTL data are permuted over the genotyped individuals while the marker data remain fixed. A sufficient amount of iterations is 1,000 and a P-value must be set to equal or less than 0.5 to give at least a 95% accuracy rate. An estimate of the frequency distribution of the maximum LOD score under the null hypothesis that there is no QTL is obtained. The interval of the permutation test where the relative cumulative count is 1 minus the p-value is the significance threshold for the linkage group tested. Any LOD score on this linkage group that is above the significance threshold is a tentative location of the QTL and the highest of these is most likely the location. Approximate multiple-QTL mapping methods (MQM) are more powerful than single-QTL methods when there are several segregating QTLs. MQM mapping is similar to interval mapping except that markers determined significant from interval mapping are chosen as cofactors to take over the roles of nearby QTLs. These cofactors are genetic background controls that absorb most of the genetic effects of their nearby QTLs from the residual variance thereby enhancing the power of the analysis (van Ooijen, 1999 & 2004).

QTL interval and MQM mapping can be done by using computer programs such

as MapQTL® 5.0. Many QTLs and major genes have been identified by using molecular markers and MapQTL programs in numerous species including arabidopsis (El-Lithy et al. 2005), onion (McCallum et al. 2006), honey bees (Lapidge et al. 2002), and mice (Srivastava et al. 2004).

2.3 Quantitative and Qualitative Trait Analysis

Many traits of agronomic and horticultural interest are controlled by a single gene and fall into a few distinct phenotypic classes. A trait controlled by a single gene is referred to as a qualitative trait. A qualitative trait is expressed with the phenotype falling into discrete categories. These categories do not necessarily have a rank order. The pattern of inheritance for a qualitative trait is typically monogenic, however there is a possibility that the single gene controlled trait is modified by other genes. Multimodal distribution is when there is more than one value that occurs more frequently in the data. If a trait is qualitative and two parents, each with unique expression of the trait, are crossed then the distribution plot of the trait expression in the F2 progeny should show bimodal distribution (Fig. 1) (Khazanie, 1996).

Other traits do not fall into discrete classes. When analyzing a segregating population for these traits, a continuous distribution is found. These types of traits are called continuous traits and cannot be analyzed in the same manner as discontinuous traits. Continuous traits are often given a quantitative value and are therefore referred to as quantitative traits and the loci controlling these traits are called quantitative trait loci (QTL). QTLs are genetic markers detecting regions of the DNA that are located in close physical proximity to genes controlling traits of interest. The tests for QTLs are known

as linked tests. These linked markers may be any of several types, including SSRs. The QTL model is based on the assumptions that each contributing gene has small, relatively equal effects, the effects of each allele are additive, there is no dominance and there is no interaction among the different loci contributing to the value of the trait. Normal distribution (Fig. 2) occurs when the averages of a certain trait distribute evenly with the larger and smaller sizes being observed among a lesser number of individuals (Khazanie, 1996). When traits are controlled by QTLs, the majority of the trees surveyed will show expression in between the average trait expression of each parent.

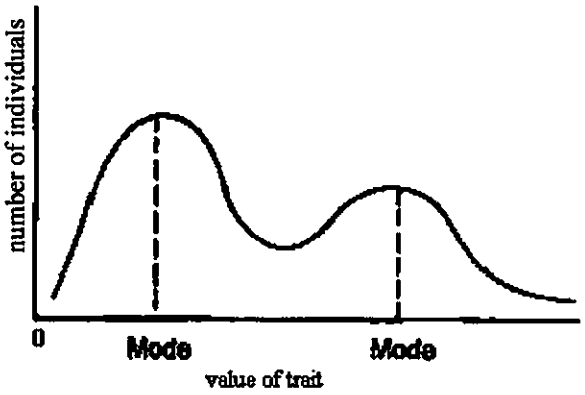


Figure1: Example of bimodal distribution patterns that should be expressed when the trait of interest is controlled by a qualitative trait locus.

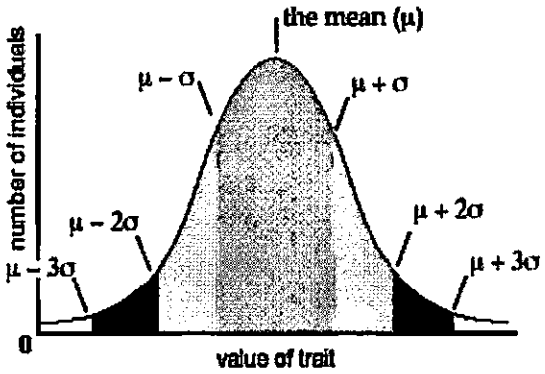


Figure 2: Example of normal distribution patterns that should be expressed when the trait of interest is controlled by QTLs. Where σ = standard deviation from mean.

Correlation among phenotypic traits can be assessed by computing a linear correlation coefficient (r) based on sample data. This computed coefficient determines whether or not two traits are linearly related. If r is equal to 1 then the traits show perfect positive correlation, on the other hand if r is equal to -1 then the traits show perfect negative correlation. The closer the r value is to 1, the more closely related are the two

traits being assessed. If the r value is near or equal to zero, then the traits are not related. An r value that is greater than zero but less than 1 shows positive correlation. This means that the trait plotted on the y-axis increases as the trait plotted on the x-axis increases (Khazanie, 1993).

2.4 Research Hypothesis

The range of fruit weight for the Khak dam X 2H94 F1 progeny is similar to the range of fruit weight for the Khak dam trees suggesting that a single gene mutated in 2H94. Fruit length, width and shape for the F2 progeny is highly varied with some trees producing long, thin fruits and other trees producing shorter, wide fruits. Flesh color of Khak dam trees is red while 2H94 is yellow. The flesh color in the F1 trees segregates indicating that the original 2H94 used in the cross is heterozygous at the flesh color locus. Therefore, fruit weight will be governed by a major gene that is modified by quantitative trait loci while fruit length, width and shape will be controlled by quantitative traits and flesh color will be a monogenic trait.

CHAPTER 3:

MATERIALS AND METHODS

3.1 Papaya Crosses

SSR markers were tested for polymorphism using five parental varieties of papaya: Kapoho, Sunup, AU9, Khak dam, and 2H94. Crosses are listed with the female parent followed by the hermaphrodite or male parent. This is done to normalize genotype scoring of the progeny.

First, a linkage map had to be created in order to locate markers to screen for correlation between fruit size, shape and flesh color phenotypes and genotypes. AU9 was crossed with Sunup and polymorphic markers were selected for high density linkage mapping using this mapping population because it had the highest estimated polymorphism rate. The parental lines Khak dam and 2H94 were crossed to produce the F2 progeny used to locate QTLs affecting fruit size and shape and the gene affecting flesh color. The F2 population derived from Khak dam X 2H94 is ideal for studying gene(s) controlling fruit size and shape. The dramatic difference of fruit size and shape between the parents ensures a clear segregation of fruit size and shape among F2 individuals. The F2 populations and parents were grown at the Kunia substation on Oahu, Hawai'i.

3.2 Phenotype Data Collection and Statistical Analysis

From the Khak dam X 2H94 F2 progeny, phenotype data was collected from 180 F2 trees, six 2H94 parent trees and five Khak dam parent trees. Five fruit were harvested per tree and fruit weight, width and length and flesh color were determined for each.

Fruit shape was determined by calculating the ratio of fruit length to fruit width (Grandillo et al. 1996; 1999). Fruit was collected just as the color break from green to yellow began showing at the blossom end signifying that the fruit had reached its full size. The data results for each trait were averaged, females were removed and histograms plotting these values per tree were constructed to determine whether the control of these traits is quantitative or qualitative. Normal and bimodal distribution plots for each histogram were verified using Microsoft Statistix® program.

Relationships among fruit traits for hermaphrodites were analyzed by running correlation tests for weight and length, weight and width, weight and shape, length and width, length and shape, and width and shape. Correlation tests were performed using Microsoft® Excel.

3.3 Plant Material and DNA Extraction

The most centrally located, semi-translucent young leaf blade and the most recently turned dark green, non-translucent leaf blade were collected from each tree to lessen the chance of sample contamination due to parasites. DNA was isolated from the combined leaf samples collected from each tree for both the AU9 X Sunup and Khak dam X 2H94 populations. Each sample was lyophilized and homogenized into fine powder. Genomic DNA was extracted using a method from Tai and Tanksley (1990) with small modifications. The DNA was treated with RNase, 15 µl of 10mg/ml RNase A per 250 µl DNA for 50 minutes at 37°C, to remove any residual RNA. The RNase-free DNA was purified by first extracting the sample with phenol:chloroform:isoamyl alcohol and followed by a second extraction with chloroform:isoamyl alcohol. DNA was

precipitated from the aqueous phase by adding 2 volumes of ice cold 95% ethanol, mixing, incubating at -80 °C for 30 minutes and centrifuging at 12,000 rpm for 10 minutes. The ethanol was decanted, the nucleic acid pellet washed with 1 mL 70% ethanol and the DNA collected by centrifugation for 5 minutes at 10,000 rpm. The DNA pellet was resuspended in 250 µl distilled and deionized water and run on a 1% electrophoresis agarose gel to determine DNA quality and concentration. Each sample was diluted to concentrations found ideal for SSR PCR amplification (5ng/µl) and AFLP analysis (10ng/µl).

3.4 MARKER DEVELOPMENT

3.4.1 Amplified Fragment Length Polymorphisms

Genomic DNA from the parents AU9 and Sunup and the 54 samples of the F2 progeny were digested with EcoRI and MseI restriction enzymes. EcoRI oligonucleotide adapters along with MseI oligonucleotide adapters were ligated to restriction digested DNA fragments for each template sample. Each template was pre-amplified by PCR with both the EcoRI and MseI oligonucleotides that had an addition of one base pair on the 3'-end of each. Next, selective amplifications were performed, using a standard PCR protocol on the pre-amplified template. IRDye -700 and -800 labeled EcoRI oligonucleotides and unlabeled MseI oligonucleotides both with a 2 to 3 base pair addition at the 3'-end were used. All selective amplifications were run on polyacrylamide gel and fluorescent labeling was viewed on Li-Cor IR2 automated DNA sequencers.

3.4.2 Simple Sequence Repeats

Simple sequence repeats or microsatellite markers were determined using BAC end sequence data obtained from the previously constructed 13.7X papaya BAC library constructed from high-molecular-weight DNA isolated from young papaya leaves (Lai et al, 2006). Other sequence sources mined for SSRs were the papaya genome Shotgun sequences obtained from the Center for Genomics, Proteomics and Bioinformatics Research Initiative (CGPBRI) at the University of Hawai'i (http://cgpbr.mhpcc.hawaii.edu/). Sequences were mined for SSRs using Microsatellite Analysis Server (MICAS), available at http://210.212.212.7/ MIC/index.html. Original criteria for SSRs were those with motif lengths ranging from di- to hexanucleotide. After screening di- to hexanucleotide SSRs for polymorphisms between AU9 and Sunup, similar criteria were used to mine for mono- to tetra-nucleotide repeats for detection of more markers.

3.5 Molecular Marker Scoring

Oligonucleotides were designed with the website Primer3 available at http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi using the following criteria: oligonucleotide length of 18-27, optimal length of 20 nucleotides, oligonucleotide Tm at 57 - 60°C, with an optimal at 60°C, amplified PCR products sized between 100-300 bp and an optimal GC content of 40-60%. Once oligonucleotides were designed to amplify repeated regions, DNA was amplified using standard PCR procedures. DNA was first amplified from AU9 and Sunup templates and then Khak dam and 2H94 templates. Products of these PCR reactions were run on 4% Super-Fine Resolution agarose gels that gave greater sensitivity to small base-pair discrepancies when compared to low-grade

agarose electrophoresis. Gels were run at 270 volts for 20 minutes followed by ethidium bromide staining. By amplifying these parent templates, small polymorphisms between the two parents of each cross were detected. Polymorphic markers found between AU9 and Sunup were sent to Dr. Ray Ming's laboratory at the University of Illinois-Champaign-Urbana where 54 F2 progeny samples were genotyped and used to create a linkage map (Fig. 5). Trees were chosen based on a 2:1 hermaphrodite: female ratio, high DNA quality and complete phenotypic data. Markers were chosen from this map and the oligonucleotide primers specific to each were used to amplify and genotype the DNA isolated from the 180 trees chosen to represent the Khak dam X 2H94 F2 population.

3.6 Map Construction and QTL Analysis

Genomic DNA was isolated from the AU9 X Sunup F2 progeny. Fifty-four samples consisting of 36 hermaphrodites and 18 females were chosen for construction of a linkage map. AFLP was done on all 54 samples and the AU9 and Sunup parents. SSR surveys were done between AU9 and Sunup. The markers mapped in the 54 F2 progeny were used to construct a papaya linkage map using JoinMap® 3.0, Software for the calculation of genetic maps with a LOD score of 3.0. A total of 38 markers found to also be polymorphic between Khak dam and 2H94 were chosen from this linkage map. Other criteria for marker choice was their being approximately equally distributed, no more than 20cM apart when possible to reduce the chance of a double cross-over event to less than 1% and allow more thorough genome coverage (Lander and Botstein, 1989), on each of 15 linkage groups. The chosen markers were used to genotype the 180 Khak dam

X 2H94 F2 trees. Genotype data was collected and formatted for input into JoinMap® 3.0. A linkage map was created using a LOD score of 3.0 from the progeny of the Khak dam X 2H94 cross. This map was constructed and compared to the first linkage map similarly created from the 54 F2 progeny genotypes of the AU9 X Sunup cross. The linkage map of Khak dam X 2H94 was used for fruit size and shape QTL mapping.

Single factor ANOVA tests were used to identify markers linked for fruit weight, length, width and shape from the 38 genotyped markers, including the 21 linked markers on the fruit size map and the remaining unlinked markers. Phenotypes for individual trees were grouped by genotype and marker genotypes were treated as classification variables for ANOVA tests. that were performed using Microsoft® Excel.

Phenotype data, genotype data and the Khak dam X 2H94 linkage map were then loaded into a MapQTL® 5.0 program. Putative QTLs were first identified using interval mapping as described in the reference manual. Permutation tests were done to develop LOD thresholds for each to verify the presence and location of QTLs phenotypic trait in question according to linkage group (van Ooijen, 1999; 2004). Permutation tests were set at 1000 permutations with a p-value of 0.02 resulting in 98% accuracy for LOD thresholds. LOD thresholds were determined for weight, length, width and shape on each linkage group and genome wide. The genome wide LOD threshold for weight and length was 3.2. For width, the genome wide LOD threshold was 3.5 and for shape, the genome wide LOD threshold was 3.1. Multiple-QTL model (MQM) mapping methods were done as described in the reference manual (van Ooijen, 2004). The closest markers before and after each putative QTL were selected as cofactors. The markers were used as

background controls in the approximate MQM. Automatic cofactor selection was also done to confirm that markers showed significant association. The estimated additive genetic effect and percentage of variance explained by each QTL were obtained using MQM mapping. The total variance explained by all the QTLs affecting a trait was obtained from the results of the automatic cofactor selection.

CHAPTER 4:

RESULTS

4.1 Polymorphism Detection

A total of 28 selective amplifications of AFLP markers were setup up from 56 primer pairs for the AU9 and Sunup parents and all 54 F2 progeny templates. Thirty-six oligonucleotide pairs were screened for AFLP markers. Each individual was genotyped at a total of 101 marker loci for use in saturation mapping.

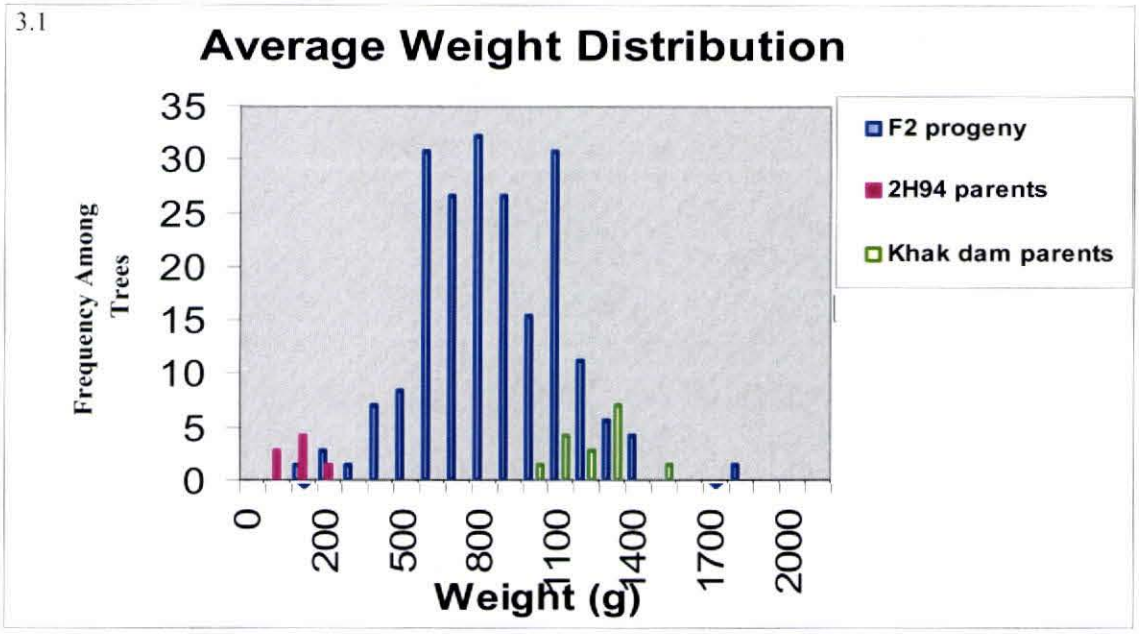
After SSRs were mined from papaya sequencing data, parents of both crosses were surveyed for polymorphisms using oligonucleotides designed to amplify the repeat regions. To develop the papaya linkage map, PCRs were setup to survey for polymorphisms between AU9 and Sunup. The template from both parents along with two hermaphrodite F2 progeny samples, two female F2 progeny samples and a template mixture of both AU9 and Sunup were amplified with each oligonucleotide pair. For the AU9 X Sunup map, 795 oligonucleotide pairs were tested for SSRs obtained through Shotgun sequencing. Of these, 230 were found polymorphic between Sunup and AU9 at a polymorphism rate of 29%. Other SSR polymorphic markers were obtained from BES and had been previously screened.

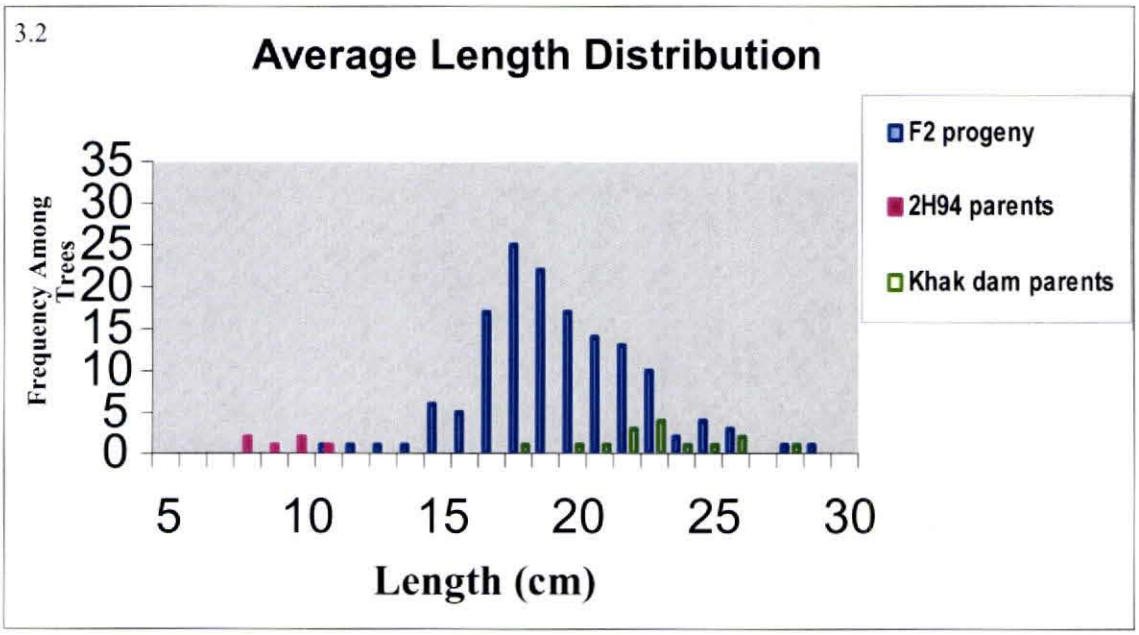
For marker-based analysis of fruit size, PCRs were setup to survey for polymorphisms between Khak dam and 2H94. Criteria were similar to that used for screening the AU9 X Sunup cross besides using a template mixture of both Khak dam and 2H94. Out of 699 oligonucleotide pairs tested, 111 were polymorphic for both AU9 X Sunup and Khak dam X 2H94 at a rate of 16%.

4.2 Phenotypic Data

Five fruit were harvested from each of six 2H94 trees, four Khak dam trees and 180 F2 progeny trees. An average fruit weight, length and width were determined for each hermaphrodite tree, totaling 137 out of the 180 original F2s, and plotted against the frequency of occurrence among all hermaphrodite trees. Fruit shape was calculated similar to that done in tomato (Grandillo et al. 1996) by figuring the ratio of length to width (Fig. 3). Linear correlation coefficients were computed among all traits (Fig. 4).

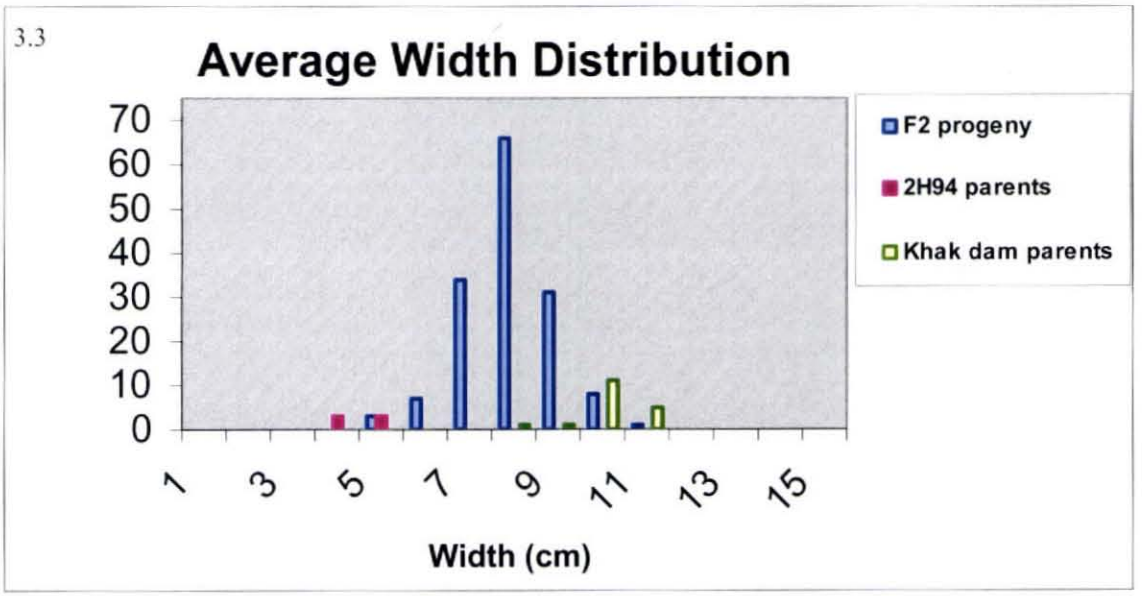
For evaluating segregation of flesh color, two different Khak dam X 2H94 F1 fruit were selfed to make up the entire Khak dam X 2H94 F2 population. KHR was the part of the population created by selfing a Khak dam X 2H94 F1 tree expressing red flesh colored fruit. Since red flesh is recessive, each tree of the F2 progeny produced red-fleshed fruit. An F2 KHY cross was also done by selfing a Khak dam X 2H94 F1 tree expressing yellow flesh colored fruit. Since yellow flesh is heterozygous dominant, flesh color segregated in the F2 progeny. Flesh color was determined for 100 trees from the KHY sub-cross, 74 trees expressed yellow flesh color while 26 trees expressed red flesh color. This shows a standard Mendelian inheritance ratio of 3:1, with yellow flesh color being dominant signifying a monogenically controlled trait.



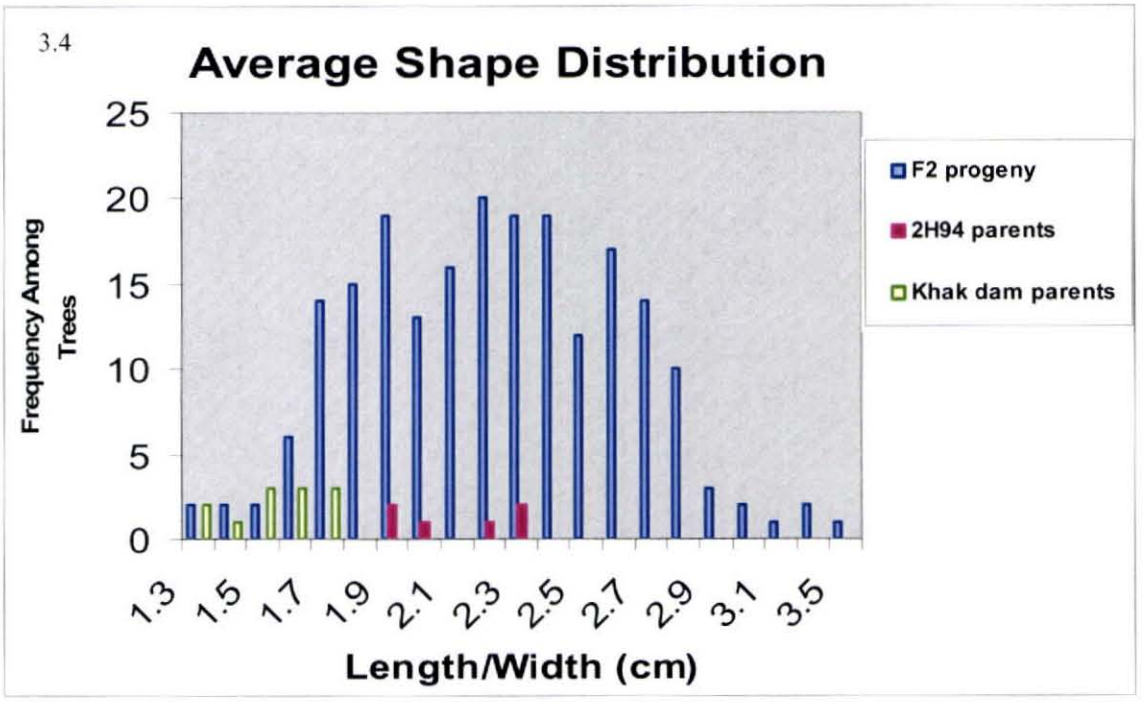


Figures 3.1 and 3.2: Variation and frequency of fruit weight and length among hermaphrodite trees.

3.3



3.4



Figures 3.3 and 3.4: Variation and frequency of fruit width and shape among hermaphrodite trees.

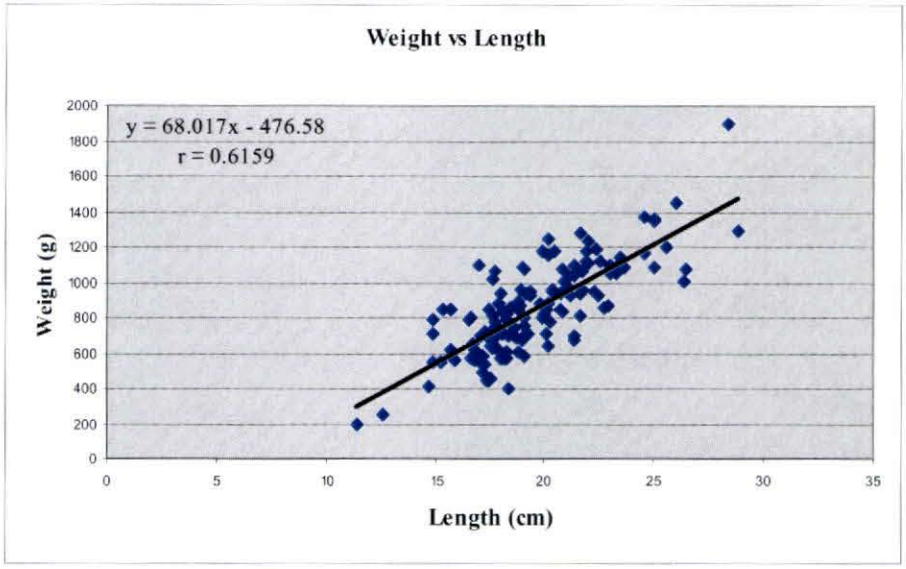


Figure 4.1: Correlation of fruit weight and length per hermaphrodite tree, where r = linear correlation coefficient.

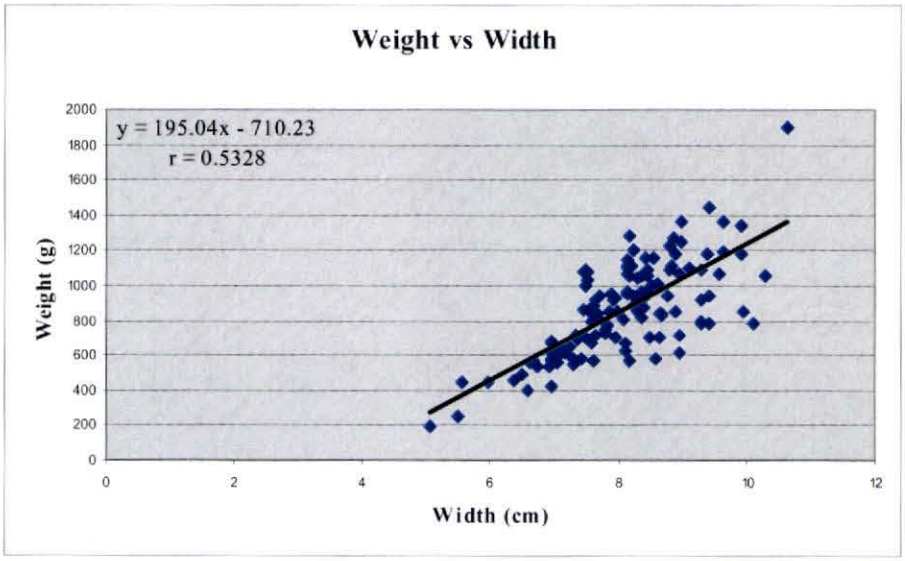


Figure 4.2: Correlation of fruit width and weight per hermaphrodite tree, where r = linear correlation coefficient.

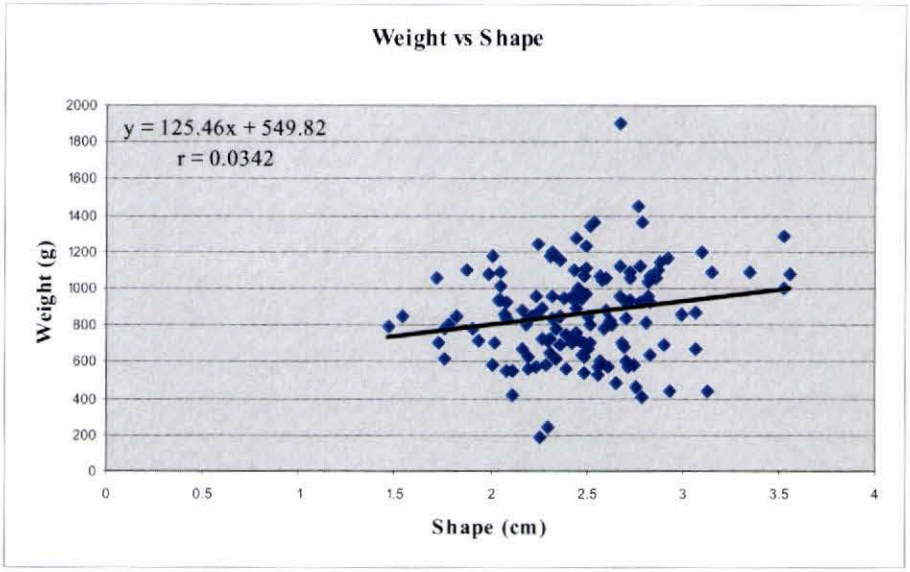


Figure 4.3: Correlation of fruit shape and weight per hermaphrodite tree, where r = linear correlation coefficient.

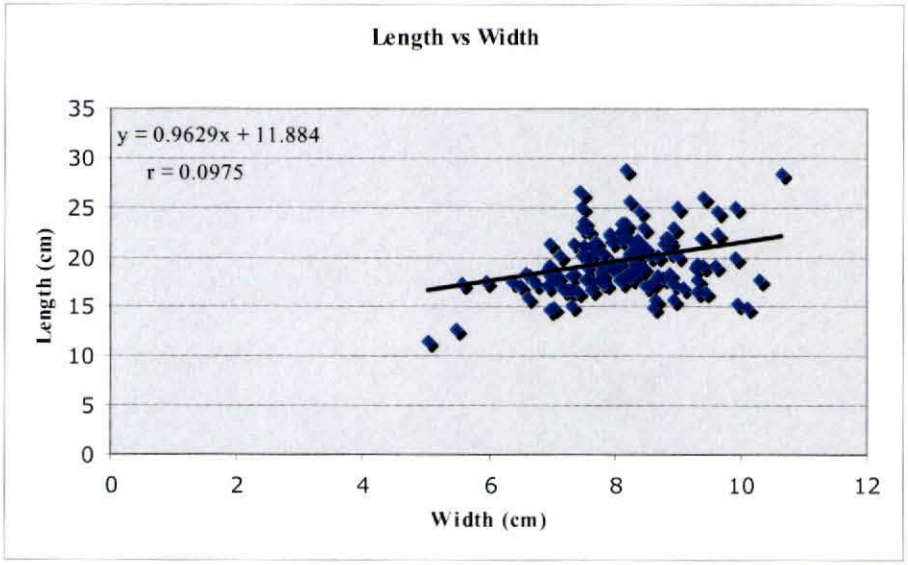


Figure 4.4: Correlation of fruit length and width per hermaphrodite tree, where r = linear correlation coefficient.

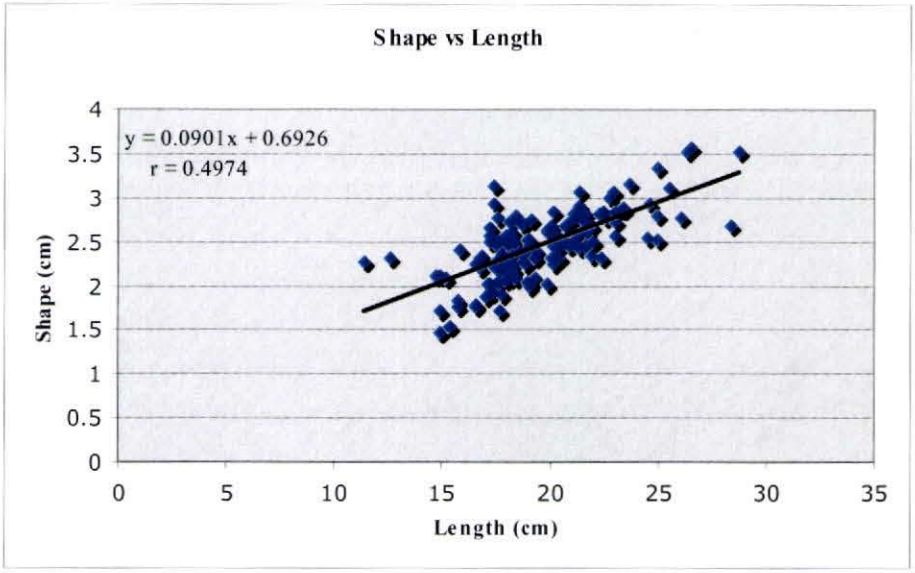


Figure 4.5: Correlation of fruit length and shape per hermaphrodite tree, where r = linear correlation coefficient.

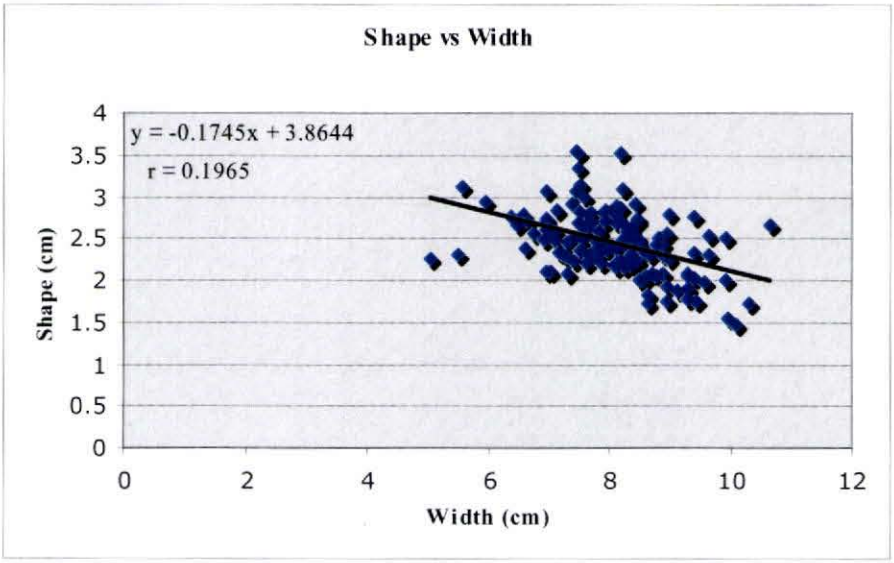


Figure 4.6: Correlation of fruit width and shape per hermaphrodite tree, where r = linear correlation coefficient.

4.3 Linkage Maps

The AU9 X Sunup linkage map (Fig. 5) was constructed by genotyping 54 F2 progeny using 199 SSR markers. The segregation data for inheritance patterns of 167 of these SSRs markers that included one MSY specific marker, P3K124Y, were mapped

into 16 linkage groups. Nine of the 16 linkage groups had seven or more markers while the seven smaller linkage groups had from one to three markers. Marker names beginning with "CPM" were mined from BES data, all other markers were mined from Shotgun sequencing data and marker names followed by "L" were generated from Li-Cor scanners. The MSY-specific marker is denoted with a "Y" following the marker name. This map was constructed to provide a tentative location for the markers so that evenly distributed markers could be chosen and Khak dam X 2H94 individuals could be genotyped at each marker locus.

After the AU9 X Sunup linkage map was created, 38 SSR markers were chosen based on afore mentioned criteria and each individual in the F2 progeny of the Khak dam X 2H94 population were genotyped at each marker locus. The results were input into JoinMap® 3.0 to construct a linkage map (Fig. 6) to be used in fruit size and shape QTL detection.

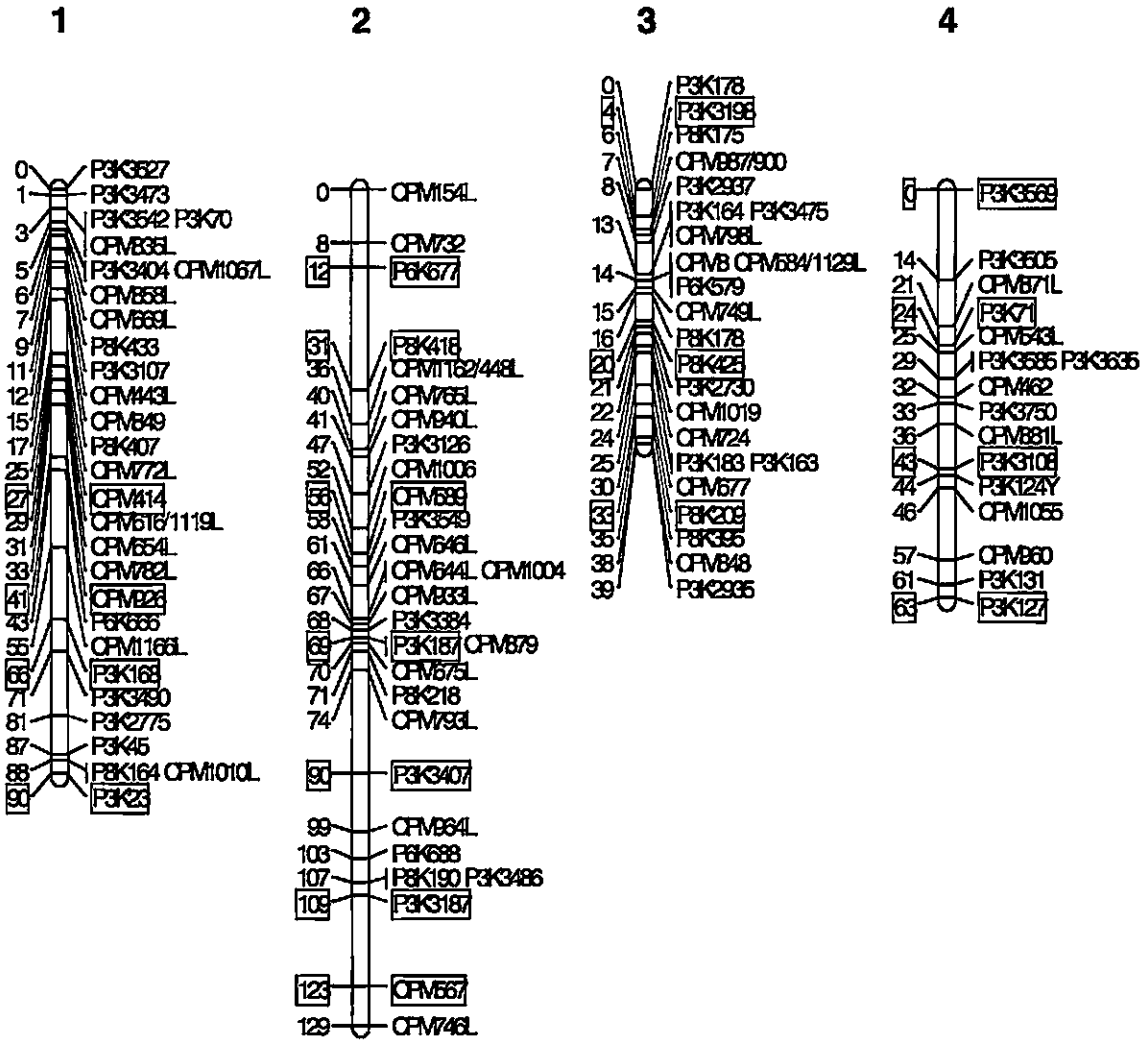


Figure 5: AU9 X Sunup *Carica papaya* linkage map showing the location of 167 SSR markers and the MSY specific marker. Li-Cor generated markers are indicated by an "L" following the marker name. BAC end sequence derived markers are indicated by "CPM". The MSY-specific marker is indicated by a "Y" following the marker name. All other markers were generated from Shotgun sequence data. The numbers to the left indicate genetic distance in cM. Markers with borders represent those chosen for fruit size QTL screening.

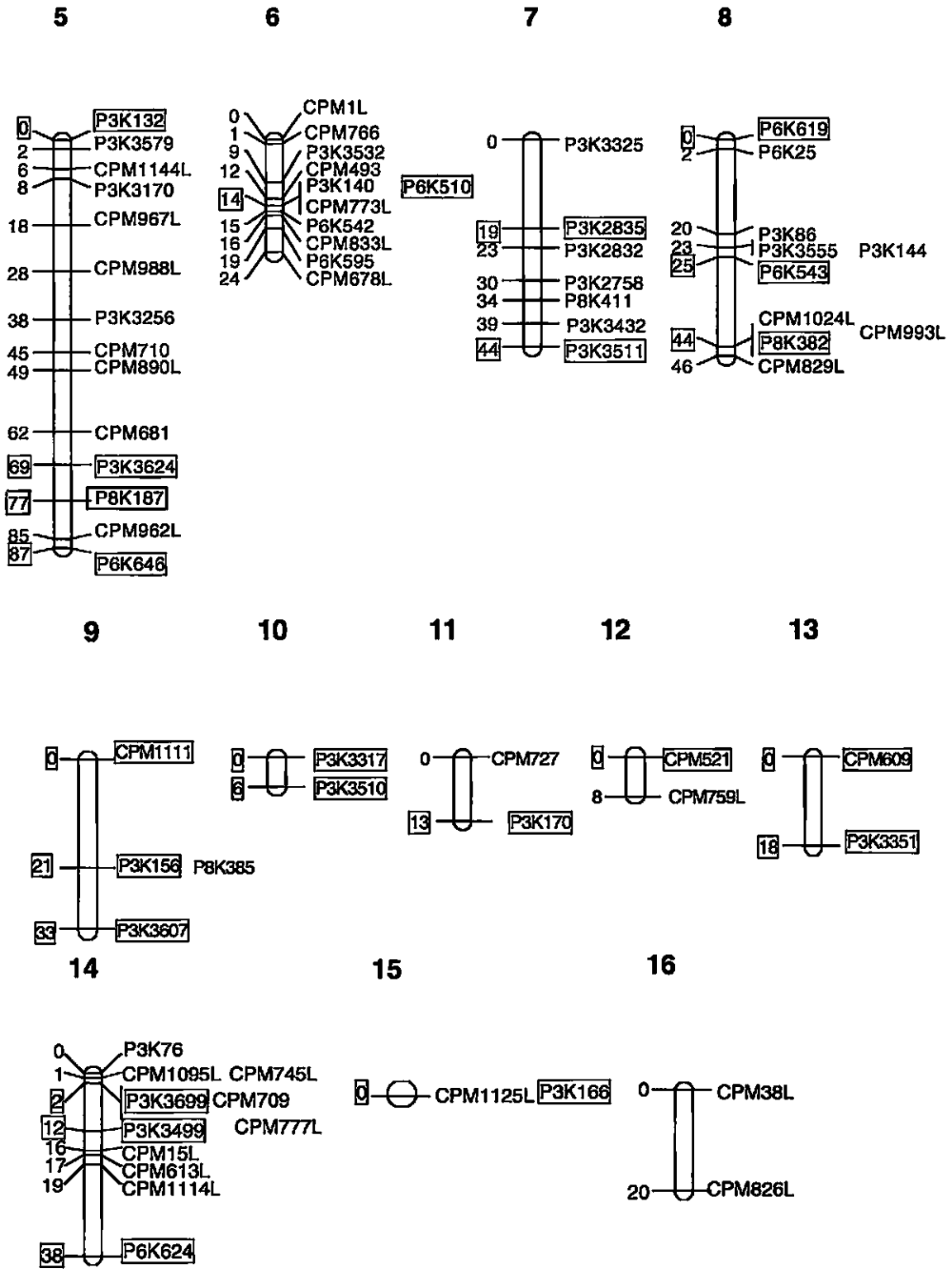


Figure 5: (continued)

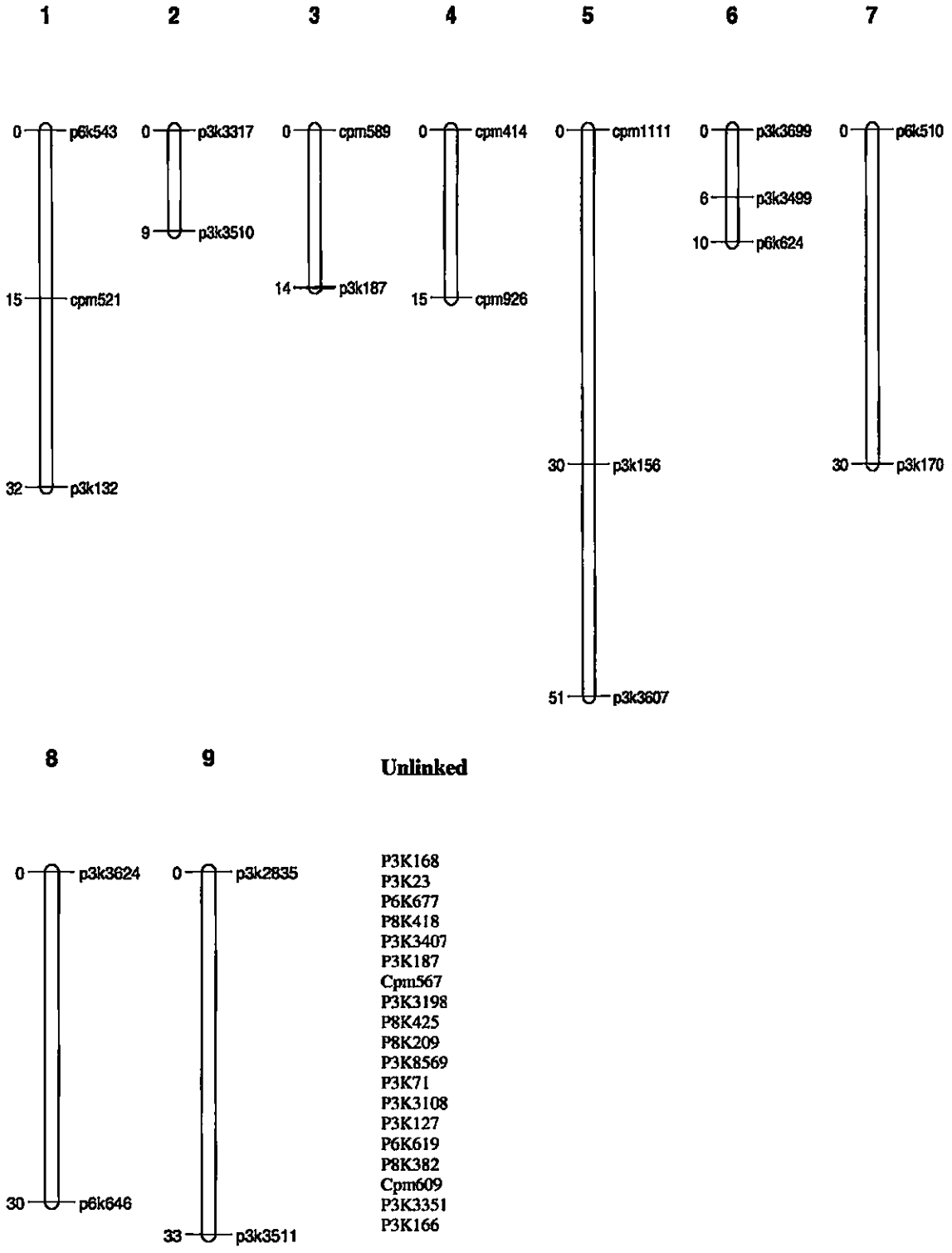


Figure 6: Khak dam X 2H94 linkage map constructed using the 38 SSR markers chosen for genotyping from the AU9 X Sunup linkage map. This map shows locations of 21 SSR marker based on the Khak dam X 2H94 cross. The numbers to the left indicate genetic distance in cM.

4.4 MAJOR GENE AND QTL MAPPING OF PHENOTYPIC TRAITS

4.4.1 Flesh Color

Genotyping results for the Khak dam X 2H94 progeny at marker P3K132 indicate a single gene segregating when comparing both the KHY and KHR F2 flesh color sects of the Khak dam X 2H94 cross (Fig. 7). However, flesh color did not correlate entirely to genotype results for individuals at marker P3K132. Some yellow-fleshed fruit have a genotype matching to Khak dam which produces red flesh. Other markers that surround P3K132 on the AU9 X Sunup linkage map could not be scored for the flesh color genotype in the Khak dam X 2H94 F2 progeny because none were polymorphic between the parents. Flesh color was found to correlate with individual genotype at 0cM on linkage group 5 in the AU9 X Sunup progeny population. This location correlates with marker P3K132. Results were verified at Dr. Ray Ming's laboratory at the University of Illinois-Champagne-Urbana.

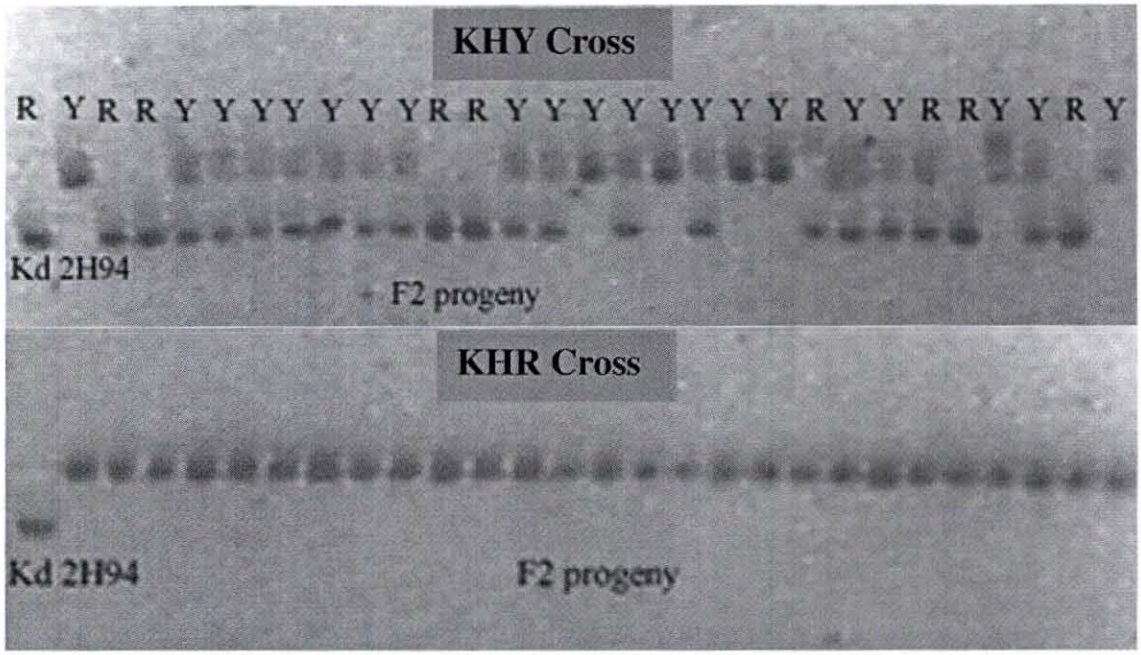


Figure 7: Sample of P3K132 gel image showing single gene segregation among Khak dam X 2H94 F2 progeny. Kd=Khak dam parent, Y = yellow flesh color, R = red flesh color.

4.4.2 Fruit Weight

Fruit weight distribution among the hermaphrodite progeny shows that this trait segregates normally indicating quantitative trait loci involvement. Single factor ANOVA tests showed significant correlation between fruit weight and markers Cpm609, P3K3499 and P3K3699 signifying these markers as locations for suggestive linkage loci (Tables1.1-1.3).

Interval mapping permutation tests (1000 permutations at p-value = 0.02) gave a genome-wide LOD significance threshold of 3.1. Significant LOD scores determined using interval and MQM mapping showed significant QTL on linkage group 6 (Fig. 8 and Table 5).

Table 1.1: ANOVA results for suggestive QTL at marker Cpm609

Genotype	*Count*	*Sum*	*Avg. Phenotype (g)*	*Variance*	*P-value*
Khak dam	22	22728.16	1033.098182	93705.90548	0.00124317
Both Parents	77	64817.32	841.7833766	49832.01516	
2H94	36	28335.395	787.0943056	70110.05936	
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	875730.6432	2	437865.3216	7.040913808	3.064760677
Within Groups	8208909.245	132	62188.7064		
Total	9084639.888	134			

Table 1.2: ANOVA results for suggestive QTL at marker P3K3699

Genotype	*Count*	*Sum*	*Avg. Phenotype (g)*	*Variance*	*P-value*
Khak dam	26	25234.82	970.57	63247.54589	0.000175683
Both Parents	78	68535.955	878.6660897	60713.86477	
2H94	33	23291.2	705.7939394	59833.94779	
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	1125575.461	2	562787.7304	9.229593556	3.063714933
Within Groups	8170842.564	134	60976.43704		
Total	9296418.025	136			

Table 1.3: ANOVA results for suggestive QTL at marker P3K3499

Genotype	*Count*	*Sum*	*Avg. Phenotype (g)*	*Variance*	*P-value*
Khak dam	24	24078.52	1003.271667	46235.94263	0.000179443
Both Parents	78	67583.995	866.4614744	64394.29295	
2H94	35	25399.46	725.6988571	63283.47134	
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	1122992.762	2	561496.3808	9.205505967	3.063714933
Within Groups	8173425.263	134	60995.71092		
Total	9296418.025	136			

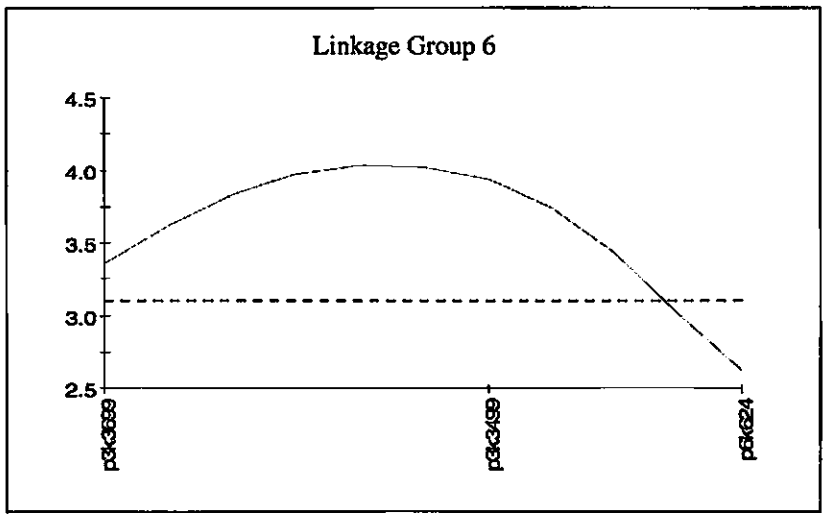


Figure 8: QTL associated with fruit weight in 137 Khak dam X 2H94 hermaphrodite F2 progeny. Horizontal axes indicate marker location on linkage group. P3K3699=0 cM, P3K3499=6 cM, P6K624=10cM. The dotted line indicates genome wide significance threshold for significant QTL. Linkage group 6 contained loci that significantly affect fruit weight.

4.4.3 Fruit Length

Fruit length distribution among the hermaphrodite progeny shows that it segregates normally indicating quantitative trait loci involvement in expression. Single factor ANOVA tests showed significant correlation between fruit length and markers P3K3317, P3K3510 and P6K510 signifying these markers as locations for suggestive linkage loci. Marker P3K170 had no significant ANOVA results (Tables 2.1-2.4).

Interval mapping permutation tests showed a genome-wide LOD significance threshold of 3.1. Significant LOD scores determined using interval and MQM mapping showed significant QTL on linkage group 7 (Fig. 9 and Table 5).

Table 2.1: ANOVA results for suggestive QTL at marker P3K3317

Genotype	*Count*	*Sum*	*Avg Phenotype(cm)*	*Variance*	*P-value*
Khak dam	36	711.79	19.77194444	8.161341825	0.019929688
Both Parents	61	1149.535	18.84483607	6.740653306	
2H94	38	782.26	20.58578947	13.51152233	
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	72.73543477	2	36.36771738	4.034023795	3.064760677
Within Groups	1190.012489	132	9.015246125		
Total	1262.747923	134			

Table 2.2: ANOVA results for suggestive QTL at marker P3K3510

Genotype	*Count*	*Sum*	*Avg Phenotype(cm)*	*Variance*	*P-value*
Khak dam	39	774.13	19.84948718	7.67309973	0.041542836
Both Parents	64	1210.915	18.92054688	7.929450887	
2H94	33	676.76	20.50787879	12.94964848	
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	59.06768064	2	29.53384032	3.258340316	3.064233814
Within Groups	1205.521947	133	9.06407479		
Total	1264.589628	135			

Table 2.3: ANOVA results for suggestive QTL at marker P6K510

Genotype	*Count*	*Sum*	*Avg Phenotype(cm)*	*Variance*	*P-value*
Khak dam	27	564.545	20.90907407	7.562103917	7.48361E-07
Both Parents	68	1375.54	20.22852941	8.730702283	
2H94	39	682.96	17.51179487	6.056604588	
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	243.1119957	2	121.5559978	15.73932784	3.065295706
Within Groups	1011.722729	131	7.723074268		
Total	1254.834725	133			

Table 2.4: ANOVA results for marker P3K170

Genotype	*Count*	*Sum*	*Avg Phenotype(cm)*	*Variance*	*P-value*
Khak dam	28	540.14	19.29071429	6.118710582	0.685714329
Both Parents	77	1512.535	19.64331169	8.512933954	
2H94	29	579.48	19.98206897	13.08883842	
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	6.809018539	2	3.404509269	0.378382906	3.065295706
Within Groups	1178.675642	131	8.997523985		
Total	1185.484661	133			

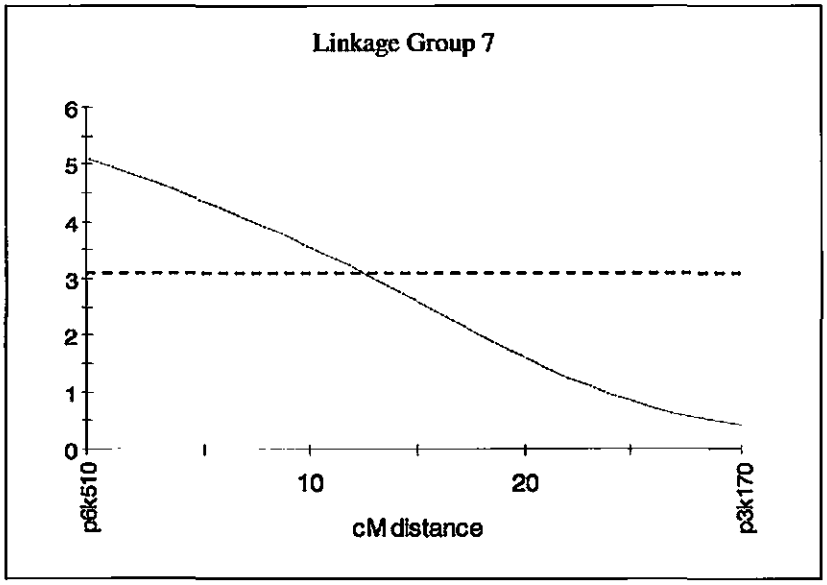


Figure 9: QTL associated with fruit length in 137 Khak dam X 2H94 hermaphrodite F2 progeny. Horizontal axes indicate marker location on linkage group and map distance in cM, the dotted line indicates genome wide significance threshold for significant QTL. Linkage group 7 contained loci that significantly affect fruit length.

4.4.4 Fruit Width

Fruit width distribution among the hermaphrodite progeny shows that it segregates normally indicating quantitative trait loci involvement. Single factor ANOVA tests showed significant correlation between fruit width and markers Cpm609, P3K3699, P3K3499 and P6K624 signifying these markers as locations for suggestive linkage loci (Tables 3.1-3.4).

Permutation tests showed a genome-wide significance threshold of 3.0. Significant LOD scores determined using interval and MQM mapping showed significant QTL on linkage group 6 (Fig. 10 and Table 5).

Table 3.1: ANOVA results for suggestive QTL at marker Cpm609

Genotype	*Count*	*Sum*	*Avg Phenotype(cm)*	*Variance*	*P-value*
Khak dam	22	186.69	8.485909091	1.139165801	0.01380156
Both Parents	77	621.9	8.076623377	0.792730554	
2H94	36	277.33	7.703611111	1.238435159	
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	8.549312585	2	4.274656293	4.42499779	3.064760677
Within Groups	127.5152345	132	0.966024503		
Total	136.064547	134			

Table 3.2: ANOVA results for suggestive QTL at marker P3K3699

Genotype	*Count*	*Sum*	*Avg Phenotype(cm)*	*Variance*	*P-value*
Khak dam	26	214.15	8.236538462	0.727973538	0.005133429
Both Parents	78	637.25	8.169871795	0.934422061	
2H94	33	248.92	7.543030303	1.17582803	
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	10.46039395	2	5.230196977	5.484946798	3.063714933
Within Groups	127.7763341	134	0.953554732		
Total	138.2367281	136			

Table 3.3: ANOVA results for suggestive QTL at marker P3K3499

Genotype	*Count*	*Sum*	*Avg Phenotype(cm)*	*Variance*	*P-value*
Khak dam	24	204.53	8.522083333	0.70326721	0.000562154
Both Parents	78	631.89	8.101153846	0.875096054	
2H94	35	263.9	7.54	1.178517647	
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	14.60958612	2	7.304793058	7.917697149	3.063714933
Within Groups	123.627142	134	0.922590612		
Total	138.2367281	136			

Table 3.4: ANOVA results for suggestive QTL at marker P6K624

Genotype	*Count*	*Sum*	*Avg Phenotype(cm)*	*Variance*	*P-value*
Khak dam	29	241.055	8.312241379	0.988752833	0.000609877
Both Parents	77	628.745	8.165519481	0.747649727	
2H94	30	222.88	7.429333333	1.345868506	
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	14.54565774	2	7.272828872	7.829953925	3.064233814
Within Groups	123.5366452	133	0.928846956		
Total	138.0823029	135			

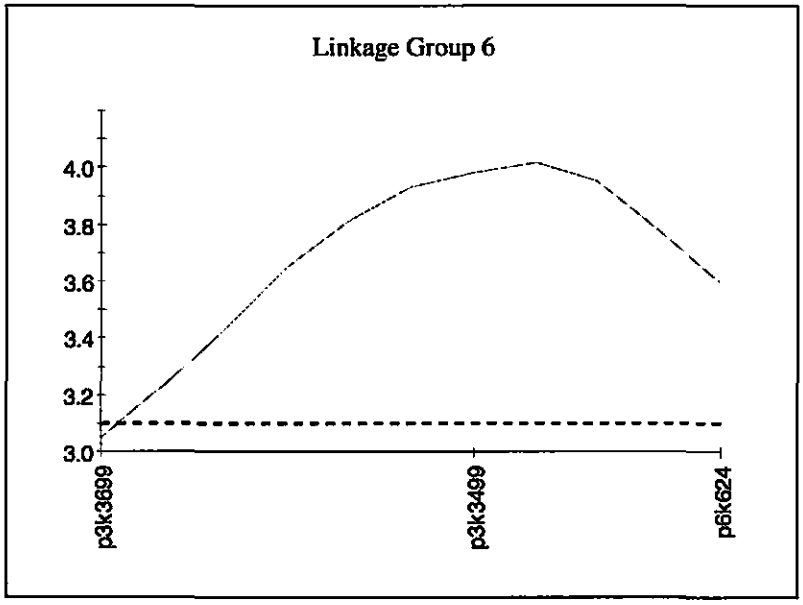


Figure 10: QTL associated with fruit width in 137 Khak dam X 2H94 hermaphrodite F2 progeny. Horizontal axes indicate marker location on linkage group. P3K3699=0 cM, P3K3499=6 cM, P6K624=10cM. The dotted line indicates genome wide significance threshold for significant QTL. Linkage group 6 contained loci that significantly affect fruit width.

4.4.5 Fruit Shape

Fruit shape distribution among the hermaphrodite progeny shows that it segregates normally indicating quantitative trait loci involvement. Single factor ANOVA tests showed significant correlation between fruit shape and marker P6K510 signifying this marker as a suggestive linkage locus. Marker P3K170 had no significant ANOVA results (Tables 4.1-4.2).

Interval mapping permutation tests showed a genome-wide significance threshold of 3.1. Significant LOD scores determined using interval and MQM mapping showed significant QTL on linkage group 7 (Fig. 11 and Table 5).

Table 4.1: ANOVA results for suggestive QTL at marker P6K510

Genotype	*Count*	*Sum*	*Avg Phenotype(cm)*	*Variance*	*P-value*
Khak dam	27	70.96343611	2.628275411	0.136225423	3.82392E-06
Both Parents	69	174.0810909	2.522914361	0.14299048	
2H94	39	86.45387762	2.216766093	0.073689262	
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	3.342331286	2	1.671165643	13.73098635	3.064760677
Within Groups	16.0654056	132	0.121707618		
Total	19.40773688	134			

Table 4.2: ANOVA results for marker P3K170

Genotype	*Count*	*Sum*	*Avg Phenotype(cm)*	*Variance*	*P-value*
Khak dam	28	66.98391633	2.392282726	0.111062033	0.373040287
Both Parents	77	189.4931069	2.460949441	0.150019905	
2H94	29	73.50091218	2.534514213	0.165301612	
Source of Variation	*SS*	*df*	*MS*	*F*	*F crit*
Between Groups	0.288633457	2	0.144316728	0.993528624	3.065295706
Within Groups	19.02863284	131	0.145256739		
Total	19.3172663	133			

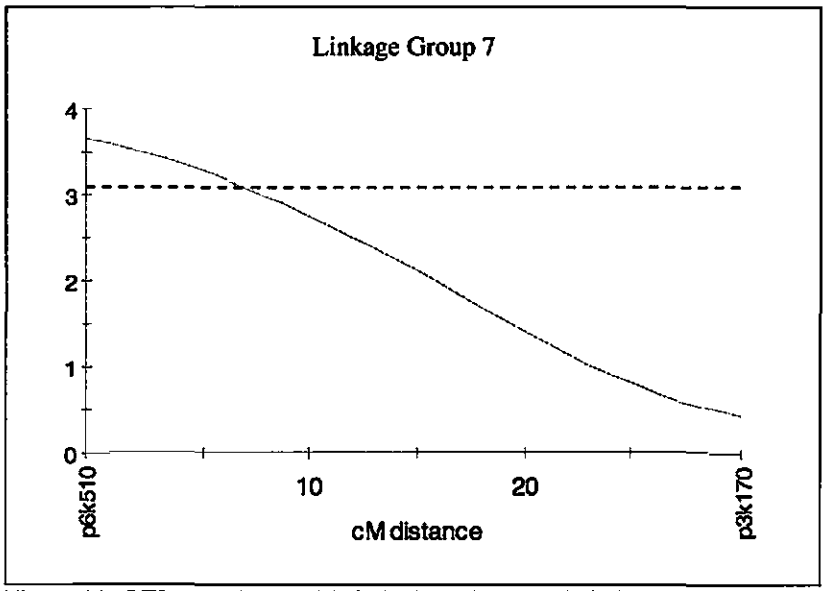


Figure 11: QTL associated with fruit shape in 137 Khak dam X 2H94 hermaphrodite F2 progeny. Horizontal axes indicate marker location on linkage group and map distance in cM, the dotted line indicates genome wide significance threshold for significant QTL. Linkage group 7 contained loci that significantly affect fruit shape.

Table 5: Significant QTLs for fruit size and shape in Khak dam X 2H94 F2 progeny trees.

Trait	Significant LOD threshold	Linkage Group	Cofactor marker	Peak Position	LOD Score	% Explained Genetic Variance	Additive Effect
Weight	3.1	6	P3K3499 @ 6cM	4 cM	4.03	10.80	60504.50
Length	3.1	7	P6K510 @ 0cM	0 cM	5.13	1.3	9.14
Width	3.0	6	P3K3499 @ 0cM	7 cM	4.02	11.6	0.89
Shape	3.1	7	P6K510 @ 0cM	0 cM	3.66	2.5	0.14

CHAPTER 5:

DISCUSSION

It is well established that papaya varieties differ greatly in fruit weight, length, width and shape (Kim et al. 2001). Papaya varieties also differ in fruit flesh color. Specific fruit traits are more desirable in certain regions of the world. For example, the smaller 'Solo' type yellow flesh papayas are preferred by the markets served by Hawai'i whereas larger, red flesh papayas are more marketable in regions such as Mexico (Purseglove, 1968). Therefore, developing papaya cultivars suited to a specific market will greatly increase its value. However, the genetic determinants that regulate fruit weight, length, width and shape are not known. This study represents the first genome-wide scan undertaken to identify the genetic loci affecting variation in these traits. The data showed that multiple genetic loci affect the variation in fruit weight, length, width and shape and that a major gene affects flesh color.

Flesh color was expected to be a major gene trait and yellow flesh was expected to be dominant. Phenotypic flesh color observations suggest that, in the KHY cross of the Khak dam X 2H94 population segregating for the trait, flesh color is conditioned by a major gene and that yellow flesh is dominant. However, the incomplete correlation of flesh color and individual genotype scores at marker P3K132 make the exact location of this major gene questionable. Still, since a major gene appears to be segregating at this marker locus and some individual genotypes correlate accurately to flesh color it is likely that the gene involved in flesh color should be mapped to the same general region on the linkage group (Lewis 2001). This theory is also likely due to the discovery that the gene

for flesh color was mapped to the same region as marker P3K132 on the AU9 X Sunup map. Since none of the markers reported in this study show complete linkage to the flesh color gene in the Khak dam X 2H94 population, it would be necessary to score more markers to tag the location more closely and enable marker-assisted selection in the future.

Fruit shape is specific to sex-type character and ranges from round, short fruit produced by female flowers to long, cylindrical fruit produced by hermaphrodite flowers (Chan and Paull, 2007). Therefore, data from female trees were removed for assessing normal distribution to avoid inaccurate results for normal or skewed distribution. Females were also removed from QTL analysis to avoid a false positive conclusion that there is a segregating QTL when, in fact, there is not.

Quantitative loci affecting fruit weight mapped to linkage group 6. It seemed possible that one gene would have a major affect on this trait as found for tomatoes (Tanksley et al. 1990). However, histograms showed normal distribution signifying that fruit weight is a quantitative trait. Still, there is a possibility that fruit weight may be affected by a major gene that is modified by a number of other genes. None of the mapped markers for fruit weight found in this study are strong candidates for being linked to a major gene. Markers P3K3699 and P3K3499 have similar LOD scores in interval mapping and similar F-values in ANOVA tests. ANOVA tests also identified marker Cpm609 as a putative QTL, since this marker is unlinked it could not be verified with interval mapping, but it should still be considered as a possibility for a significant QTL location.

Quantitative loci affecting fruit length were found on linkage groups 2 and 7. ANOVA tests show significant correlation for markers P3K3317 and P3K3510 on linkage group 2 and fruit length. However, these markers did not produce significant LOD scores after interval mapping and were therefore discarded as significant QTL locations for fruit length. Marker P6K510 was found to be highly significantly correlated to fruit length following ANOVA tests with a F-value of 15.7 and F-criteria of 3.0 Interval mapping also produced a very high LOD score of 5.1 for marker P6K510 signifying this as a significant QTL for fruit length.

Quantitative loci affecting fruit width were located on linkage group 6. Markers P3K3499 and P6K624 scored higher LOD values than the other marker, P3K3699, on the linkage group. These markers also gave more significant F-values after ANOVA analysis than markers Cpm609 and P3K3699, meaning the correlation of fruit width and individual genotype at these marker loci is greater than the that of the other markers found to significantly affect fruit width. Marker Cpm609 was unlinked on the map and therefore could not be ruled out as a as potential QTL for lack of an interval mapping LOD score.

Fruit shape was calculated and considered a separate trait because it is unknown whether or not the growth rates of width and length for the developing fruit, after flower anthesis, are equal. When these growth rates are unequal, the ultimate configuration of the mature fruit can be produced by the interaction between genetic factors ruling relative dimensional growth and any genetic change that affects final fruit weight (Kaiser, 1935; Sinnott, 1936). Marker P6K510, located on linkage group 7, gave a significant ANOVA

test statistic and significant LOD score making it a good candidate for a QTL involved in fruit shape. No other markers used in this study showed significant F-values following ANOVA tests or significant LOD scores following interval mapping.

This study showed similar QTL for width as well as weight. It is interesting that both traits showed similar significant QTL, meaning that the two traits may be correlated, since fruit weight was originally assumed to be controlled by a single gene. There was positive phenotypic correlation between weight and width, $r = 0.5328$. However, weight and length showed stronger positive correlation with $r = 0.6159$. Fruit weight varied more as a function of length (100 g to each 1cm length) than of width (100 g to each 0.43 cm width). Weight and shape had close to no correlation, $r = 0.0342$, so weight was not a function of fruit shape.

Fruit length and shape had similar significant QTL especially at marker P6K510. This could simply imply that the genes responsible for fruit length contribute to calculated shape. There was positive phenotypic correlation between length and shape, $r = 0.4974$, but a negative correlation between width and shape, $r = -0.1965$. Fruit shape is a derived or calculated trait that is computed by taking the ratio of length to width. Therefore, negative correlation with width means that as width increases length does not increase at the same rate and the resulting shape decreases. Fruit length and width have a positive linear correlation meaning that generally the greater the fruit length, the greater the fruit width. However, the r-value is very small meaning that the growth rates of width and length for the developing fruit, after flower anthesis, are not exactly equal. This relationship may indicate that the ultimate configuration of the mature fruit is

produced by an interaction between other genetic factors ruling dimensional growth.

The linkage map constructed for papaya is not complete. It was proposed by Dr. Ray Ming and Dr. Robert Paull that at least 1000 markers are required to construct a reliable linkage map as was done in tomato (Tanksley et al. 1992). A linkage map, like the ones constructed in this study, consisting of 50-300 markers that are spaced >5 cM apart is considered a low or moderate density map. While these maps are still very useful, limitations arise since the probability that the entire genome has been covered is low (Lander and Botstein, 1989). For example, the AU9 X Sunup map constructed in this study produced 16 linkage groups instead of the nine expected to equal the number of chromosomes in papaya. This is most likely due to using only 167 markers. To construct the Khak dam X 2H94 map, evenly distributed markers from the AU9 X Sunup map were sometimes unavailable to score for individual genotypes. This caused the resulting map to be somewhat deficient for QTL analysis because the entire genome could not be screened. This map was still used for QTL detection because of the great phenotypic diversity of the traits observed in this study that is present between the parents. High density linkage maps provide a greater probability that the entire genome is completely covered with molecular markers and that there are no large, marker-less gaps in the map. Avoiding marker-less gaps is especially important when using molecular linkage maps to detect and characterize loci underlying quantitative traits where assurance that the entire genome has been uniformly surveyed is necessary (Paterson et al. 1988; Lander and Botstein, 1989). Once high density linkage maps are constructed from both the AU9 X Sunup and Khak dam X 2H94 populations, additional polymorphic markers will be

available for scoring. Subsequently, the exact location of the flesh color gene will likely be determined as well as more reliable QTL locations for fruit weight, length, width and shape.

A more detailed linkage map and another round of phenotypic data collection from a new Khak dam X 2H94 cross could potentially decrease the significance of QTLs reported in this study. The phenotype expression of these traits can be altered by developmental factors such as more than one fruit born per node and varied fruit denseness on each tree. In order to avoid this problem, fruit was harvested only if it was grown singularly from a node. The denseness of the fruit on each tree should also be measured and included in the phenotype expression studies to observe any correlations that may arise in fruit size and shape and fruit denseness per tree. For the genetic analysis of natural variation, phenotype data collection should be done numerous times on the same population with each new set of fruit born per tree to allow for localization of traits of interest that show allelic variation in the same mapping population (El-Lithy et al. 2005). With each cycle of fruiting, the phenotype data collected from the same population may vary due to environmental factors such as weather conditions that will change according to year and season. For example, papayas grow well if there is a minimum monthly precipitation of ~100mm. Drought often leads to poor fruit set whereas flooding often leads to plant death due to root rot. Papaya is also very sensitive to humidity and temperature (Chan and Paull, 2007). It is also difficult to determine how much of the genetic variation in the Khak dam X 2H94 population is explained by the QTLs since only the phenotypic variance was observed. In order to ensure accurate and

reliable results, crosses must be made and grown on numerous plot sites in various locations so that it can be determined whether the variation left unexplained is due to other QTLs or the environment (Kearsey and Farquhar, 1998).

Since genotypes differ in their genetic composition, it would also be interesting to analyze the same traits of interest in different populations in order to gain insight into the genetic variation of fruit weight, length, width and shape and flesh color within the entire *C. papaya* species. In order for proper comparison to take place the same marker framework should be used and markers should be anchored to the entire papaya linkage map once it is finally constructed. Creating crosses using the mutant 2H94 as a common parent and varying the other parent would allow direct comparison of the loci segregating in the populations tested as was done in Arabidopsis (El-Lithy et al. 2005).

In conclusion, this study provides evidence that fruit weight, length, width and shape are genetically regulated and that genes influencing these traits are located on linkage groups 6 and 7 in the preliminary Khak dam X 2H94 linkage map. It is likely that fruit weight and width are both influenced by markers Cpm609, which is currently unlinked in the Khak dam X 2H94 map, P3K3499 and P3K3699. The two traits also show a positive correlation when computing the linear correlation coefficient. Fruit length and shape appear to be influenced by each other as both traits are strongly correlated with marker P6K510 and the r-value shows positive correlation. Flesh color was determined a monogenically controlled trait that is heterozygous dominant. The location of this gene was determined to be found close to marker P3K132 on both the Khak dam X 2H94 and AU9 X Sunup linkage maps.

REFERENCES

Barkley, N. A., M.L. Newman, M.L. Wang, M.W. Hotchkiss, and G.A. Pederson (2005). "Assessment of the genetic diversity and phylogenetic relationships of a temperate bamboo collection by using transferred EST-SSR markers." Genome **48**: 731-737.

Bell, C. J., and J. R. Ecker (1994). "Assignment of 30 microsatellite loci to the linkage map of Arabidopsis." Genomics **19**(1): 137-144.

Billotte, N., N. Marseillac, A.-M. Risterucci, B. Adon, P. Brottier, F.-C Baurens, R. Singh, A. Heeran, H. Asmady, C. Billot, P. Amblard, T. Durand-Gasselin, B. Courtois, D. Asmono, S. C. Cheah, W. Rohde, E. Ritter, A. Charrier (2005). "Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.)." Theoretical Applied Genetics **110**: 754-765.

Boluarte-Medina, T., and R. E. Veilleux (2002). "Phenotypic characterization and bulk segregant analysis of anther culture response in two backcross families of diploid potato." Plant Cell, Tissue and Organ Culture **68**: 277-286.

Botstein, D., K. L. White, M. Skolnick and R. W. Davis (1980) "Construction of a genetic linkage map in man using restriction fragment length polymorphisms." American Journal of Human Genetics **32**: 314-331.

Bremer, K., M. W. Chase, and P. F. Stevens (1998). "An ordinal classification for the families of flowering plants." Annals of the Missouri Botanical Garden **85**: 531-553.

Broman, K. W. (2001). "Review of statistical methods for QTL mapping in experimental crosses." Lab Animal **30**(7): 44-52.

Chan, Y. K., and R. E. Paull (2007). "Papaya *Carica papaya* L., Caricaceae." Encyclopedia of Fruits and Nuts. CAB Inc., United Kingdom, Wallingford.

Charcosset, A. (1996). "L'identification de QTL (locus affectant des caractères quantitatifs) est-elle justifiée pour la sélection?" Le sélectionneur français **46**: 35-45.

Chen, X. Y. and S.R. McCouch (2002). "Sequence divergence of rice microsattelites in *Oryza* and other plant species." Molecular Genetics and Genomics **268**: 331-343.

Chin, W. Y. (2003). Tropical Trees and Shrubs: A selection for urban plantings. Singapore, Sun Tree Publishing, Ltd.

Cobley, L. S., and W. M. Steele (1976). An Introduction to the Botany of Tropical Crops. New York, Longman Group Limited.

Coffman, C. J., R. W. Doerge, M. L. Wayne, and L. M. McIntyre (2003). "Intersection tests for single marker QTL analysis can be more powerful than two marker QTL analysis." Biomed Central Genetics **4**(10).

Dahleen, L. S., L.J. Vander Wal, and J.D. Franckowiak (2005). "Characterization and molecular mapping of genes determining semidwarfism in barley." Journal of Heredity **96**(6): 654-62.

Decraene, L. P. Ronse and E. F. Smets (1998) "The floral development and anatomy of *Carica papaya* (Caricaceae)." Canadian Journal of Botany **77**(4): 592-598.

Deputy, J. C., R. Ming, H. Ma, Z. Liu, M. M. M. Fitch, M. Wang , R. Manshardt, J. I. Stiles (2002). "Molecular markers for sex determination in papaya (Carica papayaL)." Theoretical Applied Genetics **106**: 107-111.

El-Lithy, M. E., L. Bentsink, C. J. Hanhart, G. J. Ruys, D. Rovito, J. L. M. Broekhof, H. J. A. van der Poel, M. J. T. van Eijk, D. Vreugdenhil, and M. Koornneef (2005). "New Arabidopsis recombinant inbred line populations genotyped using SNPWave™ and their use for mapping flowering time QTLs." Genetics **10.1534/genetics.105.050617**.

Fitch, M. M. M., R. M. Manshardt, D. Gonsalves, J. L. Slightom, J. C. Sanford (1992). "Virus resistant papaya plants derived from tissues bombarded with the coat protein gene of papaya ringspot virus." National Biotechnology **10**: 1466-1472.

Gallais, A. (1996). "Utilisation des marqueurs en sélection." Le sélectionneur français **46**: 47-58.

Grandillo, S., H. M. Ku, and S. D. Tanksley (1996). "Characterization of *fs8.1*, a major QTL influencing fruit shape in tomato." Molecular Breeding **2**: 251-260.

Grandillo, S., H. M. Ku, and S. D. Tanksley (1999) "Identifying the loci responsible for natural variation in fruit size and shape in tomato." Theoretical Applied Genetics **99**: 978-987.

Horovitz, S., and H. Jimenez (1967) "Cruzamientos interespecificos e intergenericos en caricaceas y sus implicaciones fitotechicas." Agronomia Tropical **17**:323-343.

Jorge, V. A. D., P. Faivre-Rampant, and C. Bastien (2005). "Genetic architecture of qualitative and quantitative *Melampsora larici-populina* leaf rust resistance in hybrid poplar: genetic mapping and QTL detection." New Phytologist **167**: 113-127.

Kaiser S (1935). "The factors governing shape and size in Cupsicum fruits: a genetic and developmental analysis." Bull Torrey Botany.

Club 62: 433-454

Kearsey, M. J., and A. G. Farquhar (1998). "QTL analysis in plants; where are we now?" Heredity **80**: 137-142.

Khazanie, R. (1996). Statistics in a World of Applications. New York, HarperCollins College Publishers.

Kim, M. S., P.H. Moore, F. Zee, M.M.M. Fitch, D.L. Steiger, R.M. Manshardt, R.E. Paull, R.A. AU9, T. Sekioka, and R. Ming (2001). "Genetic diversity of *Carica papaya* as revealed by AFLP markers." Genome **45**: 503-512.

Lai, C. W., Q. Yu, S. Hou, R. L. Skelton, M. R. Jones, K. L. Lewis, J. Murray, M. Eustice, P. Guan, R. Abgayani, P. H. Moore, R. Ming and G. G. Presting (2006). "Analysis of papaya bac end sequences reveals first insights into the organization of a fruit tree genome." Molecular Genetics and Genomics **276**(1): 1-12.

Lapidge, K. L., B. P. Oldroyd, and M. Spivak (2002). "Seven suggestive quantitative trait loci influence hygienic behavior of honey bees." Naturwissenchaften **89**: 565-568.

Lander, E. S., and D. Botstein (1989). "Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps." Genetics **121**(1): 185-199.

Lewis, R. (2001). Human Genetics: Concepts and Applications. New York, McGraw-Hill.

Liscum, M., and P. Oeller (2004). AFLP: not only for fingerprinting, but for positional cloning, University of Stanford publications.

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

Loveless, A. R. (1983). Principles of Plant Biology for the Tropics. New York, Longman Group Limited.

Luo, M. C., C. Thomas, F.M. You, J. Hsiao, S. Ouyang, C. R. Buell, M. Malandro, P.E. McGuire, O.D. Anderson, and J. Dvorak (2003). "High-throughput fingerprinting of bacterial artificial chromosomes using the SNaPshot labeling kit and sizing of restriction fragments by capillary electrophoresis." Genomics **82**: 378-389.

Ma, H., P. H. Moore, Z. Liu, M.S. Kim, Q. Yu, 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.

Malo, S. E., and C. W. Campbell (1994). The papaya. Fact Sheet HS-11. F. C. E. S. Horticultural Science Department, Institute of Food and Agricultural Sciences, University of Florida, University of Florida 1-3.

Martin, F. W. (1984). Handbook of Tropical Food Crops. Mayaguez, Puerto Rico, CRC Press, Inc.

McCallum, J., A. Clarke, M. Pither-Joyce, M. Shaw, R. Butler, D. Brash, J. Scheffer, I. Sims, S. van Heusden, M. Shigyo, and M. J. Havey (2006). "Genetic mapping of a major gene affecting onion bulb fructan content." Theoretical Applied Genetics **112**: 958-967.

Miao, X. X., S. J. Xu, M. H. Li, J. H. Huang, F. Y. Dai, S. W. Marino, D. R. Mills, P. Zeng, K. Mita, S. H. Jia, Y. Zhang, W. B. Liu, H. Xiang, Q. H. Guo, A. Y. Xu, X. X. Kong, H. X., Lin, Y. Z. Shi, G. Lu, X. Zhang, W. Huang, Y. Yasukochi, T. Sugasaki, T. Shimada, J. Nagaraju, Z. H. Xiang, S. Y. Wang, M. R. Goldsmith, C. Lu, G. P. Zhao, and Y. P. Huang (2005). "Simple sequence repeat-based consensus linkage map of *Bombyx mori*." Genetics **102**(45): 16303-16308.

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 Applied Genetics **102**: 892-899.

Ming, R., Q. Yu, and P. H. Moore "Sex determination in papaya." unpublished data.

Mohan, M. S. N., A. Bhagwat, T.G. Krishna, M. Yano (1997). "Genome mapping, molecular markers and marker-assisted selection in crop plants." Molecular Breeding **3**: 87-103.

Mueller, U. G., and L.L. Wolfenbarger (1999). "AFLP genotyping and fingerprinting." Tree **14**(10): 389-394.

Nakasone, H. Y., and R. E. Paull (1998). Tropical Fruits. CAB Inc., United Kingdom, Wallingford.

Parasnis, A. S., W. Ramakrishna, K. V. Chowdari, V. S. Gupta, and P. K. Ranjekar (1999). "Microsatellite (GATA)n reveals sex-specific differences in Papaya." Theoretical Applied Genetics **99**: 1047-1052.

Paterson, A. H., E. S. Lander, J. D. Hewitt, S. Peterson, S. E. Lincoln and S. D. Tanksley (1988). "Resolution of quantitative traits into Mendelian factors, using a complete linkage map of restriction fragment length polymorphism." Nature **335**: 72 1-726.

Paterson, A. H., J. W. DeVerna, B. Lanini, and S. D. Tanskley (1990). "Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes, in an interspecies cross of tomato " Genetics **124**: 735-742.

Purseglove, J. W. (1968). Tropical Crops: Dicotyledons 1. Trinidad, Longmans Green and Co, Ltd.

Ronse Decraene, L. P., and E.F. Smets (1999). "The floral development and anatomy of *Carica Papaya* (Caricaceae)." Canadian Journal of Botany **77**: 582-598.

Samson, J. A. (1980). Tropical Fruits. London, Longman Group Limited.

Sargent, D. J., M. Geibel, J.A. Hawkins, M. J., Wilkinson, N. H. Battey, and D.W. Simpson (2004). "Quantitative and qualitative differences in morphological traits revealed between diploid *Fragaria* species." Annals of Botany **94**(787-796): 787-796.

Sharapova, N., M.D. McMullen, L. Schultz, S. Schroeder, H. Sanchez-Villeda, J. Gardiner, D. Bergstrom, K. Houchins, S. Melia-Hancock, T. Musket, N. Duru, M. Polacco, K. Edwards, T. Ruff, J.C. Register, C. Brouwer, R. Thompson, R. Velasco, E. Chin, M. Lee, W. Woodman-Clikeman, M.J. Long, E. Liscum, K. Cone, G. Davis, and E. H. Coe, Jr. (2002). "Development and mapping of SSR markers for maize " Plant Molecular Biology **48**: 463-481.

Sinnott EW (1936). "A developmental analysis of inherited shape differences in cucurbit fruits." The American Naturalist **70**: 245-254.

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 Applied Genetics **93**: 547-553.

Srivastava, A. K., G. Masinde, H. Yu, D. J. Baylink, and S. Mohan (2004). "Mapping quantitative trait loci that influence blood levels of alkaline phosphatase in MRL/MpJ and SJL/J mice." Bone **35**: 1086-1094.

Steiger, D. L., C. Nagai, P. H. Moore, C. W. Morden, R. V. Osgood, and R. Ming (2002). "AFLP analysis of genetic diversity within and among *Coffea arabica* cultivars." Theoretical Applied Genetics **105**: 209-215.

Tai, T. H., and S. D. Tanksley (1990). "A rapid and inexpensive method for isolation of total DNA from dehydrated plant tissue." Plant Molecular Biology Report **8**: 297-303.

Tanksley, S. D., M. W. Ganal, J. P. Prince, M. C. de Vicente, M. W. Bonierbale, P. Broun, T. M. Fulton, J. J. Giovannoni, S. Grandillo, G. B. Martin, R. Messeguer, J. C. Miller, L. Miller, A. H. Paterson, 0. Pineda, M. S. Riider, R. A. Wing, W. Wu and N. D. Young (1992). "High Density Molecular Linkage Maps of the Tomato and Potato Genome." Genetics **132**: 1141-1160.

Temnykh, S., G. DeClerk, A. Lukashova, L. Lipovich, S. Cartinhour, and S. McCouch (2001). "Computaional and experimental analysis of microsatellites in rice (*Oryza sativa L.*): frequency, length variation, transposon associations, and genetic marker potential." Genome Research **11**: 1441-1452.

Van Droogenbroeck, B., Breyne P., Goetghebeur P., Romeijn-Peeters E., Kyndt T., and G. Gheysen (2002). "AFLP analysis of genetic relationships among papaya and its wild relatives (Caricaceae) from Ecuador." Theoretical and Applied Genetics **105**(2-3): 289-297.

Van Ooijen, J. W. (1999). "LOD Significance thresholds for QTL analysis in experimental populations of diploid species " Heredity **83**: 613-624.

Van Ooijen, J. W. (2004). MapQTL®5.0, Software for the mapping of quantitative trait loci in experimental populations. Wagningen, Netherlands, Kyazma B. V.

Wareing, P. F., and I. D. J. Phillips (1981). Growth and Differentiation in Plants. New York, Pergamon Press.

Xie, X., M. H. Song, F. Jin, S. N. Ahn, J. P. Suh, H. G. Hwang, and S. R. McCouch (2006). "Fine mapping of a grain weight quantitative trait locus on rice chromosome 8 using near-isogenic lines derived from a cross between *Oryza sativa* and *Oryza rufipogon*." Theoretical Applied Genetics **ePub**

Zhang, F., X. Q. Wan, and G. T. Pan (2006). "QTL mapping of *Fusarium moniliforme* ear rot resistance in maize. 1. Map construction with microsatellite and AFLP markers." Journal of Applied Genetics **47**(1): 9-15.