DEVELOPMENT OF PAPAYA CHROMOSOME-SPECIFIC CYTOGENETIC MARKERS AND CYTOGENETIC MAP OF PAPAYA SEX CHROMOSOMES

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAIʻI AT MĀNOA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

TROPICAL PLANT AND SOIL SCIENCES

DECEMBER 2011

BY

CHING MAN WAI

DISSERTATION COMMITTEE

ROBERT E. PAULL, CHAIRPERSON

RAY MING

PAUL H. MOORE

GERNOT PRESTING

QINGYI YU

Acknowledgement

First of all, I would like to thank my committee members (Drs. Robert E. Paull, Ray Ming, Paul H. Moore, Gernot Presting, and Qingyi Yu) for their guidance and support during my PhD study. Their valuable suggestions and comments help me to improve the experimental design and result interpretation. Besides, I would like to thank their effort and time on revising the dissertation and manuscripts, especially Dr. Paul H. Moore, who guides me through the difficult moment on academic writing. In addition, I would like to express my appreciation on the support from Dr. Qingyi Yu, in both research and daily life.

I would also like to thank Drs. Jiming Jiang and Wenli Zhang of the University of Wisconsin at Madison for teaching me FISH and ChIP-FISH techniques. Moreover, I would like to thank Dr. Samual Sun of the Chinese University of Hong Kong for recommending me to study PhD aboard.

I would like to thank the funding agent that this project was partially supported by a USDA T-STAR grant from the USDA-CSREES (Award # 2008-34135-19371) to the University of Hawaii at Manoa and a sub-award to the Hawaii Agriculture Research Center.

With the supports from friends in Hawaii and Hong Kong, it helps me to go through the tough time. I would especially like to thank Dr. Chris Russell, who cheers me up all the time in our geeky and nerdy conversation. Last but not the least. I would like to thank my family for their unconditional love and caring.

Abstract

Papaya (2n =2x =18) is a tropical and subtropical fruit tree grown worldwide. In recent years, the papaya genome has been sequenced and genetic maps have become available for analyzing genomic structure of the male-determining region of Y chromosome (MSY). However, research in the cytological aspect lags behind the model crops. One reason for slow cytological progress is because the papaya condensed chromosomes are small and similar in size, making it difficult to distinguish among them based on chromosomal morphology or staining pattern. By probing nine microsatellite marker-tagged BAC clones on papaya chromosome spread using fluorescence in situ hybridization (FISH) techniques, individual papaya metaphase chromosomes were distinguished and integrated with a genetic map. Then, a detailed cytogenetic map was constructed for the papaya sex chromosome (chromosome 1) by probing 15 microsatellite marker-tagged and 2 cytological-feature associated BAC clones on highly resolved papaya pachytene chromosomes. This cytogenetic map integrated the chromosomal features with genetic and physical data. Analysis of recombination rates along chromosome 1 revealed a small non-recombining region (12.2% of chromosome length) where the MSY is located, suggesting that papaya has a pair of young evolving homomorphic sex chromosomes. Sequence analysis of euchromatic and heterochromatic regions on the X chromosome showed an uneven distribution of genes and transposable elements. This cytological information enhances our understanding on the relationship between genome composition and chromosomal structure.

Table of content

List of Figures

List of Tables

Chapter 1 Introduction

Papaya is a tropical fruit tree that is grown also in the sub-tropics with a short life cycle of 9 - 15 months. The fruit contains a high content of vitamins A and C and a half papaya can fulfill the adult minimum daily requirement of vitamin C as recommended by the FDA (USDA-ARS 2011). In 2008, the total import and export of papaya to USA was 0.12 million tons at a value of US$ 72.3 million (FAOSTAT 2008). In addition to its fruit food value, the protein papain extracted from papaya is also used for industrial applications such as meat tenderizer and beer clarification.

In 2008, a 3X coverage genome sequence on transgenic Hawaiian solo papaya 'SunUp' was released (MING *et al.* 2008). This data enables us to compare papaya's genome architecture with that of other plants. Developing an understanding of the papaya sex determination mechanism has recently evolved as one of the more interesting topics in papaya research. Two genetic maps revealed that the sex determination region (male specific region of Y chromosome, MSY) is a relatively small segment of chromosome 1 showing severe suppression of recombination (CHEN *et al.* 2007; MA *et al.* 2004). Fluorescence *in situ* hybridization (FISH) of this sex chromosome with male-specific BAC clones positioned the MSY region in the middle of chromosome 1. Five darkly DAPI-stained knobs are located on the Y chromosome of hermaphrodite (Y^h), but only one knob is on the X chromosome (ZHANG *et al.*). This genetic and genomic information facilitates an in-depth study of sex determination mechanisms in plants.

Prior to now, no robust method has been developed for papaya chromosome identification because all metaphase chromosomes are similar in size and no knowledge on pachytene chromosomes is available, except for recent cytological characterization of the MSY region. FISH-based chromosome identification is one of the more common and convenient cytological methods used in animals and plants. This method has been used to develop chromosome-specific markers in rice (CHENG *et al.* 2001a), cotton (WANG *et al.* 2007), sorghum (KIM *et al.* 2002), tomato (TANG *et al.* 2009b), *Antirrhinum majus*

(ZHANG *et al.* 2005), potato (DONG *et al.* 2000) and common bean (PEDROSA *et al.* 2003). In addition, FISH has been used to integrate genetic and cytological maps in order to study the relationships between recombination frequency and chromosome structure (JIANG and GILL 2006).

In this study, we developed a set of chromosome-specific BAC clones to establish a reproducible papaya chromosome identification method. This method enables us to position any unknown BAC clones or repetitive sequences on specific chromosomes, even without having complete genome sequences of the chromosomes. In addition, we constructed a cytogenetic map for papaya sex chromosome (chromosome 1) to reveal the relationship between the genome architecture and chromosome structure. This information will help us to understand the difference in genome structure between the heterochromatic and euchromatic regions of chromosome 1. Once the sequencing of MSY and corresponding X regions is completed, a complete picture of genome composition of this evolving sex chromosome can be established.

Chapter 2 Literature Review

2.1 Introduction

Papaya, *Carica papaya* L., is a perennial fruit tree that grows in tropical areas and to a lesser extent in the sub-tropics. It is either dioecious (unisexual flowers on male or female plants) or gynodioecious (unisexual flowers on female plants and bisexual flowers on hermaphrodite plants). Male plant inflorescences have long peduncles bearing many flowers while female plants have short inflorescences with just a few flowers. Inflorescences of hermaphrodite plants are similar to those of female plants except that they bear bisexual flowers that can easily self-pollinate. Fruit are set throughout the year and each one takes 4-5 months to mature. The fruit flesh color ranges from pale yellow to red, depending on variety.

Papaya is usually consumed as fresh fruit, but is also used in drinks, candies and jams. The fruit contains high content of vitamins A and C that the content of one papaya fruit exceeds the adult minimum daily requirements of these two vitamins (USDA-ARS 2011). Papain is collected from papaya for industrial uses, such as tenderizing meat and chill proofing beer. In 2008, the total production of papaya in 56 countries was 9.1 million tonnes and the two largest producers were India and Brazil with 3.6 and 1.9 million tonnes harvested, respectively (FAOSTAT 2008).

2.2 Model organism of tree crops

Papaya has several characteristics that give it advantages as a model for tree fruit crops. Papaya plants are relatively small and can be grown at high density with 1,200 - 2,000 trees per hectare. Plant growth rate is rapid taking only 9-15 months generation time (from seed of one generation to seed of next generation). Plants produce flowers continuously throughout year and thus fruits can be collected all year around. Each fruit usually bears 800 – 1,000 seeds that provide abundance of offspring for genetic studies.

As there are three sex types of papaya trees, researchers can either allow self-pollination on hermaphrodite trees or perform controlled crossing on male and female trees.

The small haploid genome of 372 Mb (ARUMUGANATHAN and EARLE 1991) and nine pairs of chromosomes make papaya a good candidate for genomic studies. The ease of cloning individuals from cuttings and the availability of a transformation system (FITCH *et al.* 1993; FITCH *et al.* 1990) further facilitate genetic analysis in papaya. In addition, papaya is in same order, Brassicales as is the well-sequenced model plant *Arabidopsis thaliana* that diverged from papaya 72 million years ago (mya), and is thus a good candidate to serve as an outgroup for papaya in evolutionary studies.

2.2.1 Cytological resources

Papaya has nine pairs of chromosomes that are small and similar in size (HEILBORN 1921; KUMAR and ABRAHAM 1945; STOREY 1953). Due to their small size, it is difficult to distinguish individual chromosomes based on morphology (e.g. arm ratio, length) or conventional band staining (e.g. G-banding, C-banding). Unlike major cereal crops such as barley, wheat, and maize, papaya has no deletion or translocation lines, which makes it difficult to associate phenotypic traits with chromosome cytological features. Recently, sex chromosome specific bacterial artificial chromosome (BAC) clones were labeled on highly resolved papaya pachytene chromosome spreads for chromosome karyotyping by using fluorescence *in situ* hybridization (FISH) (YU *et al.* 2007a; ZHANG *et al.* 2008).

2.2.2 Genome mapping

BAC libraries are essential for sequence assembly and physical mapping of genomes. A BAC library of transgenic papaya 'SunUp' hermaphrodite plant has been constructed from two ligation reactions with an average insert size of 132 kb (MING *et al.* 2001b). In total, 39,168 BAC clones were obtained and provided 13.7X genome coverage, based on the papaya genome size of 372 Mbp. The BAC end sequences (BES), ~500 bp sequence from both ends, of 26,017 BAC clones provided initial insights of the papaya genome

organization (LAI *et al.* 2006). The resulting BESs represent 4.7% (17.5 Mb) of the papaya genome. Of the 35,472 BESs, 6,769 (19.1%) showed homology to at least one arabidopsis cDNA. Using an average gene length of 2 kb, the number of papaya genes was estimated at about 35,000. Most of the paired BESs showed a higher level of collinearity to poplar rather than to arabidopsis, which was surprising as papaya and arabidopsis are both in the order Brassicales. About 16.2% of the BESs shared homology to known plant repeat elements; class I retrotransposons (83.3%) were the largest group of repeat elements while class II transposons (7.4%) ranked second. Interestingly, for the most abundant papaya-specific repeats screened, six out of ten were repeats located in the male specific region (MSY) of papaya Y chromosome.

The entire set of 39,168 BAC clones was fingerprinted using the high-information-content fingerprinting system for physical map construction (YU *et al.* 2009). A total of 30,824 high quality fingerprints were assembled using the FPC program that assigned 26,466 BAC clones onto 963 FPC contigs while 4,358 BAC clones remained as singletons. In those assembled contigs, 535 of them (equivalent to 21,371 BAC clones) were anchored to a papaya microsatellite genetic map to facilitate the integration of physical and genetic maps. This physical map is a framework for whole genome shotgun sequence assembly and will assist map-based cloning in papaya.

A 3X coverage genome of the transgenic papaya cultivar 'SunUp' female plant was published in 2008 (MING *et al.* 2008). The 271 Mb of assembled contig sequences represent ~75% of the 372 Mb papaya genome and over 90% of the euchromatic region. The number of predicted genes in papaya (24,746) is much lower than any published flowering plant genome, being 11-20% fewer than in arabidopsis, 34% fewer than in rice, 46% fewer than in poplar and 57% fewer than in apple. When the collinearity between arabidopsis and papaya genomes was compared, most of the papaya scaffolds were collinear with two to four segments of the arabidopsis genome, which suggests that genome duplication occured in arabidopsis after papaya and arabidopsis diverged (LYONS *et al.* 2008). The MSY region on Y chromosome was estimated to be ~8 Mb with 85.6%

of sequence being composed of repeats containing *Ty1/gypsy* (54.1%) and *Ty3/copia* (1.9%) retrotransposons.

The repetitive elements in papaya were analyzed genome wide, based on whole genome shotgun sequences (NAGARAJAN *et al.* 2008). A papaya-specific repetitive element database was built, using homology-based and *de novo* methods to annotate transposable elements. The papaya genome is mainly composed of transposable elements (43.4%) with homology to other plant species, unannotated transposable elements (8.5%), tandem repeats (1.3%) that ranged from 1 - 2000 bp, and a high copy number of genes (>2%). Overall, the papaya genome has 42.8% class I retrotransposons and only 0.2% class II transposons. Similar to BES analysis, *Ty3/gypsy* is the major repetitive element which accounts for 27.8% of the genome. The ratio of *Ty3/gypsy* to *Ty1/copia* elements in papaya is close to a 2:1 ratio, which differs from the 1:1 ratio in the arabidopsis and maize genomes (MESSING *et al.* 2004).

In addition to genome sequence data and three BAC libraries, five papaya flower cDNA libraries have been constructed (Yu, Q. and Ming, R,. unpublished results). Three flower libraries derived from pre-meiosis (< 4 mm) flower buds of either male, female, or hermaphrodite plants and two flower libraries derived from mature flower buds of either female or hermaphrodite plants. ESTs from these five libraries were sequenced from the 5' end. After removal of repeat, mitochondria, chloroplast, and vector sequences, EST sequences were clustered based on local similarity scores of pairwise comparison using 88% similarity over 100 nucleotides. A unigene set of 8,571 EST contigs and singletons were assembled.

2.2.3 Genetic mapping

A high-density genetic map is important for assigning BAC clone sequences to individual chromosomes by integrating genetic marker-tagged BAC clones and physical maps. The first papaya genetic map, (HOFMEYR 1938), was a single linkage group containing three morphological markers: sex type, flower color and stem color. Fifty-eight years later, a

second map with 11 LGs was developed consisting of 62 randomly amplified polymorphic DNA (RAPD) markers, and sex type (SONDUR *et al.* 1996). Ma et al. (2004) constructed a high-density genetic map containing 1,498 amplified fragment length polymorphism (AFLP) markers, PRSV coat protein marker, sex type and fruit flesh color. These markers were developed from a 'Kapoho' X 'SunUp' F2 population and distributed along 12 linkage groups (LGs) with an average distance of 2.19 cM between markers. The total genetic length of this AFLP map is 3,294 cM.

Preliminary sequence information became available during whole genome shotgun sequencing of a 'SunUp' female plant and BAC end sequencing of a BAC library, to design sequence-tagged microsatellite (or simple sequence repeat, SSR) markers. SSRs are highly informative markers that can integrate genetic, physical and cytogenetic maps. By using an 'AU9' X 'SunUp' F2 population, a genetic map was constructed with 706 SSR markers and one fruit flesh color marker (CHEN *et al.* 2007). Compared to the earlier AFLP map, this SSR map had shorter genetic length (1069 cM), but it contained the same number of LGs (nine major LGs and three minor LGs). The average distance between SSR markers was 1.51 cM, almost 50% less than the one of the AFLP map. A good genetic map should have same number of LG as the haploid chromosome number. However, the number of LG in both maps is still higher than nine, the haploid number of papaya chromosomes, indicating a failure of linkage analysis to integrate the three minor LGs to the nine major LGs.

Attempts were made to bridge the gaps between LGs by adding additional markers to the existing SSR map. After adding 277 AFLP markers to the SSR map, the average distance between markers was decreased by one-third to 1.0 cM and the total genetic length was decreased to 945 cM (BLAS *et al.* 2009). Gaps that had been larger than 5 cM were reduced from 48 to 27 due to higher marker density. However, it resulted in 14 LGs (including nine major and five minor LGs) and thus the integrated maps were still unable to assign LGs to chromosomes.

2.3 Sex determination in papaya

Sex determination in papaya is mainly controlled by genetic factors that are heterogametic in males and hermaphrodites but are homogametic in females. Hofmeyr and Storey both suggested that the sex type is determined by a single gene having three allelic forms: M, male; M^h, hermaphrodite; and m, female (HOFMEYR 1938; STOREY 1953). The homozygous condition of either male or hermaphrodite alleles (MM, MM^h or M^hM^h) are lethal to embryo development. Therefore, in the absence of homozygous males, the seeds from self-pollinated hermaphrodites segregate hermaphrodite to female in a 2:1 ratio. For seeds from cross-pollinated females, it segregates hermaphrodites to females, or males to females, at the ratio of 1:1.

Female plants require male trees for pollination which leads to a loss of 6-10% of field area if females are used for production. On the other hand, every hermaphrodite tree produces fruit so hermaphrodites are preferred in most regions for commercial production. To circumvent the problem of having female trees in the orchard, farmers usually plant five or more seeds in a hill and wait for 4-6 months until flowering which allows for sex determination of the tree. All female trees are then removed to maximize field productivity. The process of growing extra plants to assure a field of hermaphrodites is inefficient in use of land, water, nutrients and labor so developing true breeding hermaphrodites is a worthy goal. The development of four SCAR markers linked to the sex determination locus allows determination of sex type at an earlier developmental stage (DEPUTY *et al.* 2002) but because of the large number of plants to be tested, is too costly to be used commercially.

2.3.1 Identification of primitive sex chromosomes in papaya

The high density linkage map generated from AFLP markers (see section 1.5), mapped the sex determination locus to the middle of LG 1 where 225 sex co-segregating markers were clustered (LIU *et al.* 2004; MA *et al.* 2004). This non-recombining region accounted for 66% of the 342 markers on LG 1. Clustering of sex co-segregating markers was also

observed in the SSR map and the combined SSR and AFLP map, but to a lesser extent. These maps suggest that LG 1 represents the sex chromosome in papaya and that the non-recombining region represents the sex determination locus.

The sex determination locus was fine-mapped using 4,380 informative chromosomes (two each from 2,190 female and hermaphrodite plants from three F_2 and one F_3 populations), 2 SCAR markers (W11 and T12), 3 sex-linked AFLP markers (cpsm 10, 31 and 54), and 1 BES (cpbe 55) (LIU *et al.* 2004). However, not even one recombination event was detected in this region containing a large fraction of the polymorphic markers of LG 1. This indicates that the X and Y homologous chromosomes are highly differentiated at this region. This information combined with the evidence of suppression of recombination, suggests that sex determination in papaya is controlled by a pair of primitive sex chromosomes and that the non-recombining region is the male specific region (MSY) of the Y chromosome. The non-recombining region was physically mapped using a 13.7x BAC library to produce a 2.5 Mb physical map containing 57% of the sex co-segregating AFLP markers. Random subclones from the BAC clones on the physical map were sequenced to give a glimpse of genome feature of this non-recombining region. Based on the 517 kb obtained, this region showed a 37.3% lower gene density, 27.6% higher retroelement density, and 188.9% higher inverted repeat density compared to the genome-wide papaya sequences.

2.3.2 Physical mapping and sequencing of MSY and X regions

After construction of the 2.5 Mb physical map of the MSY region, the previously identified BAC clones were used as bait to detect contigs from the genome wide physical map. Chromosome walking extended the contigs and microsatellite markers from the SSR genetic map and also facilitated its physical mapping. The positions of selected BAC clones were verified by FISH. The estimated physical size of the MSY was then calculated as 8 - 9 Mb and was mapped into four ordered contigs (YU *et al.* 2007a). Initially, five mapped Y- BAC clones (54H01, 76M08, 42B05, 41F24, and 94E22) were sequenced to characterize the genome organization of MSY (YU *et al.* 2007a). Sequence

analysis showed that these BACs contain an abundance of *Ty3/gypsy* retroelements and no protein encoding gene was detected in 715 kb, which demonstrates the paucity of genes on the MSY.

Subsequently, two X-specific BAC clones, 61H02 and 53E18, and their Y^h counterparts, 95B12 and 85B24, were sequenced and the sequence data used to calculate an estimated time of divergence between the X and Y chromosomes (YU *et al.* 2008). Direct alignment of the two pairs of X and Y^h BAC clones revealed three inversion events. Furthermore, 9.6% and 35.2% of sequence expansion was observed on the Y^h BAC clones 95B12 and 85B24, respectively. Four genes were found on two Y^h BAC clones, and all four genes had their X counterparts. These four pairs of genes showed high similarity in exon sequences that ranged from 94.8% to 100% while introns shared slightly lower similarity, ranging from 82% to 99.2%. These four X-Y^h gene pairs were used to determine the degree of synonymous (K_s) and non-synonymous (K_a) divergence between them. The K_a/K_s ratio of the four gene pairs ranged from 0.04-0.5, suggesting the sequence divergence has been constrained. By assessing the degree of silent site nucleotide divergence, the time of divergence between X-Y^h gene pairs was estimated to be 0.5-2.2 million years ago, supporting the concept of papaya having primitive sex chromosomes.

2.3.3 Molecular cytogenetics of sex chromosomes

Cytogenetics has played an important role in analyzing genome features of papaya sex chromosomes. The chromosomal location of candidate X- or Y- BAC clones was confirmed by FISH. Cytogenetic results indicated that the sequence divergence occurs on certain regions of MSY as indicated by uneven signals from labeled Y- BAC clone 54H01 on pachytene X and Y chromosomes (YU *et al.* 2007a). The Y-specific BAC clones, 76M08 and 79C23, were verified by showing signals only on the Y^h pachytene chromosome spreads.

Detailed morphology of the MSY region was revealed by FISH analysis of high-resolution pachytene chromosomes. The papaya sex chromosomes are the second

longest among the nine pairs of chromosome, and the X chromosome is more euchromatic than its corresponding Y (ZHANG *et al.* 2010). Five DAPI darkly-stained knobs can be observed on the MSY region of the Y^h chromosome (ZHANG *et al.* 2008). These five knobs are noted as K1 through K5. K1 is the largest knob and is associated with both the X and Y^h chromosomes; K2 to K5 appear only on the Y^h chromosome. A twist in pairing between K4 and K5 is always observed which may due to a larger agglomeration of DNA on that region of the Y compared to its X counterpart. The distance from K1 to K5 occupies 13% of the sex chromosome length. Based on 5-methylcytosine immunodetection and FISH, all five knobs appear highly methylated compared to their X counterparts and two knobs (K2 and K4) contain 5S ribosomal DNA sequences.

The sex determination locus in the AFLP genetic map located to near the middle of LG 1 and it was postulated that the MSY may locate on or near the centromere (MA *et al.* 2004). By simultaneously labeling meiotic metaphase I chromosome spread with knob-specific BAC clones, Zhang et al. (2008) reported that BAC clone 52H15 associated with K4 and was mapped to the most poleward position, indicating that the centromere of the sex chromosome is located on or near to K4. This supported the hypothesis that the first sex determining gene of papaya must have located within the centromeric regions where severe suppression of recombination occurs.

2.4 Plant cytogenetic

2.4.1 Introduction

The study of chromosomes started in late nineteenth century when the first detailed descriptions of mitotic chromosomes in plants were published by Strasburger in 1875 and in animals by Flemming in 1879. The word ‘chromosome’, which means color (*chroma*) and body (*soma*), was introduced by the German anatomy professor Waldeyer in 1888. After the rediscovery of Mendel’s work in 1900, scientists soon realized that the behavior of chromosomes during cell division was the key point to explain the precise distribution

of hereditary factors to offspring. In the late 1920s, McClintock discovered that a group of traits inherited together were linked to a specific maize chromosome. Later, she observed the relationship between chromosome recombination and formation of inheritable new traits in maize. A chromosome not only carries DNA and genes, its behavior also controls the mechanism of inheritance and its organization affects the activity of genes. With the emergence of new molecular technologies, cytogenetic has entered a new era that facilitates genome and evolution research in both animals and plants.

2.4.2 Fluorescence *in situ* hybridization (FISH) and resolving power of chromosomes

As indicated by the name, chromosomes take up color when dyes are applied. This has enabled cytogenetists to develop different staining methods capable of distinguishing individual chromosomes. However, stain pattern identification of chromosomes does not provide any information about the genome or chromosome organization and is limited in use to those organisms with large chromosomes. The more recent groundbreaking development of DNA-DNA *in situ* hybridization (ISH) procedures allows correlation of DNA sequences to chromosome structures. This technique was first described by Gall and Pardue (GALL and PARDUE 1969) and John et al. (JOHN *et al.* 1969). The probes used initially were radioactively labeled, but later radioactivity was replaced by non-radioactive labeling methods for easy handling. Currently, the probes are either directly labeled with fluorophores or indirectly labeled with conjugates (e.g. digoxigenin, biotin) which are later detected by fluorophore-linked antibodies.

Recently, the methods of detection were changed from the formation of color precipitates to visualization of fluorophores under the microscope. This more direct method for visualization enhanced image resolution and allowed simultaneous detection of multiple signals. To date, with the improvement of probe labeling methods and image capture systems, fluorescence *in situ* hybridization (FISH) is the most commonly used method in plant cytogenetic (reviewed by (JIANG and GILL 2006). FISH in plants has been performed on mitotic and meiotic metaphase chromosomes, interphase chromosomes,

meiotic pachytene chromosomes, super-stretched chromosomes and extended DNA fibers. Each type of chromosome spread has its own advantages and disadvantages.

Metaphase chromosomes are commonly used in FISH. To collect metaphase chromosome spreads, usually root or shoot apical meristems are treated with metaphase-arresting chemicals (such as 8-hydroxyquinoline and *n*-bromonaphthalene) which stop dividing cells at the metaphase stage to enrich the fraction of cells with metaphase chromosomes. This method has been used extensively in plant cytogenetic as root tips and shoot apical meristems are readily available. Since metaphase chromosomes are condensed and short (1 - 5 µm), this analysis has low resolving power and cannot provide details on chromosome features such as the boundaries between euchromatin and heterochromatin. Meiotic pachytene chromosomes spreads isolated from pollen mother cells are 10 - 30 times longer than metaphase chromosomes and can reveal detailed chromosome structure. In maize, it has been reported that metaphase chromosomes can resolve probes that are 3.3 - 8.3 Mbp apart, depending on the position of the probe on the chromosomes. Contrarily, pachytene chromosomes can resolve probes that are less than 0.275 Mbp apart (DANILOVA and BIRCHLER 2008). Thus, the axial resolution of FISH on pachytene chromosomes is at least 12-30 times higher than on metaphase chromosomes. Compared to metaphase chromosomes, pachytene chromosomes require a greater number of samples and more time for chromosome preparation. Also, the long pachytene chromosomes can tangle with other chromosomes causing difficulty in identifying individual chromosomes.

Alternatively, stretching metaphase chromosomes with ethanol and acetic acid, can increase their spatial resolution by more than 100 times and provide a mapping resolution of up to 70kb (VALARIK *et al.* 2004). However, this method requires a chromosome-sorting facility and is applicable only to plant species in which individual chromosomes can be sorted. Recently, Koo and Jiang (2009) developed a simple protocol to stretch maize pachytene chromosomes by 20 fold over its original length and resolve probes separated by 50 kb. This super-stretched pachytene chromosome spread provides an increasingly powerful approach for high-resolution cytogenetic mapping.

Extended DNA fibers provide the highest resolving power that is capable of separating signals a few kilobases apart. In arabidopsis and rice, every micrometer (µm) of DNA fibers could separate signals that are 2.87 - 3.27 and 3.21 kb apart, respectively (CHENG *et al.* 2002; FRANSZ *et al.* 1996; JACKSON *et al.* 1998). Fiber-FISH has been used in different plant genome mapping projects, such as the analysis of the structure and organization of repetitive DNA sequences and estimating the size of physical gaps in rice sequence maps (JIANG and GILL 2006). However, the probe signals may vary significantly depends on degree of DNA extension. Therefore, a large number of samples must be evaluated and statistical analysis is required for accurate measurements. The other limitation is that the signals from short probes are difficult to distinguish from background signals.

2.5 Use of fluorescence *in situ* hybridization in plant genome research

2.5.1 Chromosome identification

Chromosome identification is fundamental for integrating phenotypic traits, cytogenetics, and genome structure. Individual chromosomes may be distinguished by morphology (length, arm ratio, centromere position and presence of second constriction sites) or banding patterns after staining. Common staining methods include C-banding, Q-banding, and G-banding (APPELS *et al.* 1998; BICKMORE and CRAIG 1997). Since different sequences bind to each staining dye with a different stringency, and the distribution of sequences varies among chromosomes, the sequence and dye interactions allow scientists to construct a chromosomal ideogram based on the staining pattern of specific dyes. However, this method is useful only for species having large sized chromosomes. In addition, the staining patterns may vary between cultivars (HAINER and HESEMANN 1988), which limit its application to certain plant species only.

With the development of FISH, cytogenetists can hybridize chromosomes with fluorescently labeled DNA or RNA fragments to facilitate chromosome identification.

The DNA sequences used in FISH need to be sufficiently large to bind enough tag to be visualized. The two main types of DNA sequences used are repetitive sequences or large-insert genomic DNA clones. Repetitive sequences such as ribosomal DNA (rDNA) and transposable elements are dispersed throughout the plant genome. Commonly, either a single color signal representing a single repeat sequence, or multiple color signals from a probe cocktail composed of several repetitive sequences are used (LENGEROVA *et al.* 2004; MURATA *et al.* 1997; SZINAY *et al.* 2008). For example, the 10 pairs of maize mitotic chromosomes in 14 lines was distinguished by probes consisting of centromeric repeats, rDNA, knob repeats and three repetitive sequences in multicolor FISH (KATO *et al.* 2004). In addition to mitotic chromosomes, this multicolor FISH method can be used on meiotic pachytene and late prophase I chromosome spreads. A similar repeat-based FISH method was developed in maize by including chromosome length and signals from centromeric repeats for chromosome identification (CHEN *et al.* 2000; SADDER and WEBER 2001). One of the disadvantages of using repetitive sequences as probes is that the signals from multiple probes on the same chromosome spreads are difficult to distinguish when the chromosome length is short, especially on mitotic metaphase chromosome spreads. The FISH signal patterns from repeats may vary among different varieties and accessions, which limits the application of those probes in chromosome identification.

For plant species with small genome size and low repeat content, large-insert genomic DNA clones are the better choice for use as FISH probes. Genomic DNA clones, such as bacterial artificial chromosome (BAC) and yeast artificial chromosome (YAC), are commonly used in sequencing projects and can be developed as chromosome-specific cytogenetic markers for chromosome identification. For example, cucumber has a small genome size of 367 Mbp and its seven pairs of metaphase chromosome can be distinguished unambiguously by hybridization with 14 chromosome arm-specific fosmid clones (REN *et al.* 2009). Because these clones carry microsatellite markers of the genetic map, the genetic and chromosome maps were integrated for analyzing the cucumber genome organization. These analyses allowed reassigning the orientation of linkage groups based on arm lengths and direction. This approach has been successfully used for

chromosome identification and molecular cytogenetic map establishment in rice (CHENG *et al.* 2001a), cotton (WANG *et al.* 2007), sorghum (KIM *et al.* 2002), tomato (TANG *et al.* 2009b), *Antirrhinum majus* (ZHANG *et al.* 2005), potato (DONG *et al.* 2000) and common bean (PEDROSA *et al.* 2003). In arabidopsis, the complete or partial chromosomes were painted with contiguous BAC clones of a total length of 2.6 - 14.4 Mb to study chromosomal behaviour at different stages of division (LYSAK *et al.* 2001; LYSAK *et al.* 2003). This chromosome-specific cytogenetic markers approach has several advantages over the repeat-based approach: 1) It is highly reproducible as each chromosome has distinctive features; 2) The quality of chromosome preparation is not as critical since both metaphase and pachytene spreads can be used; 3) This system can be applied to most species, specifically advantageous for those species with small chromosome size. For plant species with large genome size such as maize, chromosome-specific BAC clones may be difficult to obtain due to highly abundant repetitive sequences. This limitation can be overcome when the genome sequence is available. Lamb et al. (2007) demonstrated that using repeat-free sequences as probes, chromosome karyotyping can be performed on maize metaphase spreads and these probes were also able to produce signals in wild relatives *Z. Luxurians*, *Z. Diploperennis* and *Tripsacum dactyloides* for comparative genomics.

2.5.2 Integration of genetic, physical and cytological maps

Cytological maps are created by positioning microscopically visualized structures on chromosome morphology. However, cytological maps do not contain any information about recombination rates nor the physical distance in base pairs between the visualized structures. Positions of probes on chromosomes can be represented as the percentage of distance between probe signals relative to arm or total chromosome length for quantifying the cytological distance.

A genetic map is based on the recombination frequency of DNA markers in the parents and their progeny population. However, the genetic distance between markers does not reflect the real physical distance, but only the relative level of recombination along the

chromosome. Suppression of recombination where clustering of markers occurs, is observed near the centromeric region of plant species. In the common bean chromosome 7, the markers around the centromere made up 15% of genetic length but occupied 50% of the chromosome physical length (PEDROSA-HARAND *et al.* 2009). While a physical map can give exact chromosomal position of markers, a complete physical map is difficult to achieve due to highly repetitive sequence content in the centromeric and heterochromatic regions. This problem can be overcome by integrating the cytological and genetic maps (i.e. cytogenetic map) by assigning DNA markers to chromosomes. An additional benefit is that the order of clustered genetic markers can be resolved by the cytogenetic map without data of the assembled sequences.

BAC or YAC clones are used in chromosome identification because they usually carry DNA markers such as AFLPs and microsatellites, which are good candidate clones for integrating genetic and cytogenetic maps by FISH. The position of DNA clones on chromosomes can be observed directly by FISH and thus reveal the relationship between recombination frequency and chromosomal features. The chromosomal position of DNA clones can also be used for physical distance estimation. Plant species for which physical maps were constructed for integration into the existing molecular cytogenetic maps include Arabidopsis (LIN *et al.* 1999; MAYER *et al.* 1999; SALANOUBAT *et al.* 2000; TABATA *et al.* 2000; THEOLOGIS *et al.* 2000), maize (WEI *et al.* 2009), and rice (FENG *et al.* 2002; SASAKI *et al.* 2002).

An alternative way to integrate genetic and cytological maps is through the use of cytogenetic stocks. In general, a population with abnormal chromosomes features (e.g. translocation, deletion) are used and chromosomal breakpoint can be revealed by microscopy combined with dye staining. The existence or location of molecular markers can also be studied by molecular techniques such as PCR. For example, wheat genetic markers generated by PCR were placed on a chromosome 3B deletion line to reveal the recombination rate along the chromosome (SAINTENAC *et al.* 2009). Translocation lines are another common cytogenetic stock that has been used to integrate an RFLP map with a cytological map in barley (KUNZEL *et al.* 2000). This approach has been used

intensively in wheat (SAINTENAC *et al.* 2009), barley (KUNZEL *et al.* 2000) ,and maize (SHERIDAN and AUGER 2006) which have large chromosome size and thus easily show abnormalities in chromosome structure. However, cytogenetic stocks are not available for most plant species. Fortunately, large-insert clone libraries can be constructed easily at a reasonable cost for use in FISH. Also, more time, space, and money is required to cultivate and maintain cytogenetic stocks than to employ molecular markers. Another disadvantage of using breakpoint analysis is that it is difficult to precisely determine the physical position of a chromosome break. Thus, FISH-based approach is more universal and applicable to most plant species.

FISH-based integrated maps have been developed for the sequenced plants: rice (CHENG *et al.* 2001b; KAO *et al.* 2006), sorghum (ISLAM-FARIDI *et al.* 2002; KIM *et al.* 2002; KIM *et al.* 2005a; KIM *et al.* 2005b), and maize (WANG *et al.* 2006a). The recombination frequency along chromosomes of all three species revealed suppressed recombination rate at the centromere, but the degree of suppression varied among the species. Among the ten pairs of sorghum chromosome, all centromeric regions have lower recombination frequency than the surrounding area, but in different degrees among the chromosomes (ISLAM-FARIDI *et al.* 2002; KIM *et al.* 2005a; KIM *et al.* 2005b). For example, the ratio of cytological to genetic distance of the centromere in sorghum chromosome 1 was about 80, but in chromosomes 2 and 8, the ratio was 9 and 8, respectively, which are 10-fold lower compared to the ratio of chromosome 1.

The euchromatic region also shows uneven recombination frequency between the short and long arms of specific chromosomes such as rice chromosomes 5 (KAO *et al.* 2006) and 10 (CHENG *et al.* 2001b), maize chromosome 9 (WANG *et al.* 2006a), and tomato chromosome 6 (PETERS *et al.* 2009). It is likely that each chromosome has its own recombination frequency pattern in different regions. Nevertheless, there is the general trend that the centromere has the lowest recombination rate with increasing frequency of recombination towards the two distal ends.

The quality of genetic maps obviously depends on the density of markers and the coverage of linkage groups on the chromosomes. Integrated cytogenetic maps reveal whether a large genetic distance gap is caused by a recombination hot spot or a large physical distance. This information helps determine if additional markers are needed to saturate the genetic map. In addition, coverage of a genetic map can be revealed by localizing the distal marker anchored DNA clones on chromosomes by FISH. Kim et al. (2005b) showed by FISH that LG 2 and 8 of the sorghum genetic map covered >95% of chromosome length.

Physical maps contains details on gene and repetitive sequence distribution, this information combined with cytogenetic map enables calculations of the correlation of recombination rate, gene and repeat content, with chromosomal morphology, such as telomeres, centromeres, and heterochromatic knobs (PETERS *et al.* 2009). Moreover, the condensation of chromatin in different regions can be revealed by integrating cytological and physical maps (CHENG *et al.* 2002; WANG *et al.* 2006a; ZHONG *et al.* 1998). In the short arm of the arabidopsis chromosome 4, the condensation degree of euchromatin ranged from 230 to 550 kb/µm and the decondensation rate on euchromatin was higher than on the heterochromatic knob (FRANSZ *et al.* 2000). The average chromatin condensation degree can be measured in plant species even without a completely assembled physical map, which is useful for estimating the physical size between two DNA clones and gaps on physical map (BUDIMAN *et al.* 2004; FENG *et al.* 2002; SASAKI *et al.* 2002; TANG *et al.* 2009a).

An integrated cytogenetic map unifies recombination frequency with chromosomal features. Such maps facilitate quantitative trait locus (QTL) mapping by estimating the physical distance of markers flanking the genetic determinates of important agronomical traits. DNA clones of one species can be used as probes for FISH mapping in a related species through comparative genomics. Direct visualization of DNA clones on chromosomes enables one to define the centromere position, as well as the boundary between euchromatin and heterochromatin on genetic and physical maps. This

information gives a better understanding on genes and repeats distribution along chromosomes.

2.5.3 Study of intergenomic and intragenomic genome architecture

The sorghum genome map shows extensive macrocollinearity with rice (KLEIN *et al.* 2003) even though these two species diverged ~50 mya. Six pairs of sorghum chromosomes have regions of collinearity to six pairs of rice chromosomes (MOORE *et al.* 1995; WILSON *et al.* 1999). The cytogenetic map of the 10 sorghum chromosome pairs was constructed by probing 18 - 30 BAC clones per chromosome. The resulting map revealed that rates of recombination and architecture of chromosomes are differed between rice and sorghum (KIM *et al.* 2005a). Among the collinear regions, the average size of euchromatin region in sorghum is 1.8-fold larger than the one in rice and the size of pericentromeric heterochromatin in sorghum is 3.6-fold larger than the one in rice. However, the recombination rate in the heterochromatin of sorghum was much lower (6.4 to 36.3-fold) than the homologous regions in rice. This difference may caused by the higher repetitive DNA content in sorghum than in rice, especially in heterochromatic regions.

The overall collinearity between potato and tomato was also demonstrated by FISH mapping potato BAC clones on tomato pachytene chromosome spreads (IOVENE *et al.* 2008). The FISH-based approach revealed a chromosomal inversion in the euchromatin of the short arm of potato chromosome 6 compared to the same region of tomato chromosome 6.

In addition to chromosomal rearrangement between species, intragenomic duplication can also be analyzed by FISH. Seven soybean BAC clones gave two pairs of signals on two different pachytene chromosomes, one on chromosome 19 and another on an unidentified chromosome that shared synteny with chromosome 19. This data indicates a small chromosomal segment duplicated within the paleopolyploid soybean (WALLING *et al.* 2006).

To compare the composition difference between euchromatin and pericentromeric heterochromatin, BACs containing fragments in euchromatin and pericentromeric heterochromatin were identified by direct visualization of BACs on chromosomes using FISH and then sequenced. Eleven sequenced tomato BAC clones were hybridized to pachytene spreads for localization. BAC clones located in euchromatin contained fewer transposons and greater non-transposable element gene density than did the BAC clones that located in the heterochromatin (WANG *et al.* 2006b). Detailed genomic composition of euchromatin and heterochromatin on tomato chromosome 6 was similarly studied by positioning 139 sequenced BAC clones using FISH (PETERS *et al.* 2009). Surprisingly. high gene content was discovered in the pericentromeric heterochromatic regions on this chromosome with uneven repetitive sequence distribution on the short and long arms.

2.5.4 Complementary tools of genome sequencing

As the cost of whole genome sequencing declines, more plant species are being sequenced. From 2008 to early 2011, genome sequences of eight crops have been published (FEUILLET *et al.* 2011). However, gaps remained on physical maps due to tandem repeat sequences, especially at the centromeric and telomeric regions. FISH can be a useful tool for positioning unknown BAC clones and resolving the order of unassembled BAC clones.

The tomato genome (2x =2n = 24; 917 Mbp) is currently being sequenced using BAC-by-BAC approach. At last report, 77% of its genome is highly repetitive heterochromatin (PETERSON *et al.* 1996). Ninety percent of tomato's genes are located in the 23% euchromatin which is equivalent to 211 Mbp, slightly larger than the entire arabidopsis genome. Thus, the tomato genome sequencing consortium decided to focus on sequencing euchromatic regions only. To achieve this goal, it was necessary to position the BAC clones by FISH and then select which BAC clones to sequence. FISH also helps locating BAC clones near the boarders of telomeres and pericentromeric regions so that the extension of contigs could be terminated. Some BAC clones with high

repeat content created difficulty in sequence assembling. Stack and colleague (2009) studied the location of 201 “problem” BAC clones on 12 pairs of chromosomes by FISH with the aid of C_ot-1 DNA. 147 (73%) of the BAC clones were successfully placed on specific chromosomal locations.

A similar FISH approach was used for tomato chromosome 6 to confirm the cytological position of 75 candidate seed BAC clones (SZINAY *et al.* 2008). The seed BAC clones were retrieved from libraries with the aid of genetic markers for chromosome 6. Among the 75 BAC clones, 51 of them were located on the euchromatin of chromosome 6. These 51 BAC clones showed high chromosomal coverage. Nevertheless, three large gaps remaining in the short arm were not covered by any BAC clone and five gaps were observed in long arm. The gaps could be closed by BAC walking or additional genetic markers around the gap regions.

The 51 tomato BAC clones were used to build the backbone for sequencing chromosome 6 and additional BAC clones were added to increase chromosome coverage. The location and order of the BAC clones were verified by FISH on pachytene chromosomes. In total, 139 sequenced BAC clones were assembled into 25 BAC contigs and 8 singleton contigs which covered 29% and 6% of euchromatin and heterochromatin, respectively (PETERS *et al.* 2009). The genome organization of euchromatin and heterochromatin were compared in detail based on BAC clone sequences. Uneven distribution of retrotransposons along chromosome 6 and an unexpected high gene density in heterochromatin were revealed.

Another important application of FISH in genome research is estimation of the physical sizes of gaps. Jackson et al. (1998) demonstrated an accurate estimation of physical size of two gaps on *Arabidopsis thaliana* chromosome 2 by hybridizing BAC clones adjacent to the gaps on DNA fibers. The sizes of seven gaps along rice chromosome 4 were also estimated on pachytene chromosomes using similar methods (FENG *et al.* 2002). In the positional cloning of the tomato *jointless-2* gene, the locus was flanked by two genetic markers, 2.3 cM and 0.1 cM on either side of the locus. Based on the measurement of chromosomal positions of markers by FISH, the physical distance between the two

markers was estimated to be ~60 Mb. Due to the large physical distance involved, chromosomal walking would not be the ideal method for positioning cloning of the *jointless-2* gene. Consequently, the authors switched to genome-wide BAC-based physical mapping to identify putative genes on this locus (BUDIMAN *et al.* 2004).

Chapter 3 Hypotheses and Objectives

3.1 Hypotheses

Fluorescence in situ hybridization (FISH) has been proved to be useful for chromosome identification and in revealing relationships between physical and genetic maps. Papaya chromosomes are small and look similar and thus require a sensitive and reproducible method for karyotyping. In addition to integrating cytological and genetic maps, FISH can be used to align different genetic elements along papaya chromosomes. We will prove this concept on the papaya chromosome 1, which carries the sex-determining region. Applying FISH technology with previously published genetic maps and DNA sequences, we will test the following hypotheses:

1) The nine major linkage groups of papaya SSR genetic map are equivalent to nine chromosomes. The three minor linkage groups were disconnected from the major linkage groups due to insufficient markers in gap regions.
2) Genomic compositions (gene density and repetitive elements) and chromosomal positions, contribute to variation in recombination rate across the genome.

3.2 Objectives

1) Develop a set of chromosome-specific cytogenetic markers to identify the nine pairs of papaya chromosomes
2) Assign the linkage groups of papaya SSR genetic map to individual chromosomes by FISH
3) Position ribosomal DNAs on genetic map by FISH using chromosome-specific cytogenetic markers
4) Construct a molecular cytogenetic map of papaya chromosome 1
5) Study the relationship of recombination rate variation and genomic contexts of papaya chromosome 1

Chapter 4 Development of chromosome-specific cytogenetic markers and mapping of linkage groups to chromosomes in papaya

4.1 Abstract

Carica papaya L. is a tropical and sub-tropical fruit-tree crop with a small genome and nine pairs of chromosomes. The transgenic cultivar SunUp has been sequenced and three high-density genetic maps are available for mapping agronomically and economically-important traits. However, the small size and similar morphology of papaya chromosomes hinder their identification and few genetic resources are available for integration of genetic and cytogenetic information. Fluorescence in situ hybridization (FISH) was performed on mitotic metaphase chromosomes using BAC clones harboring mapped simple sequence repeat (SSR) markers as probes. A total of 104 BAC clones covering all 12 linkage groups (LGs) were tested and 12 of them that gave a single specific signal were chosen as representative of the 12 LGs of the simple sequence repeat (SSR) genetic map. This set of chromosome-specific DNA markers acted as a foundation for papaya chromosome karyotyping and re-assigning orientation of mapped LGs. Chromosome-specific markers allowed us to assign the minor LGs 10, 11, and 12 to major LGs 8, 9, and 7, respectively. We thus reduced the number of LGs in the genetic map to nine, corresponding to the haploid number of papaya chromosomes. We also tested the relative order of DNA markers on minor LGs 10 and 11 to place them on top of LGs 8 and 9 in the correct orientation. Ribosomal DNAs, a set of major cytogenetic markers, were positioned on specific papaya chromosomes. The 25S rDNA showed strong signals at the constriction site of a single pair of chromosomes identified as containing LG 2 by the colocalization of LG 2-specific BAC clones. The 5S rDNA showed strong signals on two pairs of chromosomes that are syntenic with LG 4- and LG 5. This integrated map will facilitate genome assembly, quantitative trait locus (QTL) mapping, and to the study the relationships of cytological, physical and genetic distance of papaya chromosomes.

4.2 Introduction

Cytogenetics was used before the discovery of DNA structure (WATSON and CRICK 1953) to correlate visibly distinct chromosomal segments with a number of genetic traits in maize (CREIGHTON and MCCLINTOCK 1931; MCCLINTOCK 1931; MCCLINTOCK and HILL 1931). Chromosome identification allowed scientists to study recombination events and the spatial arrangement of each chromosome (LYSAK *et al.* 2001) within and between species (HIZUME *et al.* 2002; LIM *et al.* 2000). Traditionally, chromosome banding techniques (e.g. C- and G-banding) and morphological characteristics (e.g. chromosome size, arm ratio) played important roles in chromosome identification.

Repetitive sequences such as microsatellites or simple sequence repeats (SSRs) and ribosomal DNA (rDNA), have been used as landmarks for chromosome karyotyping (ISLAM-FARIDI *et al.* 2007; REN *et al.* 2009; SZINAY *et al.* 2008). Due to its conserved sequences in plant, rDNA has primarily been used with data on chromosome lengths and arm ratios to characterize specific chromosomes (ISLAM-FARIDI *et al.* 2007; SZINAY *et al.* 2008; WANG *et al.* 2008). The tandem array of 45S rDNA forms the nucleolus organizing region (SALANOUBAT *et al.*) associated with the nucleolus while 5S rDNA is a component of large subunit of ribosome for translation. Although rDNA sequences are conserved, the number and position of rDNA sequences vary among species or cultivars (CHUNG *et al.* 2008; KOORNNEEF *et al.* 2003; ROSATO *et al.* 2008). The cytogenetic location of ribosomal DNA combined with geographical co-distribution, predicted the A genome ancestor of tetraploid *Arachis* specie (ROBLEDO *et al.* 2009). This method also facilitated the study of polyploidization of *Glycine* species at the chromosomal level (KRISHNAN *et al.* 2001).

Chromosome identification in plants gives a broad view of chromatin structure organization but detailed information on gene ordering and recombination can only be provided by genetic maps. The DNA markers, such as amplified fragment length polymorphism (AFLP), single nucleotide polymorphism (SNPs) and sequence-tagged microsatellites, are commonly used in genetic maps. These are important resources for

mapping quantitative trait loci (QTL) controlling agronomic traits (COOPER *et al.* 2009; JANSEN *et al.* 2009) and they facilitate genome assembly (IRGSP 2005; MING *et al.* 2008; PATERSON *et al.* 2009; PROJECT 2005). Integrated genetic and cytogenetic maps have been produced in plants in an attempt to assign LGs to chromosomes and to elucidate the relationship between different types of chromosome structures and recombination rates for genetic maps (IOVENE *et al.* 2008; KIM *et al.* 2005a; PETERS *et al.* 2009; WANG *et al.* 2006b).

Currently, there are two main approaches for integrating genetic and cytogenetic maps: (1) DNA markers are physically mapped to chromosome regions in deletion or translocation stocks; and (2) large-insert clones with DNA markers are assigned directly to chromosomes by FISH. The former method has been used mainly in the analysis of plant species having a large genome size and established cytogenetic stocks such as barley (KUNZEL *et al.* 2000), maize (CHAO *et al.* 1996) and wheat (DIEGUEZ *et al.* 2006; HOHMANN *et al.* 1994; MILLA and GUSTAFSON 2001; SAINTENAC *et al.* 2009). Chromosome morphological differences in deletion or translocation lines can be distinguished easily in these species due to the large size of their chromosomes. Chromosome arms of maize can be distinguished by the pattern of C-banding, G-banding (KIM *et al.* 2009) and HKG-banding (DE CARVALHO 1997). These patterns facilitate the linkage of phenotypic traits to chromosome segments.

However, plant species with small chromosomes, such as *Arabidopsis*, papaya and potato (KATO *et al.* 2005), cannot be analyzed visually, but must depend on newer methods based on large-insert clone libraries (such as BAC libraries). Storey (1953) reported that papaya chromosomes (2n = 18) are similar in size during metaphase and they appear to be metacentric or submetacentric (C. M. Wai, R. Ming, P. H. Moore, R. E. Paull, Q. Yu, unpublished). Recently, high resolution papaya pachytene spreads showed that the distal ends of chromosomes are mostly euchromatic (MING *et al.* 2008; ZHANG *et al.* 2008) with centromeric heterochromatic block. The small size and chromosome morphological uniformity makes it difficult to identify individual papaya chromosomes based on banding pattern, arm ratio, or chromosome length. With the development of fluorescence

in situ hybridization (FISH), chromosome identification became more versatile and accurate by using repetitive sequences and large-insert clones as probes (JIANG *et al.* 1995; LENGEROVA *et al.* 2004). FISH gives reliable and routine results regardless of the chromosome size and the quality of chromosome preparations (JIANG and GILL 2006). For example, the 12 pairs of *Silene latifolia* chromosomes were distinguished by simultaneously hybridizing mitotic chromosomes with three repetitive sequences and five BAC clones (LENGEROVA *et al.* 2004). A set of chromosome-specific BAC clones were developed in potato and have been used for mapping the chromosomal position of a potato late blight resistance gene (DONG *et al.* 2000). Chromosome identification using chromosome-specific cytogenetic DNA markers have been demonstrated in rice (CHENG *et al.* 2001a), cotton (WANG *et al.* 2007), sorghum (KIM *et al.* 2002), tomato (TANG *et al.* 2009b), *Antirrhinum majus* (ZHANG *et al.* 2005), potato (DONG *et al.* 2000) and common bean (PEDROSA *et al.* 2003). The cytogenetic markers used in these species have been mainly developed from BAC clones containing microsatellite or AFLP genetic markers.

Recently, a draft genome sequence of transgenic papaya SunUp cultivar was released (MING *et al.* 2008) . Genomic resources, such as BAC libraries (MING *et al.* 2001a), EST sequences and three high-density genetic maps (BLAS *et al.* 2009; CHEN *et al.* 2007; MA *et al.* 2004) are also available. A high-density sequence-tagged genetic map of papaya was constructed using 706 microsatellite (or simple sequence repeat, SSR) markers and the one morphological marker of fruit color (CHEN *et al.* 2007). However, due to variation in the recombination rate along chromosomes, gaps remained that complicated the assembling of genetic markers on specific chromosomes. In this SSR genetic map, nine major linkage groups (LG 1-9), which were hypothesized to represent the nine pairs of papaya chromosomes, and three minor LGs (LG 10-12) were constructed with a total length of 1069.9 cM. In attempts to increase the genome coverage of this genetic map, an additional 277 AFLP markers were integrated into the SSR genetic map but these additions were still unable to bridge the gaps between the major and minor LGs (BLAS *et al.* 2009). The resulting incomplete genetic map could confound positional cloning genes of interest and QTL mapping of important traits.

Chromosome-specific cytogenetic markers derived from the existing SSR genetic map could be used to identify individual chromosomes with FISH techniques. Since the cytogenetic markers are sequence-tagged, they will facilitate the physical assignment of the genetic map on chromosomes. This set of cytogenetic markers can be used to assign the three minor LGs of SSR genetic map to the nine major ones as it allows direct visualization of FISH signals on chromosomes (JIANG *et al.* 1995). A complete assembled genetic map will facilitate the papaya genome assembly, enhance the capacity for positional cloning of genes with important agronomic value, contribute to the comparative genomics between papaya and other Brassicales species and allow for the study of relationships between chromatin structure and recombination rate along the chromosomes.

4.3 Methods

4.3.1 Plant materials

The papaya cultivar SunUp was germinated and planted at Hawaii Agriculture Research Center Kunia Substation (Kunia, HI, USA). Mature trees were grown in the field and hermaphrodite flower buds 6-8 mm in length were collected for pachytene chromosome preparation. Shoot apical meristems of six to eight-week old seedlings were harvested for mitotic metaphase chromosomes preparation.

4.3.2 Chromosome preparation

Immature flowers (6-8mm in length) were collected from hermaphrodite trees and fixed in 3:1 (100% ethanol: glacial acetic acid) Carnoy's solution. Microsporocytes at the pachytene stage were squashed in 60% acetic acid on Fisherbrand Superfrost Plus® slides (Fisher scientific, PA). Glass coverslips were removed with a razor blade after immersing slides in liquid nitrogen. The chromosome slides were stored at 4ºC until hybridization and analysis.

Mitotic metaphase chromosomes were prepared from shoot apical meristems (SAMs) of 6 to 8-week old plants. SAMs were treated with saturated 1,4-dichlorobenzene and 1-bromonaphthalene for 2 to 3 hrs to accumulate prometaphase and metaphase cells, and then fixed in Carnoy's solution. Samples were then squashed on slides in 60% acetic acid and air-dried. All slides were stored at 4º until used.

4.3.3 Probe selection and labeling

Microsatellite marker-anchored BAC clones of each linkage group were chosen for testing signal specificity in FISH studies. Five to 16 markers were selected and tested for each of the nine major LGs, while one to three markers were tested for each of on the three minor LGs. The microsatellite markers were selected based on papaya SSR genetic

map (CHEN *et al.* 2007) and then used as probes to screen the 'SunUp' hermaphrodite BAC library (MING *et al.* 2001b) to obtain BAC clones containing mapped microsatellite markers for FISH analysis. BAC DNA was isolated using the Qiagen plasmid midi kit (QIAGEN, CA) according to the manufacturer's instructions, and then labeled with either digoxigenin (DIG)-11-dUTP, or biotin-16-dUTP using DIG-, or Biotin-nick translation mix (Roche Diagnostics, IN).

4.3.4 Molecular cloning of ribosomal DNAs

Ribosomal DNA sequences in the papaya genome were searched by using *Arabidopsis thaliana* 45S rDNA sequences (Accession: X52320) as query. Primers were designed on putative rDNA in the papaya supercontig 0 sequence (2,423,715-2,424,400) by Primer 3 program (ROZEN and SKALETSKY 2000) to amplify partial 25S rDNA sequence. The sequence of the forward primer was 5' ctc ccc att cga ccc gtc 3' and the reverse primer was 5' gcc atc att ttc ggg g 3'. The PCR reactions consisting of 1X Taq polymerase Buffer, 1 unit EconoTaq® polymerase (Lucigen, WI), 10ng 'SunUp' genomic DNA, 0.2mM dNTPs, 0.2μM forward primer and reverse primer were denatured at 95°C for 5 min, a 30 cycle of 95°C for 30 sec, 55°C for 30 sec and 72°C for 1 min, followed by 72°C for 7 min and stored at 4°C. The expected size band of 685bp was eluted from agarose gel by Illustra GFX™ PCR DNA and gel band purification kit (GE Healthcare, PA) before being labeled by nick translation.

Papaya genome sequence was searched using *Arabidopsis thaliana* partial 5S rDNA sequence (Accession: AY130728.1) as query. The 5S rDNA sequence was amplified by specific primer (5' aac ctg cgg acc ata att taa 3' and 5' gta gta cta gga tgg gtg ac 3') with same reaction conditions as used for 25S rDNA. The expected band size of 140bp was cloned into pGEM-T® vector (Invitrogen, CA) by *E. coli* (BL21) transformation. Transformed plasmid was amplified in *E. coli* and 1μg was used for probe labeling.

4.3.5 Fluorescence in situ hybridization (FISH)

Mitotic and meiotic chromosomes slides were pretreated with RNase (100µg/mL) at 37°C for 1 hr, followed by pepsin (5µg/mL) at 37°C for 15 min and 4% paraformaldehyde at room temperature for 10 min. After washing three times at 5 min interval with 2X saline-sodium citrate (SSC), slides were denatured in 70% deionized formamide at 80°C for 2 min and immediately dehydrated in an cold ethanol series (70%, 90% and 95%) for 2 min each. Hybridization mixtures consisted of 50% formamide, 2X SSC, 10% dextran sulfate, 400ng papaya C_0t-1 DNA, 100-250ng DIG-11-dUTP and/or biotin-16-dUTP labeled BAC. The DNA preparation was denatured at 80°C for 5 min before applying it to chromosome preparations. The slides were incubated in OmniSlide thermal cycler (Thermo Electron Co., MA) at 37°C for 18 hr. DIG- and biotin-labeled BAC clones were detected using 10ng/mL anti-DIG-fluorescein antibody (Roche Diagnostics, IN) and streptavidin Alexa Fluor® 568-conjugated antibody (Invitrogen, SD), respectively. Chromosomal DNA was counterstained with 4, 6-diamidino-2-phenylindole (1ug/mL) in antifade Vectorshield solution (Vector Laboratories) and covered with a glass coverslip.

4.3.6 Cytological analysis

Images were captured by 2.0 megapixels monochrome QuantiFire® camera (Optronics) attached to an Olympus BX51 epifluorescence microscope. Grey-scale images on each color channel were captured, pseudo-colored and merged by PictureFrame™ 2.2 image system (Optronics).

4.4 Results

4.4.1 Development of chromosome-specific cytogenetic markers

The sequence-tagged high density genetic map generated from an 'AU9' x 'SunUp' F2 population was used to identify markers that might be useful for tagging of individual chromosomes (CHEN *et al.* 2007). A total of 104 microsatellite markers were selected from this genetic map that covered all 12 LGs. With assistance from the integrated map (YU *et al.* 2009), the BAC clones harboring these microsatellite markers were identified. Most microsatellite markers were chosen from non-clustering regions or at the distal ends of each LG to avoid any BAC clones containing highly abundant repetitive sequences, which were mostly confined to marker-clustered proximal regions of the LGs. BAC DNA was isolated and hybridized on mitotic metaphase chromosomes with the aid of C_ot-1 DNA to suppress background signals produced from repetitive sequences. Among the 104 BAC clones selected, 16 of them were located at genetic positions with a cluster of five or more markers; while the remaining 88 BAC clones were located at genetic positions having fewer than five markers. A total of 54 BAC clones gave specific FISH signals on only one pair of chromosomes. Five of them were located at genetic positions with a cluster of five or more markers; while the remaining 44 clones that gave unambiguous signals were located at genetic position with fewer than five markers (Table 4.1).

Only BAC clones giving reproducible and unambiguous FISH signals were chosen as chromosome-specific markers. Among the 55 BAC clones showing unambiguous FISH signals, 12 were chosen as the most reliable representations of LG-1 to LG-12 of SSR genetic map (Fig. 4.1). The genetic position of 12 LG-specific BAC clones were located primarily near, or at, the distal ends of the genetic map (Table 4.2). Notable exceptions to this observation were with LG 8 and LG 9 where the BAC clones were located nearer the middle of the LGs.

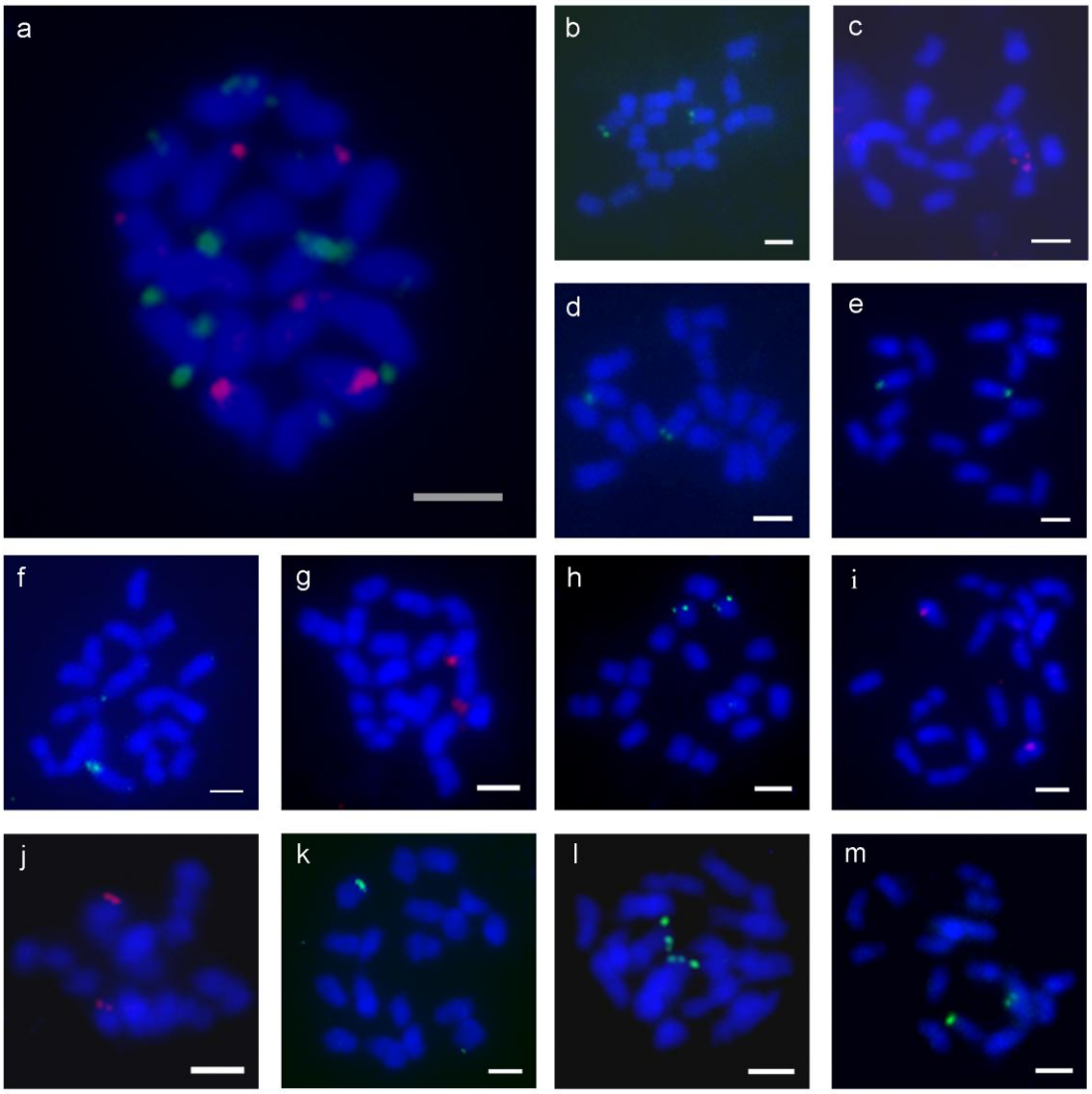

Fig. 4.1 Papaya chromosomes labeled with linkage group-specific BAC clones

(a) Papaya chromosomes simultaneously hybridized with nine BAC clones that represent the major LGs 1 to 9. BAC clones 15O14, 28B18, and 78D03 specific for LGs 2, 6 and 8, respectively, were labeled red with biotin. BAC clones 96C17, 7H21, 43N18, 57E17, 12M21, and 39C20 specific for LGs 1, 3, 4, 5, 7 and 9 were labeled green with DIG. Each LG-specific BAC clone showed signal on only a single pair of chromosomes. (b-m) Papaya chromosomes hybridized with single LG-specific BAC clones. (b) Papaya chromosomes hybridized with LG 1- specific BAC clone 96C17; (c) Papaya chromosomes hybridized with LG 2-specific BAC clone 15O14; (d) Papaya

chromosomes hybridized with LG 3-specific BAC clone 7H21; (e) Papaya chromosome hybridized with LG 4-specific BAC clone 43N18; (f) Papaya chromosomes hybridized with LG 5-specific BAC clone 57E17; (g) Papaya chromosomes hybridized with LG 6-specific BAC clone 28B18; (h) Papaya chromosomes hybridized with LG 7-specific BAC clone 12M21; (i) Papaya chromosomes hybridized with LG 8-specific BAC clone 78D03; (j) Papaya chromosomes hybridized with LG 9-specific BAC clone 39C20; (k) Papaya chromosomes hybridized with LG 10-specific BAC clone 99D21; (l) Papaya chromosomes hybridized with LG 11-specific BAC clone 39P03; and (m) Papaya chromosomes hybridized with LG 12-specific BAC clone 1P02. Bar, 2μm.

Table 4.1 Relationship between clustering of microsatellite markers and specificity on FISH

	No. of marker	Specificity on FISH		
		***	**	*
Marker clustered with <5 markers at same genetic position	88	49	14	25
Marker clustered with ≥ 5 markers at same genetic position	16	5	5	6
Total	104	54	19	31

*** Specific single signals on one pair of chromosomes

** Major signal on one pair of chromosomes with weak background on other chromosomes

* Strong non-specific signal on more than one pair of chromosomes

Table 4.2 List of papaya chromosome-specific BAC clones

Linkage group	Total length of linkage group (cM)	Selected microsatellite marker	Genetic position[a] (cM)	BAC clone containing microsatellite marker
1	145.0	CPM1737	112.6	96C17
2	138.8	CPM976	129.6	15O14
3	132.4	CPM746	0	7H21
4	120.6	CPM1111	91.0	43N18
5	103.6	CPM1657	20.7	57E17
6	100.2	CPM1010	3.4	28B18
7	96.4	CPM1846	27.9	12M21
8	91.8	CPM1564	46.9	78D03
9	64.5	CPM584/1129	31.2	39C20
10	27.1	CPM2098	0	99D21
11	26.8	CPM1125	26.6	39P03
12	21.4	P3K3510	0	1P02

[a] The genetic position of microsatellite marker on corresponding linkage group according to SSR genetic map (CHEN *et al.*).

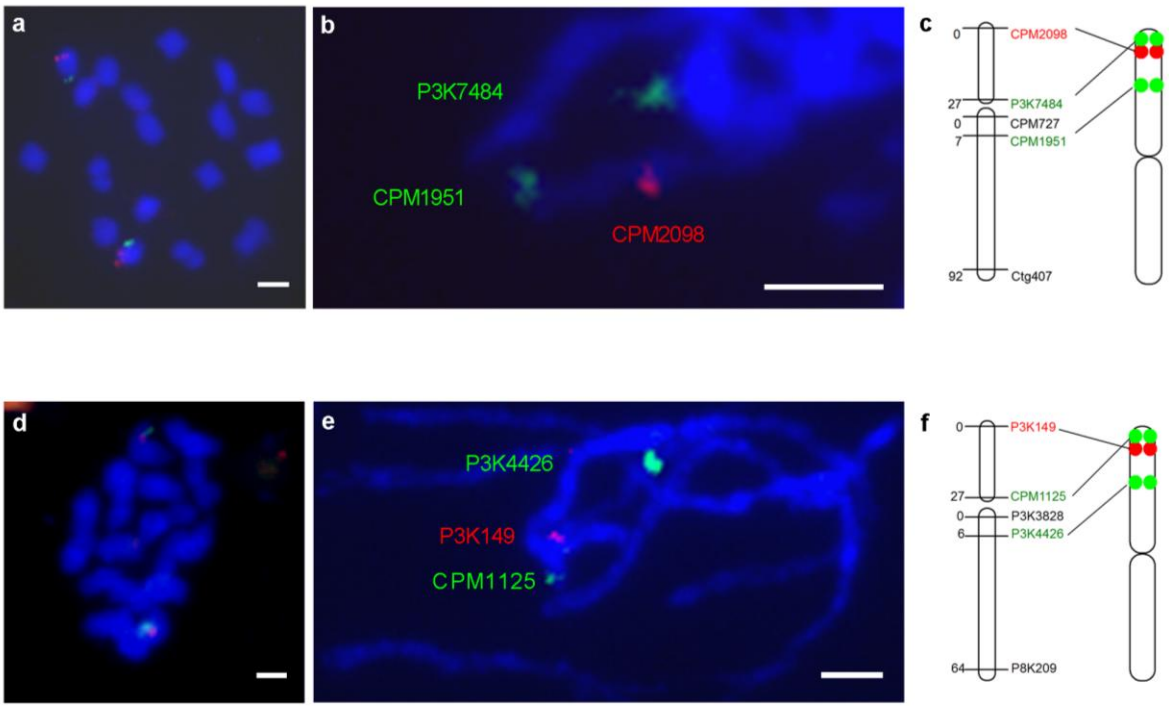

Fig. 4.2 Assignment of minor LGs to major LGs on mitotic and meiotic chromosomes (a) LG 8 BAC clone 35I09 (red) and LG 10 BAC clone 99D21 (green) colocalized on the same metaphase chromosome by FISH. (b) The order of LG 8 BAC clone 51E11 harboring microsatellite marker CPM1951 (red) and two LG 10 BAC clones, 25C09 (P3K7484, green) and 99D21 (CPM2098, red) were revealed on pachytene chromosome. (c) The genetic and cytological positions of LG 8 marker CPM1951 (red) and two LG 10 markers P3K7484 (green) and CPM2098 (red) are presented diagrammatically (not to scale). (d) LG 9 BAC clone 67F13 (red) and LG 11 BAC clone 39B03 (green) colocalized at the distal end of a single pair of metaphase chromosomes. (e) The order of LG 9 BAC clone 87B15 (P3K4426, green) and two LG 11 BAC clones, 39P03 (CPM1125, green) and 09J08 (P3K149, red) are shown on pachytene chromosome. (f) The corresponding genetic and cytological positions of LG 9 marker P3K4426 (green) and two LG 11 markers CPM1125 (green) and P3K149 (red) were indicated on diagrammatically (not to scale). Bar, 2μm.

Nine BAC clones, corresponding to the nine major LGs, were simultaneously hybridized on mitotic chromosome of a single nucleus in an attempt to integrate the genetic map LGs with the chromosomes. Each BAC clone signal was observed on a single chromosome, suggesting the nine major LGs represent papaya's nine pairs of chromosomes (Fig. 4.1a) and that each of three minor LGs may be linked to different major LGs. All signals observed were near or at the distal ends of chromosomes.

4.4.2 Integration of linkage groups and chromosomes

Once papaya chromosome-specific cytological markers were established, the 12 LGs of the SSR genetic map were assigned to nine pairs of papaya chromosomes. BAC clones from the three minor LGs (10, 11 and 12) were screened with LG-specific cytogenetic markers from each of the nine major LGs (1 through 9) by FISH. This allowed colocalization of LG signals on the same mitotic metaphase chromosome.

By using this strategy, BAC clone 99D21 containing microsatellite marker CPM2098 from LG 10 and BAC clone 35I09 containing microsatellite marker CPM766 from LG 8 were observed as located at the opposite ends of one single pair of chromosomes. Thus, LG 10 was recognized as the north arm of LG 8 (Fig. 4.2a). The orientation of LG10 on top of LG 8 was analyzed by simultaneously hybridizing two BAC clones, 25C09 and 99D21, from LG 10 and BAC clone 51E11 from north arm of LG 8 to a meiotic pachytene chromosome preparation (Fig. 4.2b). LG10 BAC clone 25C09 containing microsatellite marker P3K7484 (27 cM) was observed at the chromosome end and BAC clone 99D21 carrying marker CPM2098 (0 cM) located nearer to the LG 8 BAC 51E11 than did BAC clone 25C09. This observation indicated that LG 10 was linked to the top of LG 8 in an inverted direction (Fig. 4.2c).

Minor LG 11 was merged with another linkage group using the same methods as used for assigning LG 10 to a chromosome. BAC clone 39P03 from LG 11 and BAC clone 67F13 from LG 9 were visualized at the distal end of a single pair of mitotic metaphase chromosomes (Fig. 4.2d). The relative order of microsatellite markers was resolved on

meiotic pachytene chromosome (Fig. 4.2e). Specifically, LG 11 BAC clone 39P03 harboring microsatellite marker CPM1125 (27cM) and BAC clone 09J08 containing marker P3K149 (0cM) were pseudo-colored in red and green, respectively. BAC clone 39P03 in green color was close to the telomeric region of a pachytene chromosome while BAC clone 87B15 containing microsatellite marker P3K4426 (6 cM) of LG 9 in green color was observed adjacent to LG 11 BAC clone 09J08 at the distal end of chromosome. Thus, the inverted minor LG 11 was placed on top of LG9 (Fig. 4.2f).

For the smallest LG, LG 12, FISH on a preparation of mitotic metaphase chromosomes showed that LG 12 BAC clone 01P02 and LG 7 BAC clone 24J25 were located at the opposite ends of the same chromosome. This observation suggested that LG 12 is located at north arm of LG 7 (Fig. 4.3). Due to the small physical distance between the microsatellite markers at the opposite ends of LG 12 and the limited number of chromosome-specific BAC clone available for LG 12, the orientation of LG 12 on the north arm of LG 7 could not be verified with additional markers.

4.4.3 Localization of ribosomal DNAs on genetic map

Arabidopsis thaliana rDNA sequences were used as a query in a nucleotide BLAST search against the papaya genome sequence database to obtain papaya rDNA sequences. Eight hits with an expectation value lower than e^{-10} were obtained for the 25S rDNA. A papaya partial 25S rDNA sequence was cloned and used for localizing the nucleolus organizing region (NOR) (SALANOUBAT *et al.*) on papaya chromosomes. Only one pair of chromosomes was labeled and the signals were located at a constriction site (Fig. 4.4a), where it is close to the middle of the chromosome. The 25S rDNA signals were syntenic to LG 2 BAC clone 15O14, suggesting that the NOR located at the center of LG 2.

More than 100 BLAST hits were obtained using *Arabidopsis thaliana* partial 5S rDNA sequence as the query. The hit with the highest expectation values (supercontig 1275) contained more than 50 tandem repeats of the query sequence. A partial 5S rDNA sequence was cloned from papaya genomic DNA and the plasmid was used as a probe for

FISH. Four strong signals were observed at interstitial positions of two pairs of chromosome (Fig. 4.4b). The 5S rDNA signals were syntenic to LG 4 and 5. On LG 4, the 5S rDNA locus was flanked by BAC clones containing microsatellite markers P3K398 (31 cM) and CPM1024 (48 cM) (Fig. 4.4b). On LG 5, the locus were flanked by BAC clones harboring markers CPM988 (44 cM) and P6K911 (63 cM) (Fig. 4.4c).

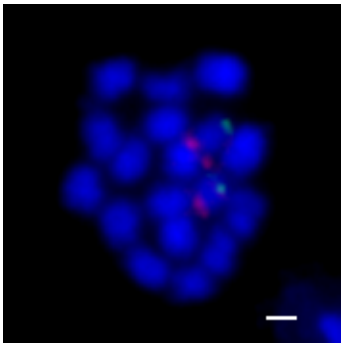

Fig. 4.3 Physical assignment of LG 7 to LG 12 by FISH

LG 7 BAC clone 24J25 (red) and LG 12 BAC clone 01P02 (green) were located at opposite distal end of same chromosome. Bar, 2μm.

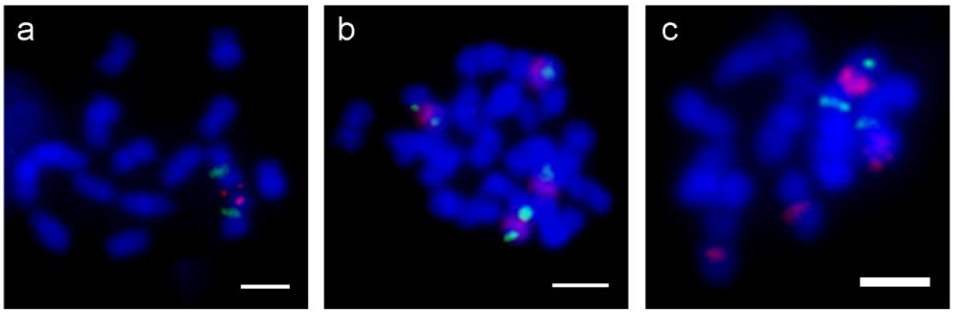

Fig. 4.4 Positioning of ribosomal RNA on genetic map by FISH

Ribosomal RNAs were hybridized with chromosome-specific BAC clones simultaneously on mitotic metaphase chromosomes to reveal their location and occurrence. (a) A single strong signal of 25S rDNA was observed near the middle of LG 2 by simultaneously hybridizing a plasmid carrying partial 25S rDNA labeled green and a LG 2-specific BAC clone 15O14 labeled red. (b) A partial 5S rDNA sequence clone (red) showed four major signals on two pairs of chromosomes. Two LG 4 BAC clones (green), 67N22 and 29G19, hybridized with 5S rDNA (red) on metaphase chromosomes simultaneously. BAC clone 29G19 harboring microsatellite marker P3K398 (31 cM) showed specific signals while four signals were observed from BAC clone 67N22, which contains microsatellite marker CPM1024 (48 cM). The 5S rDNA signal was flanked by two LG 4 BAC clones. (c) The papaya metaphase chromosomes were hybridized simultaneously with the red 5S rDNA clone red and two green LG 5-specific BAC clones green, 60F11 and 17E10. The 5S rDNA signals were located between the two LG 5-specific BAC clones, which carried microsatellite markers CPM988 (44 cM) and P6K911 (63 cM). Bar, 2μm.

4.5 Discussion

A saturated genetic map is essential for assembling the genome sequences into pseudomolecules. Such a map should have the identical number of LGs with the number of chromosomes. The papaya high density genetic map contains 706 SSR markers at a density of 1.9 markers per Mb (CHEN *et al.* 2007). However there were nine major and three minor LGs for the nine pairs of papaya chromosomes. The addition of 277 AFLP markers could not close any of the three gaps (BLAS *et al.* 2009). The inability to bridge the gaps by linkage mapping might be due to highly skewed marker segregation on those minor LGs, or an insufficient number of markers, or a large physical distance between gaps. Using BAC clones containing mapped markers and FISH techniques, the three minor LGs 10, 11, and 12 were assigned to LGs 8, 9, and 7, respectively, consolidating the total number of LGs to nine. This new version of the genetic map will help to finish sequencing the papaya genome in the foreseeable future as sequencing cost continues to decline. In addition, it is a valuable resource for gene tagging and QTL mapping of agronomic traits in genome wide scale.

Meiotic pachytene chromosomes provide considerable higher powers of resolution for differentiating among chromosomes than do mitotic metaphase chromosomes. In *Arabidopsis*, metaphase can resolve probes that are 3Mb apart while pachytene chromosomes can resolves probes that are only 75kb apart (SCHUBERT *et al.* 2001). In maize chromosome 9, the resolution of metaphase chromosomes varied from 3.3 to 8.2Mb while the resolution of pachytene chromosomes was 275kbp, which is 12 times greater than the resolution with metaphase chromosomes (DANILOVA and BIRCHLER 2008). Due to their higher resolving power, pachytene chromosome preparations have been used in FISH to resolve marker order on genetic maps (ISLAM-FARIDI *et al.* 2002; KIM *et al.* 2005b), to construct cytogenetic maps (KOO *et al.* 2008; PETERS *et al.* 2009), and to study chromosomal rearrangements (IOVENE *et al.* 2008; TANG *et al.* 2008; WALLING *et al.* 2006). In the present study, the relative orders of DNA markers of LG 9 with LG 11, and LG 8 with LG 10 were observed on extended meiotic pachytene

chromosomes. These observations enabled us to place the minor LGs on major LGs in the well defined orientation that could not be resolved on mitotic metaphase chromosomes. All three minor LGs were located at the distal end of assigned major LGs.

The possible reasons of the disconnection of the three minor LGs from major LGs are 1) insufficient DNA markers on the SSR genetic map, 2) high repetitive sequence contents on gap regions which do not contain generic SSR marker, 3) Lack of polymorphism between parents at gap regions. To tackle the first possible reason, additional AFLP markers were integrated into existing SSR genetic map, it was still unable to connect any minor LGs to the major one. Besides, in our pachytene FISH results, no highly DAPI stained region was observed between the gaps of LG 9 and 11 nor between the gaps of LG 8 and LG 10. In addition, we had the genomic sequence of the gap between LG 8 and 10 and used it to analyze the gene density and repeat content by RepeatMasker (Smit AFA, Hubley R and Green P, http://www.repeatmasker.org). The repetitive sequence density and gene content at the gap and its surrounding regions were similar (Wai CM, unpublished results). Thus, it is unlikely that inability to map across the gap was due to high repeat content or paucity of genes. Therefore, the disconnection of minor LGs is likely due to lack of polymorphism in the gap regions between the parents. If genes with high selective pressure locate at the gap regions, the sequence will be highly conserved and thus lack of polymorphism between varieties. The genomic sequences at the gap regions will provide information for this hypothesis.

In the artichoke female genetic map, 20 LGs were constructed for 17 pairs of chromosomes. The authors of that paper suggested that the unassigned LGs may due to the lack of polymorphism in the gap regions between the parents, or it might be due to the use of overly stringent parameters in genetic map construction (PORTIS *et al.* 2009). These hypotheses might also explain the appearance of minor LGs in the several papaya genetic maps.

This is the first report of using selected papaya BAC clones for integrating mapped LGs to individual chromosomes. In the present study, the use of chromosome-specific markers

for FISH-based chromosome identification is highly reproducible and does not require any specific stages of chromosome preparation as is required for band staining and arm ratio measurements. This approach can be used at different stages of mitosis or meiosis. In addition, arm measurements so useful with large chromosome plants are unreliable for small chromosomes since chromosome lengths can vary so greatly to confound their nomenclature (KIM *et al.* 2009). Thus, the use of chromosome-specific markers in FISH is the most certain method for papaya chromosome identification.

Papaya sequence-tagged genetic map has been produced from microsatellite markers developed from whole genome shotgun sequences, BAC end sequences and subclones of BAC clones. It was thus straight forward to anchor DNA markers to their corresponding BAC clones. The individual LGs of SSR genetic map could be distinguished on mitotic chromosomes by using only 12 LG-specific BAC clones as probes. As suggested by Chen *et al.* (2007), our FISH result confirmed that the nine longest LGs represent the nine pairs of papaya chromosomes.

Microsatellite markers at various regions of the papaya SSR genetic map were chosen and tested for their ability to identify specific papaya chromosomes. Previous studies showed that the distal ends of chromosomes are mainly DAPI light-stained euchromatic regions containing low-repeat DNA (MING *et al.* 2008; ZHANG *et al.* 2008). Such regions might be expected to show highly specific signals on chromosomes in FISH. This hypothetical chromosome structure is consistent with our results showing that 24 of 54 chromosome-specific cytogenetic markers were located at distal ends, where fewer microsatellite markers were clustered. The phenomenon of chromosome specific markers near the chromosome ends has also been observed in rice (CHENG *et al.* 2001a), sorghum (KIM *et al.* 2005a) and *Antirrhinum majus* (ZHANG *et al.* 2005). In sorghum, among 22 markers chosen at the distal end of LGs, 19 showed specific FISH signals and only three showed signals with moderate background without C_ot-1 suppression. Heterochromatic regions have low recombination rates and thus lead to clustering of genetic markers. Heterochromatic regions also contain large amounts of repetitive sequences that produce non-specific signals when using FISH (PETERS *et al.* 2009; WANG *et al.* 2006b).

Surprisingly, the chromosome-specific markers of LG 8 and LG 9, which located in a marker clustering region, also gave unambiguous signals on FISH. This result might be due to small amount of repeat sequences (LAI *et al.* 2006; NAGARAJAN 2008), or low percentage of heterochromatin area in these regions (MING *et al.* 2008). This may also explain why we obtained a higher number of LG-specific signals than expected from 31% of microsatellite markers chosen at genetic positions having clusters of five or more markers.

In this present study, only one NOR site was observed in papaya which is similar to other plants in which one to two NOR sites are reported. For example, one NOR site was reported in soybean (GRIFFOR *et al.* 1991) and two NOR sites were observed in *A. thaliana* (KOORNNEEF *et al.* 2003) and barley (TAKETA *et al.* 1999). In the BLAST search using *Arabidopsis thaliana* 25S rDNA as query against papaya genome sequence, only five hits were obtained with an expectation value less than e^{-10}. This suggested that the copy number of 25S rDNA in papaya genome is either low, or may not be assembled into scaffolds. For the papaya 5S rDNA analyzed in this study, two major sites were observed on metaphase chromosomes while 13 sites were observed on high-resolution pachytene chromosome (Zhang *et al.*, same issue). Our BLAST search showed that more than 100 scaffolds contained multiple copies of the 5S rDNA sequence. This data supported the possibility of a high number of rDNA sites in the papaya genome. The difference in number of 5S rDNA sites observed is likely due to the compaction of chromosome structure at metaphase which leads to decreased resolution. In addition, the two very strong signals observed on LG 4 and 5 may very well mask the weak signals from other sites.

Since ribosomal DNA consists of highly repetitive sequences, it creates difficulties in genome assembly. The ability to identify the exact position of ribosomal DNA in papaya genome enabled us to design different sequencing strategy for those regions to fill the gap. Furthermore, it has been reported that ribosomal DNA number and position can be different among varieties or closely related species (DE MELO and GUERRA 2003; ROSATO *et al.* 2008; SHISHIDO *et al.* 2000). Thus, the NOR and 5S rDNA may be good

landmarks for karyotyping and studying the evolutionary relationship of chromosome structure between papaya and its related species.

The distribution of microsatellite markers in wheat, barley and rye were showed to be specific and correlated with certain chromosomal structures (heterochromatin, euchromatin and centromeres) (CUADRADO *et al.* 2008). With the aid of chromosome-specific cytogenetic markers, we were able to locate ribosomal DNA on the genetic map and will be able to extend this method to study the distribution of other tandem repeat sequences, such as microsatellites and knob-associated repeats, on each chromosome without the complete genome sequence. This finding will enhance the knowledge of papaya genome organization at the chromosomal level.

Chapter 5 Integration of cytogenetic and physical maps by fluorescence *in situ* hybridization reveals uneven genome composition and recombination rates along the papaya sex chromosome

5.1 Abstract

Papaya is a model for analyzing the evolution of sex chromosomes in plants since it possesses a pair of recently evolved sex chromosomes. The Y chromosome is distinguished from the X chromosome by having a small, non-recombining region, characterized as the male specific region of the Y chromosome (MSY), which is not present in the X chromosome. The resources available for the analysis of the papaya sex chromosomes include a draft genome sequence, a sequence-tagged high-density genetic map, and a BAC-based integrated physical map. Nevertheless, the cytological structure and function of the papaya genome remain largely unknown and chromosomal features have not been linked with any genetic or genomic data. A direct way to integrate chromosome morphology with genetic or physical data might be to use fluorescence *in situ* hybridization (FISH) to localize microsatellite marker-tagged BAC clones on individual chromosomes. In the present study, we constructed a cytogenetic map of the papaya Y chromosome (chromosome 1) by hybridizing 15 microsatellite markers and 2 cytological feature-associated markers on pachytene chromosome spreads. This map revealed the physical and genetic distribution of heterochromatic and euchromatic regions of chromosome 1. Analysis of the recombination frequency along chromosome 1, revealed a heterochromatic region with suppressed recombination that occupied about 12.2% of the chromosome, however, the rest of the chromosome exhibited a recombination frequency the same as the genomic average. Although the papaya X chromosome is mostly euchromatic, it contains two heterochromatic knob-like structures that are shared with the Y chromosome. Sequence analysis of assembled contigs of the X chromosome revealed an uneven distribution of genes and transposable elements within the euchromatin and between the euchromatic and centromeric regions. The cytogenetic map of the papaya chromosome 1 produced in this study provides information essential

for the evolutionary analysis of sex chromosomes in the Caricaceae family and will facilitate the assembly of genomic sequences of the papaya sex chromosome.

5.2 Introduction

Papaya, *Carica papaya* L., is a nutritious fruit crop grown in all tropical and many subtropical countries of the world. The papaya plant is a small, herbaceous tree, primarily grown as a garden crop, but also in commercial production for local and export markets. Papaya is trioecious having three sex types (hermaphrodite, female, and male) and it has several properties that facilitate its use in genetic studies. It has a short generation time of approximately 15 months from seed-to-seed and produces numerous fruit containing thousands of seeds. Papaya produces a fruit at approximately ten day intervals twelve months of the year over a period of two to four years. Its large flowers allow for individual hybridization to guarantee parentage for genetic studies. It has a small genome size (372 Mbp) that has been sequenced to a 3X draft (MING *et al.* 2008). Several high-density genetic maps have been produced (BLAS *et al.* 2009; CHEN *et al.* 2007; MA *et al.* 2004) and a BAC-based physical map has been integrated with genome sequence and genetic map (YU *et al.* 2009).

Analysis of amplified fragment length polymorphism (AFLP) and microsatellite marker genetic maps revealed a Y chromosome that contains a small non-recombining region that co-segregates with gender and was characterized as a male-specific region of the Y chromosome (MSY) (LIU *et al.* 2004; MA *et al.* 2004). Further sequence analysis of four X/Y gene pairs located at MSY and the homologous region on the X chromosome revealed that the papaya Y chromosome evolved 0.5-2.2 million years ago (MYA) (YU *et al.* 2008).

In addition to genetic linkage group identification, a nomenclature system was developed for papaya chromosomes based on fluorescence *in situ* hybridization (FISH) (WAI *et al.* 2010; ZHANG *et al.* 2010). The sex chromosome has been designated as chromosome 1

(ZHANG *et al.* 2010) and linkage group (LG) 1 in both the AFLP and microsatellite marker (or simple sequence repeat, SSR) genetic maps (BLAS *et al.* 2009; CHEN *et al.* 2007; MA *et al.* 2004).

According to Zhang *et al.* (2008 and 2010), chromosome 1, with five heterochromatic knobs in the proximal region and one heterochromatic knob close to the distal end, is the second largest chromosome in papaya. The rest of chromosome 1 is stained lightly by DAPI, but one of the chromosome arms is always stained more intensely than the other (ZHANG *et al.* 2010). The five proximal knobs are defined as K1 through K5 and their location at MSY region was confirmed by FISH. Both chromosomes X and Y^h (the hermaphrodite Y chromosome) contain the largest knob K1, while the remaining four knobs appear only on the Y^h chromosome. The greater DNA content of Y^h than of X chromosomes may be responsible for a twist of the Y chromosome between K4 and K5. The DNA associated with K2 to K5 is highly methylated compared to the homologous region on the X chromosome (ZHANG *et al.* 2010), but the reason for differential methylation remains unknown. The centromere of chromosome 1 was shown to be located on or close to K4 (Zhang et al. 2008).

Although the genomic resources of papaya are fairly extensive, the cytological features of papaya chromosomes have not been linked with the existing genome sequence, nor with the genetic and physical maps. The most direct way to integrate the chromosomal features of papaya with its genetic and physical maps would be to generate a cytogenetic map for localizing microsatellite-tagged BACs directly on pachytene chromosomes by FISH. Meiotic pachytene chromosomes are more relaxed than somatic metaphase chromosomes giving them much higher resolving power (CHENG *et al.* 2002; DANILOVA and BIRCHLER 2008; JIANG and GILL 2006). Localization of labeled BACs on pachytene chromosomes to integrate genetics and cytology has been applied in sorghum, tomato, potato, cucumber, rice, maize, and arabidopsis (CHENG *et al.* 2001b; FRANSZ *et al.* 1996; IOVENE *et al.* 2008; KIM *et al.* 2005b; KOO *et al.* 2008; KOUMBARIS and BASS 2003; PEDROSA-HARAND *et al.* 2009; PETERS *et al.* 2009; REN *et al.* 2009; WANG *et al.* 2006a). A molecular cytogenetic map combines chromosome structures with recombination rate

and physical distance, thus providing integrated biological information on genome organization. This information is important in that it allows for an estimation of physical size and distances of trait loci for map-based cloning (BUDIMAN *et al.* 2004; KIM *et al.* 2005b), it enables comparative genomic studies between closely related species (IOVENE *et al.* 2008; KIM *et al.* 2005a), and it contributes to development of genome sequencing strategies (SZINAY *et al.* 2008).

In our previous study, we demonstrated the feasibility of BAC-FISH on high-resolution papaya pachytene chromosome spreads to resolve problems about the orientation of linkage groups (WAI *et al.* 2010). Here, we report the construction of a cytogenetic map of chromosome 1 by labeling 15 genetically mapped microsatellite markers and 2 cytological feature-associated markers on pachytene chromosomes. The recombination frequency along chromosome 1 was analyzed. The data obtained enabled us to analyze the degree of chromosome condensation in selected regions and the relationships between chromosomal morphology, transposable elements and gene content.

5.3 Materials and Methods

5.3.1 Plant materials and chromosome preparation

Carica papaya varieties SunUp and AU9 were planted and maintained at the Hawaii Agriculture Research Center substation located at Kunia, Hawaii. Male flowers (5-7mm in length) were collected and fixed in Carnoy's solution (100% ethanol: glacial acetic acid= 3:1) with 15 min vacuum infiltration. Meiotic pachytene chromosomes were prepared as described in Wai *et al.* (2010). Microscopic slides holding chromosome preparations were pretreated by incubating on 60ºC slide warmer for 30 min, then fixed in 4% paraformaldehyde (in phosphate-buffered saline, pH8.0) for 10 min followed by 2X SSC for 3 times at a interval of 5 min. The slides were then dehydrated in 70% and 95% ethanol for 2 min each.

5.3.2 Probes and fluorescence in situ hybridization

Microsatellite markers previously mapped to LG 1 of papaya SSR genetic maps were chosen randomly with an average distance of 10 cM apart. BAC clones that harbored corresponding microsatellite markers were chosen based on the assembled papaya genome sequence database (version 2007-04-27) available at the Advanced Studies of Genomics, Proteomics and Bioinformatics (ASGPB) homepage (http://asgpb.mhpcc.hawaii.edu). Corresponding BAC clones positions for knob 2 and knob 4 at the MSY region were selected from Zhang *et al.* (2008). All BAC clones used in FISH were obtained from the SunUp hermaphrodite BAC library (MING *et al.* 2001b). The specificity of BAC clones to a single chromosome site was tested on pachytene chromosome spreads before proceeding to cytological measurements. Only BAC clones with a single clear signal were used for further cytological measurement. BAC DNA extraction, probe labeling and fluorescence *in situ* hybridization methods were described in Wai *et al.* (2010), except that C_0t-1 DNA was not added to the hybridization mixture.

5.3.3 Cytological measurement

The distribution of euchromatin and heterochromatin of the short and long arms of chromosome 1 were measured in 18 pachytene chromosome spreads. Short arm euchromatin refers to the region from the telomeric end carrying the microsatellite marker P6K72 to knob 5 of the MSY. Long arm euchromatin refers to the region from knob 1 of the MSY to the telomeric end carrying the microsatellite marker CPM1737.

To measure the cytological position of BAC clones giving unambiguous signals, two to four BAC clones were labeled simultaneously on meiotic pachytene chromosomes. FISH images were captured by a 2.0 megapixels monochrome QuantiFire® camera (Optronics) attached to an Olympus BX51 epifluorescence microscope. Grey-scale images on each color channel were captured and merged by PictureFrame™ 2.2 image system (Optronics) with final adjustment done by Adobe Photoshop 7. The length of chromosome and position of signals from chromosome ends carrying the microsatellite marker P6K72 were measured on merged images by Rincon®. The relative cytological position (CP) was calculated as:

$$CP \text{ (in percentage)} = (S/T) \times 100\%,$$

where S is the distance between signals and telomeric end carrying marker P6K72 and T is the total length of the chromosome. The fraction length (SALANOUBAT *et al.*) was calculated as:

$$FL = (S/T) \times \text{total genetic distance of LG 1},$$

where the total genetic distance of LG 1 is 145 cM. For each marker, a minimum of 10 pachytene chromosome spreads were measured. The measurement data variance was analyzed by Microsoft Excel 2003.

The average cytological distance, i. e. the distance in µm, between the telomere of the short arm and the probe signal was determined on 8 pairs of markers by linear regression analysis. The average chromosome length of 45.35 µm was used to calculate the degree of condensation of the euchromatin and heterochromatin on chromosome 1.

5.3.4 Repetitive sequences and gene density analysis

Scaffold sequences used for analysis of chromosome 1 were based on the physical map published by Yu *et al*. (2009). The sequences are available at ASGPB (http://asgpb.mhpcc.hawaii.edu/papaya/) and only sequences larger than 100 kb were included in this analysis. Transposable elements and repetitive sequences in 18 scaffolds were identified using RepeatMasker version 3.2.8 (Smit, A.F.A and Hubley, R., http://www.repeatmasker.org) and cross match version 0.990329 (http://www.phrap.org) with the combined libraries composed of RepBase library (http://www.girinst.org/repbase, 2009-06-04 version) (JURKA *et al.* 2005), TIGR plant repeat database (OUYANG and BUELL 2004) and *Carica papaya* specific annotated transposable elements database (ftp:/ftp.cbcb.umd.edu/pub/data/CPR-DB) (NAGARAJAN *et al.* 2008).

For putative gene content, the presence of genes on each scaffold was either predicted by a combination of Augustus, Fgenesh, Genscan, GlimmerHMM, SNAP, and TWINSCAN or supported by presence of EST sequences. The predicted gene location and number were accessed through ASGPB homepage (http://asgpb.mhpcc.hawaii.edu/papaya/).

5.4 Results

5.4.1 Morphology of chromosome 1

Under DAPI staining, we observed that the papaya pachytene chromosome 1 was mainly euchromatic, except for the middle part of the chromosome. Five darkly stained knobs were observed at the center of this chromosome. The first knob was the largest in size, and appeared on both chromatids. The other four knobs alternated with euchromatin and were found on only one of the chromatids. In addition to these five knobs, one heterochromatic knob was also observed on the distal end of the euchromatic arm (Fig 5.1a). The observed morphology of chromosome 1 in this study is similar to that reported by Zhang *et al.* (2008 and 2010). Zhang (2008) named the five knob-like structures covering the central region of the MSY as K1, K2, K3, K4, and K5 with K1 as the knob that appears on both the X and Y chromosomes. The chromatid bearing K1 through K5 is recognized as the Y chromosome and the chromatid with only K1 is the X chromosome. FISH labeling of metaphase I bivalent chromosomes showed that the centromere of chromosome 1 lies on or around K4. Thus, the region between K1 and K5 of the Y chromosome and its corresponding region on the X chromosome are defined as centromeric/pericentromeric heterochromatin.

Based on the papaya sequence-tagged genetic map, 16 microsatellite markers randomly distributed on LG1 were selected for cytogenetic mapping (Table 5.1). BAC clones harboring the selected microsatellite markers were identified by searching for microsatellite sequences and papaya BAC end sequences against the papaya genome sequences. BAC clones were labeled on pachytene chromosome spreads by FISH and the distances between specific marker and the telomeric end were measured. The relative cytological position (CP) of each marker was calculated and the results are shown in Table 2. We measured the length of the two arms of chromosome 1, using K4 as the centromere (Zhang et al. 2008). The distance between the telomere end close to marker P6K72 and K4 occupied 42% of the chromosome length and the distance between the other telomere end and K4 is about 58% of the chromosome length. Thus, the

chromosome arm containing marker P6K72 and K5 is defined as the short arm and the other arm is defined as the long arm. Chromosome 1 is sub-metacentric with an arm ratio of 0.71. Based on the measurement of 18 pachytene chromosome spreads, 84.71% ± 2.35% of the Y chromosome is euchromatin and 51% ± 4.6% of the euchromatin lies on the short arm and 49% ± 4.6% of it lies on the long arm (Supplementary file 1). The remaining 15.29% of the Y chromosome is heterochromatic, which locates in the middle of chromosome.

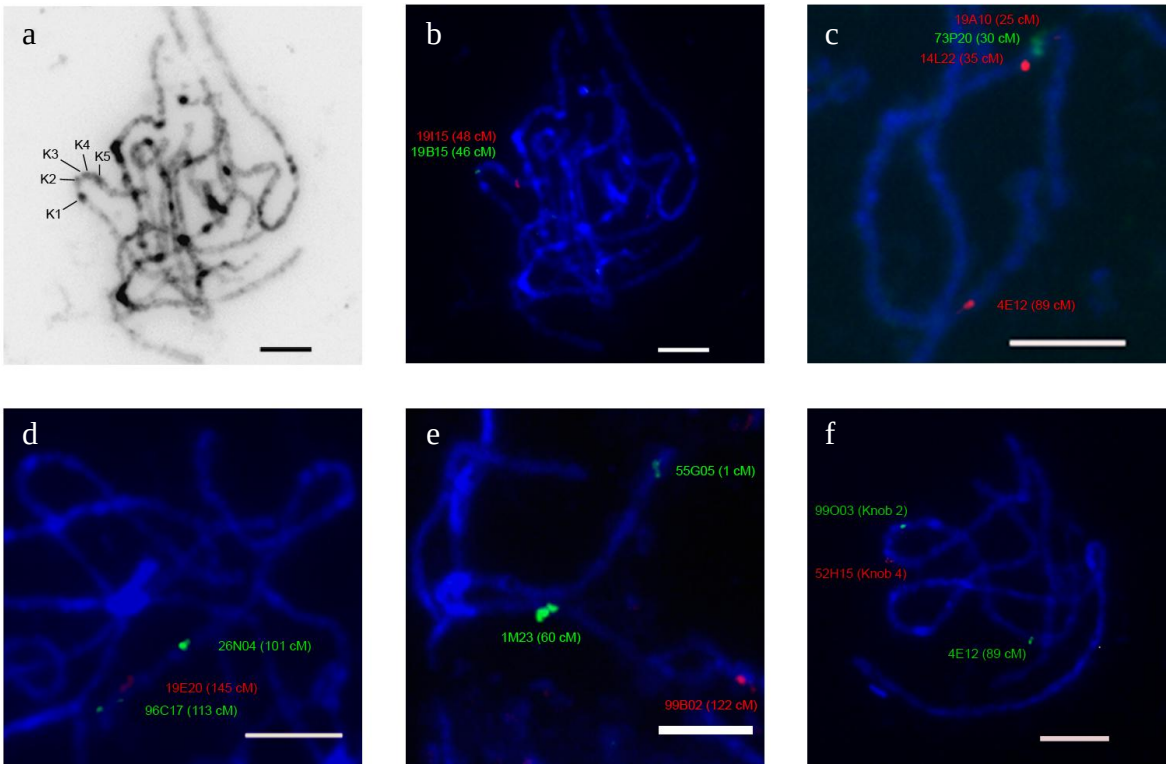

Fig. 5.1 FISH mapping of microsatellite marker-anchored BAC clones on papaya chromosome 1

(a) DAPI-stained pachytene chromosome spread was converted into a black-and-white image. The five heterochromatic knobs on chromosome 1 are labeled K1 through K5. (b) The same chromosome spread as in (Fig. 5.1a) was hybridized with BAC clone 19B15 (46 cM) and 19I15 (48 cM), which spans knobs K2 to K5 of the MSY region on chromosome 1. (c) Simultaneous FISH mapping of four BAC clones, 19A10 (25 cM), 73P20 (30 cM), 14L22 (35 cM) and 4E12 (89 cM) reveals the orientation of the genetically-mapped markers on chromosome 1. (d) BAC clone 19E20 harboring the most distal microsatellite marker ctg566 (145 cM) is positioned between BAC clones 26N04 (101 cM) and 96C17 (113 cM). (e) Simultaneous FISH mapping of three BAC clones, 55G05 (1 cM), 1M23 (60 cM) and 99B02 (122 cM), on pachytene chromosome 1. Most of chromosome 1, except for the telomeres, is spanned between BAC clones 55G05 and 99B02. (f) BAC clones 99O03 and 52H15, associated with knobs K2 and K4, respectively, was co-mapped with BAC clone 4E12 (89 cM). Bars, 5 µm.

Table 5.1 Genetic position of fifteen selected microsatellite markers and two knob-associated markers used for cytogenetic map construction

Microsatellite marker	Genetic position (cM)[a]	BAC clone
P6K72	1	55G05
P6K1066	25	19A10
P3K3585	30	73P20
P3K4292	35	14L22
P3K2981	46	19B15
P6K1460	48	19I15
P3K2388	60	01M23
P3K3051	76	88M12
P3K3143	82	86F03
P3K3750	89	04E12
P3K71	96	13I21
P3K629	101	26N04
CPM1737	113	96C17
P3K418	116	90J10
P3K7101	122	99B02
Ctg566	145	19E20
K2[b]	47	99O03
K4[b]	47	52H15

[a] According to Chen et al. (2007).

[b] K2 and K4 represent knob 2 and knob 4 on MSY region, respectively, and the BAC positions were confirmed by Zhang *et al.*, 2008.

The condensation degree of euchromatin and heterochromatin can vary as a function of the pachytene stages and thus affect the precision of our cytological measurements. We measured the lengths of the euchromatin and centromeric heterochromatin of 18 chromosome spreads to record lengths ranging from 29 – 75 µm (Supplementary file 1). The correlation between the length of the euchromatin and heterochromatin was high and significant ($r = 0.703$, $p \leq 0.01$), suggesting both regions change proportionately when the chromosome length varies in different stages of relaxation. We conclude that pachytene chromosome 1 spreads of different lengths are suitable for constructing the cytogenetic map.

5.4.2 Cytogenetic map of chromosome 1

BAC clones harboring microsatellite markers were chosen and labeled on pachytene chromosomes by FISH to integrate the genetic and cytological maps. In a previous study (Wai et al., 2010), we reported that BAC clones located at genetic regions without marker clustering are likely give informative signals on FISH. We selected 16 microsatellite markers from non-clustering regions on LG 1 (Table 5.1) to construct a cytogenetic map. Two morphological feature associated markers, BAC 99O03 on K2 and BAC 52H15 on K4 were also included to allow us to study the MSY region. The total genetic length of papaya chromosome 1 (LG 1 of the genetic map) is 145 cM (CHEN *et al.* 2007). The selected 16 markers were located between 1 and 145 cM of LG 1. The genetic distance between markers ranged from 5 to 24 cM with an average distance of 9 cM.

The BAC clone signals of 16 markers and 2 knobs that hybridized by FISH were measured on at least 10 pachytene chromosome spreads (Fig. 5.1). FISH results revealed that all the 16 markers except P3K2981 were located in the euchromatic region, four on the short arm, one around the euchromatin-heterochromatin boundary of the short arm, and ten on the long arm. Marker P3K2981 located in the DAPI lightly-stained region between K1 and K2 of centromeric heterochromatin (Figure 5.1a, 5.1b and 5.2). The order of BAC clone signals on chromosome 1 was, concordant with the order of microsatellite markers on the genetic map with three exceptions, P3K2981, P6K1460 and

Ctg566 (Fig. 5.2). On the genetic map, markers P3K2981 and P6K1460 located at 46 and 48 cM, respectively, which is on or adjacent to the MSY region. Note that the position of the BAC clones corresponding to these two markers was on a reversed order on the cytological map (Table 5.2). The FISH signals of these two BAC clones flanked K2 to K5 of the centromeric heterochromatin (Fig. 5.1a and 5.1b). Zhang *et al.* (2008) showed that a BAC clone harboring the marker CPM1055 located at 50cM in the euchromatin of the long arm adjacent to K1. Thus, the five proximal knobs were located between markers P3K2981 (46 cM) and CPM1055 (50 cM) of the genetic map.

The orientation of LG 1 on chromosome 1 was revealed by simultaneously hybridizing multiple BAC clones on pachytene chromosomes. By comparing the FISH signals and DAPI-staining pattern, BAC clone 19A10 containing the marker P6K1066 (25 cM) located at the north arm of LG1, was labeled on the short arm of chromosome 1 (Fig. 5.1c). The BAC clone containing the other most distal end marker on LG 1, Ctg566 (145cM), showed a signal close to the heterochromatic knob on the long arm of chromosome 1 and located between markers P3K629 and CPM1737 (Figure 5.1d). FISH result indicated that marker Ctg566 had been mapped wrongly on the genetic map and the second most distal end SSR genetic marker of LG 1, P3K7101 (122cM) located closer to telomeric end of the chromosome than did the marker Ctg566. About 96% of the length of chromosome 1 was covered by markers P6K72 (1cM) and P3K7101 (122cM) (Figure 5.1e and Table 5.2), indicating a high coverage by the genetic map of this chromosome.

Table 5.2 Genetic and cytological positions of microsatellite and knob-associated markers on chromosome 1

Marker	Genetic position (cM)[a]	Chromosome position	Cytological position		Number of measurements
			(% ± S.D.)[b]	(FL ± S.D.)[c]	
P6K72	1	Short arm	2.69 ± 1.04	3.90 ± 1.50	47
P6K1066	25	Short arm	21.35 ± 0.84	30.96 ± 1.22	11
P3K3585	30	Short arm	24.02 ± 1.12	34.83 ± 1.62	33
P3K4292	35	Short arm	26.65 ± 1.17	38.64 ± 1.70	12
P3K2981	46	Heterochromatin	51.69 ± 1.93	74.73 ± 2.50	11
Knob 4	47	Heterochromatin	41.66 ± 1.71	60.41 ± 4.27	10
Knob 2	47	Heterochromatin	48.53 ± 2.37	70.37 ± 3.43	10
P6K1460	48	Short arm	39.50 ± 2.95	57.76 ± 4.27	11
P3K2388	60	Long arm	67.41 ± 2.64	97.75 ± 3.82	12
P3K3051	76	Long arm	74.40 ± 2.49	107.87 ± 3.60	10
P3K3143	82	Long arm	77.25 ± 2.85	112.01 ± 4.13	10
P3K3750	89	Long arm	79.60 ± 2.73	115.31 ± 3.95	11
P3K71	96	Long arm	83.57 ± 2.96	121.18 ± 4.30	10
P3K629	101	Long arm	86.46 ± 3.15	125.37 ± 4.57	19
CPM1737	113	Long arm	95.02 ± 1.54	137.78 ± 2.23	32
P3K418	116	Long arm	96.20 ± 0.98	139.49 ± 1.42	10
P3K7101	122	Long arm	98.52 ± 1.18	142.52 ± 2.23	34

[a] According to microsatellite marker genetic map (Chen *et al.*, 2007).

[b] Percentage (%) of cytological position is calculated as (distance of signals from end of short arm/ total length of chromosome) X 100%. S. D. represents standard deviation.

[c] Fraction length (FL) of cytological position is calculated as (distance of signals from end of short arm/ total length of chromosome) X total genetic distance of LG 1 (*i.e.* 145cM).

5.4.3 Recombination frequency along chromosome 1

Although a physical map of chromosome 1 is available (YU *et al.* 2009), gaps remained between several scaffolds so that the actual physical distance between some of the genetic markers was unknown. The BAC clones used in FISH are tagged with microsatellite markers so their cytological map can be integrated with a genetic map without additional genetic analyses to compare recombination frequency of different regions along the chromosome.

Along the euchromatin of chromosome 1, the ratio of genetic to cytological distance was nearly constant, ranging from 0.61 to 1.98 cM/ FL (0.51 to 1.64 FL/ cM) (Fig. 5.2, Supplementary file 2). The average recombination frequency between marker P6K72 (1cM) and P3K7101 (122cM) was 0.87 cM/ FL. For the short arm between markers P6K72 and P6K1460, the mean ratio was 0.87±0.07 (47 cM/ 53.86 FL) while for the long arm between markers P3K2388 and P3K7101, the mean ratio was 1.39±0.14 (62 cM/ 44.77 FL), which was higher than the one on the short arm. Cytological distance is an indirect indicator of physical distance, thus, the lower ratio of genetic to cytological distance on the short arm indicated that it has a lower recombination rate than does the long arm.

The centromeric heterochromatic region between markers P6K1460 and P3K2981 showed a severe suppression of recombination. Although this region is about 12.2% (16.97 FL/ 145 FL) of cytological length of the chromosome, it contributes only 1.3% (2 cM/ 145 cM) of the genetic map length. The ratio of genetic to cytological distance in this region is 0.12±0.02 cM/ FL (2 cM/ 16.97 FL) which was 6- to 18-fold lower than that of the euchromatic regions.

5.4.4 Detailed comparison of cytological, genetic and physical distances

A detailed physical picture of chromosome 1 was not possible due to the presence of gaps between scaffolds. Thus, we are unable to discern the relationships of cytological, genetic, and physical distances along the entire chromosome. Instead, we were able to analyze a set of seven pairs of markers and a pair of heterochromatic knobs for which the physical distance information was available to obtain a glimpse of chromosome 1 structure.

The mean chromosome length of 108 measured pachytene spreads was 45.35 ± 9.55 µm. The mean absolute cytological distance of each marker (in µm) was plotted against chromosome length in a linear regression analysis ($p < 0.05$) to predict the position of each marker at given pachytene chromosome lengths.

Aligning the cytogenetic map with the physical map of eight pairs of markers allowed us to calculate the condensation degree along chromosome 1. The condensation in euchromatin showed significant variation ranging from 273 – 806 kb/ µm with a mean of 474 kb/ µm (Table 5.3). The marker pair of P3K418 and CPM1737, which located close to the telomere, were more condensed (806 kb/ µm) than was any other euchromatic region. The condensation degree between K2 and K4 in the centromeric heterochromatin was 1711 kb/ µm, nearly four-fold higher than was the condensation degree of the euchromatin. The two pairs of markers from the short arm (112 and 132 kb/ cM) had a lower recombination frequency than all, but one, pair of markers from the long arm (75 - 103 kb/ cM).

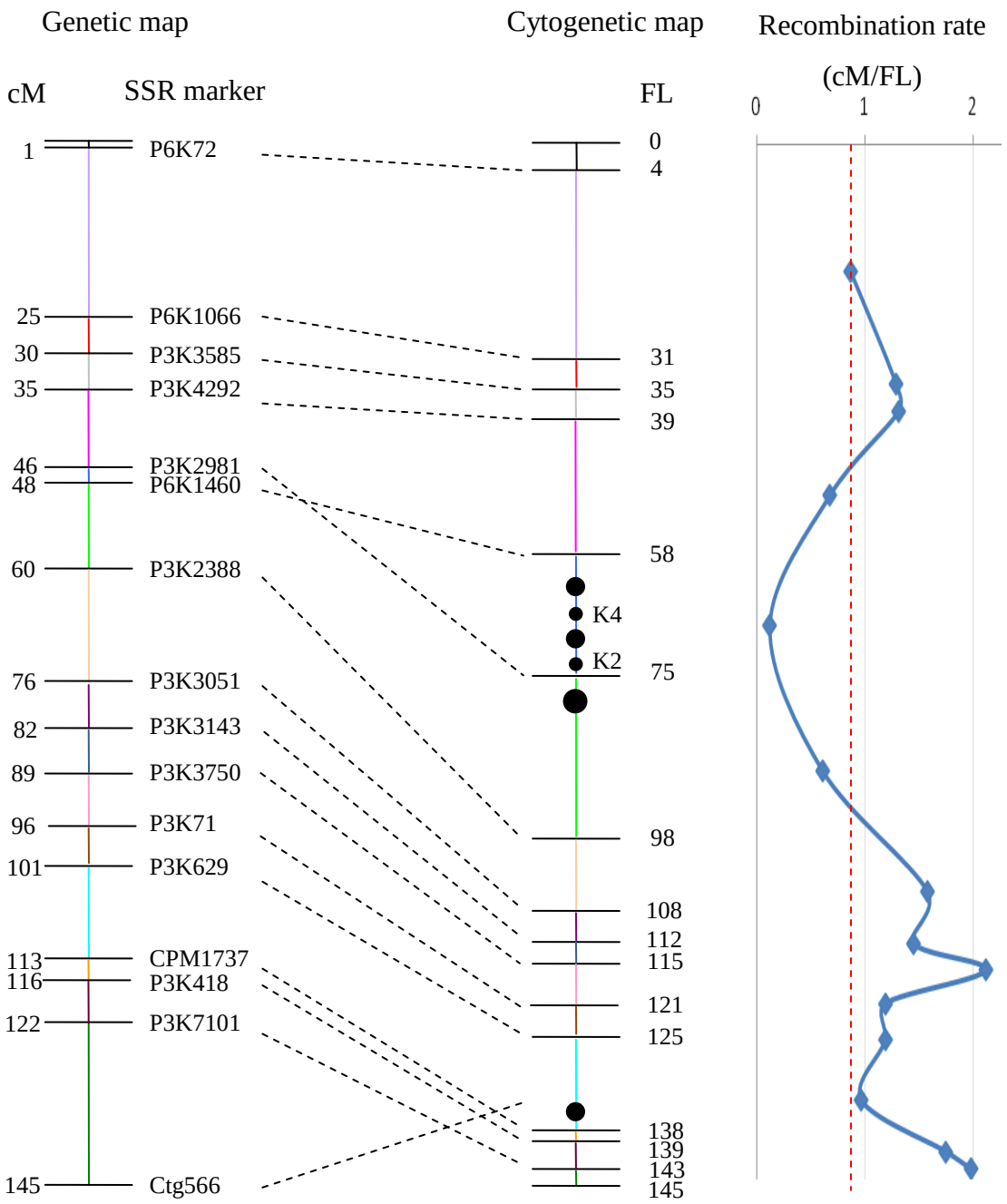

Fig. 5.2 Genetic maps, cytogenetic maps and recombination rate of chromosome 1

The chromosomal positions of 15 microsatellite markers (except Ctg566) were measured by FISH and diagrammed to scale. The genetic positions of markers are based on the microsatellite marker genetic map published by Chen *et al.* (2007). The dark circles on cytogenetic map represent the six heterochromatic knobs on chromosome 1. On the right panel, the recombination frequency was calculated as genetic to cytological distance (cM/

FL, x-axis) and plotted against cytological distance (in fraction length, FL) in y-axis. The average recombination frequency (0.87cM/ FL) of chromosome 1 is marked as the red dashed line.

Table 5.3 Physical, genetic and cytological relationships of eight pairs of markers

Marker pair		Genetic distance (cM)[a]	Physical distance (kb)[b]	Mean absolute cytological distance (µm)[c]	Condensation degree (kb/µm)	Recombination frequency (kb/cM)
P6K1066	P3K3585	5	560	1.27	441	112
P3K3585	P3K4292	5	662	1.26	524	132
P3K2388	P3K3051	16	1506	3.29	458	94
P3K3143	P3K3750	7	523	1.18	443	75
P3K71	P3K629	5	390	1.43	273	78
CPM1737	P3K418	3	308	0.65	473	103
P3K418	P3K7101	6	873	1.08	806	146
K2	K4	-	5408	3.16[d]	1711	-

[a] According to microsatellite marker genetic map (Chen *et al.*, 2007).

[b] Physical distance between BAC clones harboring markers.

[c] Absolute distance between markers on an average sized pachytene chromosome (length = 45.35 µm) based on linear regression analysis.

[d] Unpublished results of Q. Yu and R. Ming.

5.4.5 Abundance of repetitive sequences and genes on chromosome 1

The genome of transgenic papaya female plant, 'SunUp', was sequenced (MING *et al.* 2008) and the papaya scaffold sequences were assembled based on SSR genetic markers and high information-content fingerprinting (YU *et al.* 2009). Twenty-five scaffolds in total length of 21,968,080 bp were anchored on chromosome 1 (Yu et al. 2009). Among the 25 anchored scaffolds, 18 of them with size over 100 kb were included in this cytogenetic analysis. We divided the 18 scaffolds into three groups based on their cytological position: short arm euchromatin, centromeric heterochromatin of X chromosome, and long arm euchromatin. As sequencing of MSY is ongoing, its centromeric heterochromatic region is not included in the present study. The locations of scaffolds were determined by the cytological position of genetic markers on chromosome 1 that were revealed by FISH. The centromeric heterochromatic region of X chromosome contains four scaffolds (scaffolds 39, 66, 177 and 215) and is flanked by the markers P3K2981 and CPM1055 on the genetic map. There are four scaffolds located at the short arm (scaffolds 14, 99, 116, 123 and 140) and nine at the long arm (scaffolds 21, 26, 36, 49, 64, 142, 160, 249 and 286). After removing the gaps within scaffolds, a total of 15,299,791 bp of DNA sequence was analyzed with RepeatMasker, 4,333,240 bp of sequence on the short arm, 2,869,702 bp on the centromeric heterochromatin, and 8,096,849 bp on the long arm (Supplementary file 3).

The total percentage of repetitive sequences in the short arm, centromeric heterochromatin of X chromosome, and long arm was 39%, 49%, and 26%, respectively (Table 5.4). The abundance of repetitive sequences increased in the scaffolds located closer to the centromeric heterochromatin (Fig. 5.3, Supplementary file 3). Transposable elements were the most abundant repetitive sequences. Class I retrotransposons and Class II DNA transposons contributed 31% of the genomic sequence in the short arm, 44% of the centromeric heterochromatin of X chromosome and 20% of long arm (Table 5.4). Among all transposable elements, Ty3/Gypsy retrotransposons were the most abundant repeats in all three regions, occupied 13 to 29% of the region. The *Ty1/Copia* retrotransposons (2-8%) were the second most abundant transposable elements.

The total amount of DNA transposons in three regions was similar, about 0.4-0.6% of each region. However, the distribution of types of DNA transposons was different in the three regions. For example, there was 0.07% of En/Spm in the short and long arms but only 0.03% in centromeric heterochromatin. MuDr showed similar abundance in the long arm (0.014%) and centromeric heterochromatin of X chromosome (0.013%) and slightly greater in the short arm (0.28%).

There were 526 predicted genes on the four scaffolds located on the short arm and 1299 predicted genes on the nine scaffolds located on the long arm (Table 5.4). In contrast, there were only 284 predicted genes on the centromeric heterochromatin of the X chromosome. To compare the gene content in these three regions, the number of predicted genes per 100 kb was calculated. The long arm had the highest gene density of 15.2 genes per 100 kb; the short arm had a lower density of 12.1 genes per 100 kb; and the centromeric heterochromatin had the lowest density of 9.9 genes per 100 kb. Along the length of chromosome 1, the gene density increased in the scaffolds toward the distal ends of each arm and was greatest in the scaffolds located at the end of each arm (Fig. 5.3, Supplementary file 2).

Table 4 Repeat content of chromosome 1

			Short arm Euchromatin (%)	Centromeric heterochromatin on X chromosome (%)	Long arm Euchromatin (%)
Class I retrotransposons			**32.79**	**44.14**	**19.66**
	LTR retrotransposons		30.55	41.86	18.06
		Ty1/Copia	5.36	8.28	2.74
		Ty3/Gypsy	22.24	29.79	13.82
		Unclassified	2.95	3.79	1.51
	Non-LTR retrotransposons		2.09	2.15	1.51
	Unclassified retrotransposons		0.15	0.13	0.10
Class II DNA transposons			**0.42**	**0.56**	**0.40**
	CACTA		0.113	0.065	0.142
	Harbinger		0.005	0	0
	hAT		0.004	0	0.002
	Helitron		0.033	0.057	0.046
	Micropron-like		0.019	0.005	0.017
	MuDR		0.028	0.013	0.014
	MITE		0.018	0.025	0.032
	Unclassified		0.198	0.394	0.147
Unclassified repeat			**0.554**	**0.277**	**0.330**
Low complexity			**4.19**	**3.20**	**4.89**
Simple repeat			**0.90**	**0.73**	**1.11**
Total % of repetitive sequences			**38.85**	**48.91**	**26.40**
Total % of transposable elements			**33.20**	**44.70**	**20.06**
Total no. of gene			**526**	**284**	**1229**
No. of gene/ 100kb			**12.14**	**9.90**	**15.18**
Ratio of *Ty3/Gypsy*: *Ty1/Copia*			4.1	3.6	5.0

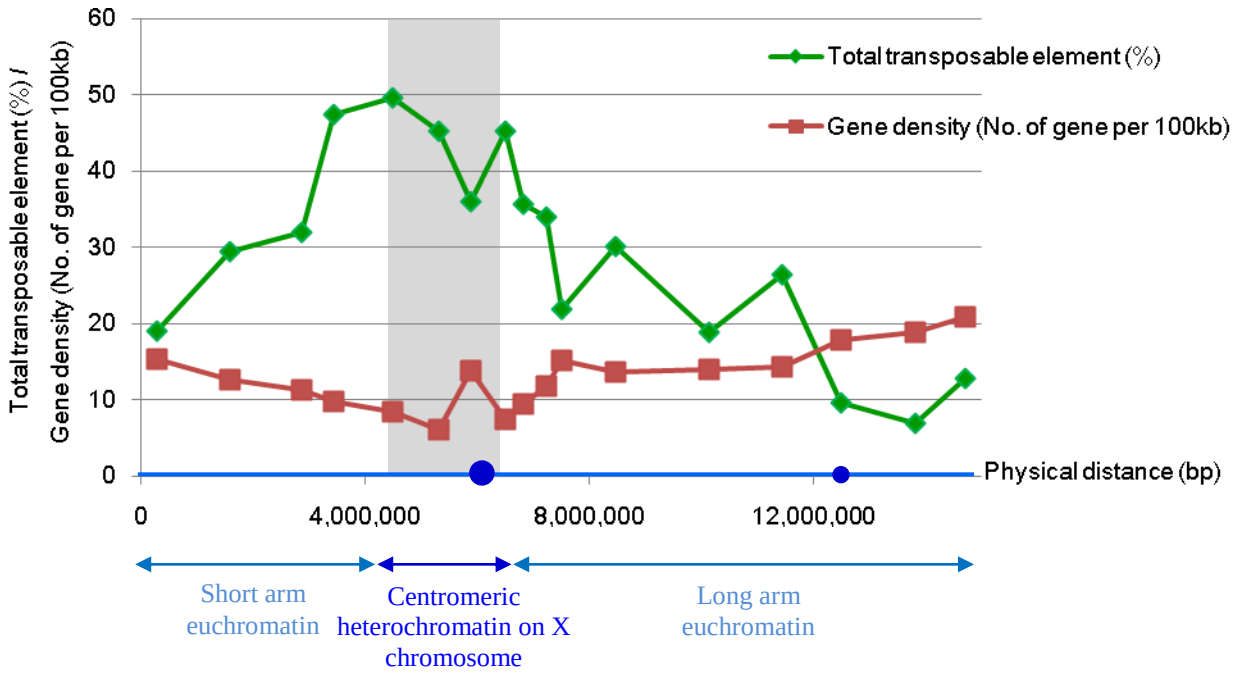

Fig. 5.3 Distribution of transposable elements and gene density along papaya chromosome 1

The 18 scaffolds were joined together as a pseudomolecule and subjected to transposable element distribution and gene density analysis. The percentage of total transposable element (■) and number of gene per 100kb (◆) in each scaffold were plotted against the physical distance from telomere of short arm (in bp). The morphological characters of the X chromosome were represented diagrammatically on the X-axis (blue line) with knob 1 on the centromeric heterochromatin of X chromosome (large blue circle) and a heterochromatic knob on long arm (small blue circle). Chromosome 1 was divided into three regions: short arm euchromatin, long arm euchromatin, and the centromeric heterochromatin shown as the grey color vertical bar.

5.5 Discussion

5.5.1 MSY represents a small chromosomal region on papaya chromosome 1

The first step in the evolution of heteromorphic sex chromosomes is postulated to be a suppression of recombination surrounding a sex determination locus located within an autosome (CHARLESWORTH and MANK ; MING *et al.* 2011; MING and MOORE 2007). Suppression of recombination is not only the result of having the sex determination locus surrounded by repetitive sequences and transposable elements. The lack of recombination does not allow correction of chromosome replication errors which results in the accumulation of additional repetitive sequences and transposable elements in the non-recombination locus. This accumulation of non-coding sequences may eventually expand across the entire chromosome so that what was initially an autosome with a small sex locus becomes a heteromorphic sex chromosome with limited recombination and relatively few functional genes. Such a pathway for sex chromosome evolution has been postulated for animals and plants including papaya (CHARLESWORTH *et al.* 2005; CHARLESWORTH and MANK 2010; MING and MOORE 2007), S*ilene latifolia* (KEJNOVSKY and VYSKOT), grape (DALBO *et al.* 2000), and poplar (YIN *et al.* 2008), based on genetic maps. This hypothesis is supported by the chromosome structure of sex chromosome deletion mutants for *S. latifolia* (LEBEL-HARDENACK *et al.* 2002), however, the non-recombining region of the MSY is difficult to position on the *S. latifolia* sex chromosome due to its large genome size and insufficient cytogenetic information. Although papaya does not have any known sex chromosome deletion mutants, the genome size of the sex determination region of papaya evolving sex chromosome is sufficiently small that it has been genetically mapped and could be mapped cytogenetically to provide evidence for probable mechanisms in the evolution of plant sex chromosomes. This thesis reports the properties of a papaya sex chromosome cytogenetic map that is the first for plants and that will thus act as a foundation for analyzing homomorphic plant sex chromosomes.

Recently, the papaya SSR and AFLP genetic maps revealed a large number of sex-linked markers that co-segregate as a single locus on a pair of chromosomes (CHEN *et al.* 2007; LIU *et al.* 2004; MA *et al.* 2004). This large block of markers may imply either a large physical size or, a sufficient time for divergence between the X and Y chromosomes so there can be no recombination within this block. Our cytogenetic results ruled out large physical size as the reason for the co-segregating markers of the MSY. Our cytogenetic map of papaya chromosome 1 showed that the suppressed recombination is limited to a small MSY region which occupies only 12% of chromosome length. There was no indication of inhibition of recombination immediately beyond 15% of chromosome length adjacent to the MSY region. The small non-recombining MSY region supports the hypothesis that papaya has a pair of young evolving homomorphic sex chromosome (LIU *et al.* 2004; YU *et al.* 2008).

5.5.2 Small recombination gradient is observed on chromosome 1

The centromere is a key component of chromosome structure as it is involved in segregation of sister chromatids. Similar to the MSY, the centromere is a region devoid of crossing over, containing only a few genes, but an abundance of highly repetitive sequences. It has been reported that the non-recombining area of the centromeric region is generally smaller in small genome plant species than in large genome plant species. Rice (*Oriza sativa*) has a small genome of 425 Mbp. The centromeres of *O. sativa* chromosomes 5 and 10 are located at highly suppressed recombination regions that occupy only 6.7% of the length of chromosome 5 and 4% of the length of chromosome 10 (CHENG *et al.* 2001b; KAO *et al.* 2006). For maize, with a relatively large genome (~2,500 Mbp/ 1C), severe genetic recombination suppression over a large region of chromosome 9 occupying 28% of chromosomal length was observed. This region's low recombination rate calculates to be 1.9% of the length of its genetic map (WANG *et al.* 2006a). In wheat (*Triticum aestivum*, ~16,000 Mbp/ 1C) chromosome 3B, 27% of the chromosome is devoid of crossing over (SAINTENAC *et al.* 2009). These studies support the hypothesis that the area of magnitude of the non-recombining region is proportional

to genome size. However, this is not always true since a recent study of the common bean (*Phaseolus vulgaris*), which has small genome (~637 Mbp/ 1C), reported suppressed recombination extending over one-third of the area around the centromere of chromosome 3 (PEDROSA-HARAND *et al.* 2009).

In papaya chromosome 1, the centromere is located within the MSY region where recombination is completely suppressed over 12% of the pachytene chromosome length. The recombination frequency adjacent to this region on each side towards the telomere increased dramatically by 5-fold, and the recombination frequency was similar over the rest of chromosome. The highest (between markers P3K3143 and P3K375) and lowest recombination rate (between K1 and K4) on chromosome 1 is about an 18-fold difference. The small area with suppressed recombination and uniform recombination frequency in regions other than MSY suggested that papaya chromosome 1 had characteristics similar to the chromosomes of small genome plants and are in contrast to the large regions of non-recombination in the large genome plant species. In addition to the possible effects of genome size on the area of suppressed recombination, it also affects the gradient of recombination frequency along the chromosome. Maize chromosome 9 shows a steep gradient of recombination rates with the centromere having the lowest recombination frequency and higher rates of recombination towards the distal ends. The recombination rate between centromere and telomeric ends varies 145-fold (WANG *et al.* 2006a). A large recombination gradient was also revealed in the cytogenetic map of wheat chromosome 3B (AKHUNOV *et al.* 2003; SAINTENAC *et al.* 2009). On the other hand, rice chromosome 5 shows a low recombination rate only at the centromere and uniform distribution of recombination over both arms (KAO *et al.* 2006). Similar overall recombination rate along two chromosomal arms is reported in *Arabidopsis thaliana* chromosome 1 and 2 (LIN *et al.* 1999; THEOLOGIS *et al.* 2000). More research is needed to determine the validity of the hypothesis that the area of non-recombining region is proportional to the genome size.

5.5.3 Recombination rate is related to chromosomal position and genome compositions

We used FISH to combine chromosome features with genetic and physical distances to categorize chromosome 1 into three specific regions: 1) a short arm euchromatic region between 0 cM and 46 cM; 2) a centromeric heterochromatin (MSY or homologous region of X chromosome) between 46 cM and 50 cM; and 3) a long arm euchromatic region between 50 cM and 145 cM. The 18 marker-tagged scaffolds that assembled the X chromosome physical map were then categorized in same way for repetitive sequence and gene density analysis. This cytological information enabled an analysis of the difference between the euchromatic arm and the heterochromatic pericentromere/ centromere of the X chromosome which could have application to the other chromosomes. We found difference in the recombination frequency, gene density, and repetitive sequences content between the centromeric heterochromatin and euchromatin of the chromosome arms. Although all euchromatic regions appear the same as visualized under microscopy, the genome composition and recombination rate of the two arms were highly variable.

The short arm euchromatin of the SSR genetic map occupied 46 cM and the long arm euchromatin occupied 95 cM, or about twice the length of the short arm. However, our cytogenetic map indicated that the cytological lengths of the short and long arms were similar. If we assume K4 is the centromere, then the cytological distance of the short arm was 60.41 FL and long arm was 84.59 FL. The p/q ratio of chromosome 1 would thus be 0.71, indicating that chromosome 1 is sub-metacentric even though the genetic map shows a 2-fold difference in genetic length between the short and long arms. The average recombination frequency of long arm is 40% higher than short arm. The discrepancy between cytological and genetic length may due to differences in the gene and repetitive sequence contents of the two arms.

Although both short and long arms of chromosome 1 are mainly euchromatic, the long arm had 25% less transposable element sequences and 40% higher gene density than did the short arm. In previous studies (MING *et al.* 2008; ZHANG *et al.* 2008), it was reported that one of the euchromatic arms of chromosome 1 is more densely stained by DAPI. That observation might be reconciled by the higher amount of transposable elements and repeats we report in present study. In addition, more transposable elements and repeats may contribute to the lower gene density and lower recombination rate in the short arm. It has been reported that recombination rate in the rice genome is negatively correlated with gene density (TIAN *et al.* 2009a). In certain arabidopsis chromosome, the recombination rate was positively correlated with transposable element content (WRIGHT *et al.* 2003; WU *et al.* 2006) .

When we examined the genome composition in each scaffold closely, a negatively correlation between gene density and transposable elements content was observed, which is the same as arabidopsis and rice genomes (TIAN *et al.* 2009b; WRIGHT *et al.* 2003). Besides, the distribution of transposable elements and genes is strongly affected by its chromosomal position. The density of gene was the lowest at pericentromeric and centromeric region, and then increased toward both arms and reached the highest (up to 21 genes per 100 kb) at the distal ends. In an opposite manner, the percentage of transposable elements varied from 50% at the centromeric heterochromatin down to 8% at the distal ends. At the same time, the recombination frequency is the lowest in centromeric/ pericentromeric region and increased towards the distal ends of both arms. This suggests that the recombination rate may be related to the gene density and transposable elements distribution. Also, the chromosomal position may have influence on the deposition of gene and transposable element. This study gives an overall view of the genome composition of the papaya sex chromosome. In order to study the relationships between genome composition, recombination frequency and chromosomal position statistically, the genome composition should be analyzed in a smaller portions and a high-density cytogenetic map should be constructed which will give a more precise information.

Centromere is a gene-poor region in plant and animal species. In this study, only four assembled contigs were included in this study to represent the centromeric heterochromatic region. These four contigs may locate at the pericentromeric region, which usually has higher gene density, compared to the centromere. Secondly, contigs with high abundance of repetitive sequences are difficult to assemble, thus may not be included in this analysis. Therefore, the gene density of this region is likely overestimated while the percentage of repetitive sequence is underestimated. A complete sequencing of MSY and homologous region on X chromosome will give more accurate information. Although the genome coverage at heterochromatic region needs to be increased, we can still observe a trend of gene and transposable element distribution along chromosome 1.

5.5.4 Transposable elements distribute unevenly on chromosome 1

A detail examination of transposable elements in regions distributed along chromosome 1 identified preferential regions for transposable elements, which was not shown in whole genome analysis. In maize and rice genomes, most DNA transposons are enriched in gene-rich, recombinationally active region while retrotransposons are enriched in gene-poor, recombinationally inactive region (2005; SCHNABLE *et al.* 2009; WEI *et al.* 2009). Papaya genome showed similar patterns as these two genomes on retrotransposons distribution but not on DNA transposons. Examination of the distribution of different classes of transposable elements on the papaya X chromosome showed that although *Ty3/Gypsy* was the most abundant transposable elements over the entire region, more *Ty1/Copia* was observed in the centromeric heterochromatin than in the euchromatic arms. A genome-wide analysis of transposable element in papaya, showed the overall ratio of *Ty1/Copia* to *Ty3/Gypsy* to be about 1:5 (NAGARAJAN *et al.* 2008), which is similar to the ratio we observed in the long arm euchromatin of chromosome 1. The ratio increased to 1:4 in the short arm euchromatin and centromeric heterochromatin. Peters *et al.* (2009) reported that the heterochromatin of tomato chromosome 6, contains more *Ty3/Gypsy* than *Ty1/Copia* and the opposite phenomenon was observed in euchromatin. The *Ty3/Gypsy* is concentrated more in the pericentromeric heterochromatin of the maize genome, while *Ty1/Copia* is more common in the euchromatin of distal arms (BAUCOM *et*

al. 2009). We did not observe a similar dominant abundance of *Ty1/Copia* in the euchromatin of papaya chromosome 1.

Overall, the abundance of DNA transposons was similar (0.4-0.6%) in both arms and the centromeric heterochromatin of chromosome 1. However, there was two times more CACTA sequence in each of the arms than in the centromeric heterochromatin. This distribution of CACTA in chromosome 1 was similar to the situation in papaya chromosome 6, where three times more CACTA was observed in each euchromatin arm, than in the centromeric heterochromatin (Wai, C.M. and Yu, Q., unpublished data). The mechanism for differential distribution of CACTA is unknown, but cytogenetic studies of *Triticeae* species, reported a significant reduction of En/Spm element around centromeres when visualized by FISH (ALTINKUT *et al.* 2006). Preferential location of CACTA in pericentromeric region was also observed in arabidopsis (MIURA *et al.* 2004). Contrarily, uniform distribution with cluster of En/Spm signals around middle of chromosomes was reported in a sequence and cytological analysis of maize (WEI *et al.* 2009; YU *et al.* 2007b).

For another DNA transposon MuDr, it showed the highest abundance at short arm euchromatin with similar amount in both centromeric heterochromatin and long arm euchromatin. This suggests that the distribution of this transposable element may not related to gene density, recombination rate, nor chromosomal structure. Uneven distribution of specific retrotransposons and DNA transposons indicate a preferential insertion of genetic elements along chromosomes, or a slower elimination rate from the genome, during evolution. It is also possible that the distribution of transposable elements could be species-, or chromosome-specific. Additional papaya chromosomes would need to be studied to draw a firm conclusion. Recently, it has been reported that each transposable element has its own specific methylation environment which affect its location in the rice genome (TAKATA *et al.* 2007). This specificity may contribute to the diverse location of individual transposable elements among different species.

5.5.5 Cytogenetic map unravels discrepancy of physical and genetic maps

Chromosomal recombination occurs primarily in euchromatin and rarely in heterochromatin. Therefore, it is difficult to map the order of DNA markers in heterochromatic regions, such as the centromere and the sex-determination locus, due to little or no recombination and a high content of repetitive sequences. The order of microsatellite markers sometimes differs between genetic and physical maps thus necessitating the development of a cytogenetic map to validate the accuracy of either type of maps. Discrepancies of DNA markers at cytological and genetic position have been reported in some plants. For example, in maize chromosome 9, a probe that cytologically located in the pericentromere was mistakenly mapped to the telomeric end in physical and genetic maps (DANILOVA and BIRCHLER 2008).

Our FISH mapping of 16 microsatellite marker-tagged BAC clones on pachytene chromosome spreads showed the same order of markers for all but three of the markers as on the SSR genetic map (CHEN *et al.* 2007). That map reported the two markers (P3K4292 and P6K1460) in the pericentromeric and centromeric regions were in reversed order on the cytogenetic map. The third marker (Ctg566) that had mapped to the telomere was cytologically within the long arm. The discrepancy between the SSR map and our cytogenetic map is likely due to a mistake in genetic mapping due to the low recombination rate in the highly heterochromatic pericentromeric, centromeric, and telomeric regions.

Comparing our cytogenetic map to the AFLP marker-enriched SSR genetic map constructed by Blas et al. (2009), showed markers located on the short arm of our cytogenetic map were in the same order as on the AFLP enriched map (BLAS *et al.* 2009). However, the nine cytogenetic markers that located on the long arm of our cytogenetic map were scattered over both arms of enriched genetic map in a different order. This contrast suggests that the SSR genetic map predicts marker order more accurately than either the AFLP or the AFLP marker-enriched SSR integrated map. The difference

between the enriched genetic map and our FISH results may be caused by quality of data used for genetic map construction.

There are two major gaps on LG 1 of SSR genetic map where the adjacent markers are more than 15 cM apart. The physical distance of the gap located between 8 cM and 24 cM could be determined by measuring the distance of FISH signals from corresponding BAC clones. This information will help us to developing strategies to increase marker density in this region of the genetic map. The other genetic gap located between markers P3K7101 (122 cM) and Ctg566 (145 cM), did not appear in our cytogenetic map. Marker Ctg566 was cytologically mapped towards the middle of the long arm adjacent to marker CPM1737 at 113 cM instead of at the telomeric end as it was mapped genetically. Since the apparent genetic gap was not confirmed by our cytological gap, no additional markers are required to fill it.

Recently, the analytical power of FISH for cross species analyses has been demonstrated in *Cucumis* spp. (LIU *et al.* 2010), maize (AMARILLO and BASS 2007; FIGUEROA *et al.* 2011), and *Solanum* spp. (IOVENE *et al.* 2008; LOU *et al.* 2010; TANG *et al.* 2008). Lou *et al.* (2010) used 13 BAC clones isolated from potato to study the evolution of chromosome 6 in seven distantly related *Solanum* species, including tomato and eggplant. Similarly, the BAC clones from *Carica papaya* could be used as surrogate clones in FISH analysis of closely related *Vasconcellea* species. The cytogenetic map of this present study can become a foundation for the comparative and evolutionary studies of sex chromosomes in *Vasconcellea* species.

Supplementary File 1 Distribution of euchromatin and heterochromatin on chromosome 1

Sample	Chromosome length (µm)	Short arm euchromatin length (µm)[a]	Hetero-chromatin length (µm)[b]	Long arm euchromatin length (µm)[c]	Euchromatin total length (µm)[d]	Short arm of euchromatin (%)	Long arm of euchromatin (%)	Euchromatin (%)	Hetero-chromatin (%)
1	39.699	17.444	6.315	15.940	33.384	52.25	47.75	84.09	15.91
2	46.824	19.940	6.009	20.875	40.815	48.85	51.15	87.17	12.83
3	38.479	12.563	6.154	19.762	32.325	38.86	61.14	84.01	15.99
4	60.933	28.990	7.769	24.174	53.164	54.53	45.47	87.25	12.75
5	41.757	16.720	6.976	18.061	34.781	48.07	51.93	83.29	16.71
6	53.684	22.325	6.606	24.753	47.078	47.42	52.58	87.69	12.31
7	42.943	17.087	6.431	19.425	36.512	46.80	53.20	85.02	14.98
8	41.632	15.095	7.406	19.131	34.226	44.10	55.90	82.21	17.79
9	42.333	14.531	7.253	20.549	35.080	41.42	58.58	82.87	17.13
10	59.386	22.602	6.401	30.383	52.985	42.66	57.34	89.22	10.78
11	74.680	32.951	14.280	27.449	60.400	54.55	45.45	80.88	19.12
12	39.892	18.509	6.133	15.250	33.759	54.83	45.17	84.63	15.37
13	44.379	17.023	7.419	19.937	36.960	46.06	53.94	83.28	16.72
14	36.791	15.615	6.998	14.178	29.793	52.41	47.59	80.98	19.02
15	39.301	16.584	5.042	17.675	34.259	48.41	51.59	87.17	12.83
16	33.647	15.017	4.588	14.042	29.059	51.68	48.32	86.36	13.64
17	28.934	12.069	4.549	12.316	24.385	49.49	50.51	84.28	15.72
18	39.165	16.849	6.132	16.184	33.033	51.01	48.99	84.34	15.66
				Mean		48.52	51.48	84.71	15.29
				Standard deviation		4.64	4.64	2.35	2.35
				correlation of short and long arm length		0.733			
				correlation of euchromatin and heteromatin length		0.703			

[a] Chromosome distance between knob 5 and closest telomere

[b] Chromosome distance between knob 1 and knob 5

[c] Chromosome distance between knob 1 and closest telomere

[d] Total length of short and long arm euchromatin

Supplementary File 2 Recombination frequency along papaya chromosome 1

Marker pair		Chromosome position	Genetic distance (cM)[a]	Cytological distance (FL±S.D.[b])	Ratio of genetic to cytological distance (cM/ FL)
P6K72	P6K1066	Short arm	24	27.06 ± 1.93	0.87
P6K1066	P3K3585		5	3.87 ± 2.03	1.29
P3K3585	P3K4292		5	3.81 ± 2.35	1.31
P3K4292	P6K1460		13	19.12 ± 4.60	0.68
P6K1460	P3K2981	Heterochromatin	2	16.97 ± 4.95	0.12
P3K2981	P3K2388	Heterochromatin + Long arm	14	23.02 ± 4.57	0.61
P3K2388	P3K3051	Long arm	16	10.12 ± 5.25	1.58
P3K3051	P3K3143		6	4.14 ± 5.48	1.45
P3K3143	P3K3750		7	3.30 ± 5.71	2.12
P3K3750	P3K71		7	5.87 ± 5.84	1.19
P3K71	P3K629		5	4.19 ± 6.27	1.19
P3K629	CPM1737		12	12.41 ± 5.09	0.97
CPM1737	P3K418		3	1.71 ± 2.64	1.75
P3K418	P3K7101		6	3.03 ± 2.64	1.98

[a] According to microsatellite marker genetic map (Chen *et al.*, 2007).

[b] S.D. represents standard deviation.

Supplementary File 3 Transposable element and gene content of scaffolds on chromosome 1

Chromosome	scaffold	estimated physical size (bp)	masked sequence size (bp)	retro-transposon (%)	DNA transposon (%)	low complexity repeat (%)	simple repeat (%)	unclassified (%)	Total transposable element (%)	Total repeat (%)	predicted genes (no.)	predicted gene density (no./100kb)
Short arm euchromatin	116	955,180	575,408	17.32	0.74	6.35	1.52	0.92	18.06	26.85	88	15.3
	14	2,884,448	2,041,850	29.68	0.37	4.07	0.91	0.66	30.05	35.69	257	12.6
	140	828,544	497,008	33.93	0.56	4.41	0.74	0.37	34.50	40.02	56	11.3
	99	989,767	664,089	49.42	0.34	3.56	0.77	0.31	49.77	54.40	65	9.8
	123	736,463	554,885	39.31	0.20	2.94	0.56	0.25	39.52	43.27	60	10.8
Centromeric Heterochromatin	39	2,371,359	1,445,656	47.74	0.76	2.80	0.64	0.11	48.50	52.05	121	8.4
	215	291,840	196,739	44.79	0.37	2.69	0.43	1.15	45.16	49.43	12	6.1
	66	1,257,971	928,489	37.52	0.39	3.50	0.83	0.24	37.91	42.49	129	13.9
	177	657,316	298,818	46.84	0.26	4.58	0.99	0.61	47.10	53.29	22	7.4
Long arm euchromatin	160	608,068	349,964	35.78	0.37	5.25	1.27	0.27	36.15	42.94	33	9.4
	142	726,689	466,720	33.81	0.24	4.95	0.89	0.23	34.05	40.13	55	11.8
	249	327,176	111,716	23.18	0.45	6.07	1.01	0.20	23.63	30.91	17	15.2
	21	2,510,081	1,804,902	30.16	0.37	5.32	1.01	0.05	30.53	36.92	246	13.6
	36	1,794,408	1,496,058	19.68	0.28	5.04	1.12	0.36	19.96	26.48	209	14.0
	49	1,506,978	1,150,572	25.71	0.59	4.94	1.15	0.23	26.30	32.63	165	14.3
	64	1,200,704	948,862	8.65	0.64	4.56	1.31	0.58	9.29	15.75	170	17.9
	26	2,069,338	1,696,177	7.86	0.32	4.36	1.09	0.58	8.18	14.21	319	18.8
	286	108,169	71,878	12.98	0.30	3.35	1.04	0.25	13.28	17.92	15	20.9
Short arm euchromatin		6,394,402	4,333,240								526	12.1
Heterochromatin		4,578,486	2,869,702								284	9.9
Long arm euchromatin		10,851,611	8,096,849								1229	15.2
Entire chromosome 1		21,824,499	15,299,791								2039	13.3

Chapter 6 Conclusion

Papaya flowers throughout the year which provides large quantity of easily accessible meiotic tissues for cytological analyses. However, the similar small size of papaya chromosome makes identification of individual chromosomes difficult. With the development and proof of a set of BAC-based cytogenetic markers, we now have a reproducible method to identify each papaya chromosome pair. This set of markers will facilitate analyses of the distribution patterns of transposable elements and repetitive sequences on individual chromosome and reveal the relationship between these sequences and chromosomal structure.

Although the papaya genome has been sequenced, gaps remain in the assembly, mainly in the heterochromatic regions such as centromere and telomere, In addition, prior to the presently reported work, there was no information relating chromosome structure to genome sequence. By hybridizing genetic-marker tagged BAC clones to the papaya sex chromosome, the genetic map was integrated with chromosome structures. The genome composition in different chromosome regions was also studied through sequencing and physically mapping the selected BAC clones. We found that the papaya sex chromosome shares certain genome structure with some previously reported plant species. Firstly, there are significant difference in recombination rate and condensation degree between the chromosomal euchromatic and heterochromatic regions. Secondly, the transposable elements and repetitive sequences are mostly found in heterochromatin while more genes are found in euchromatin. Lastly, the recombination rate is similar along euchromatic short and long arms while the condensation pattern varied.

However, some unexpected genome structures occurred on this sex chromosome. Instead of a random distribution of transposable elements along chromosome, certain ones showed preferential locations on other parts of the papaya genome. For example, the DNA transposons family CACTA was preferentially located in the euchromatin rather than in the heterochromatin. Also, an uneven distribution of genes was observed with a

higher gene density in the long arm compared to the short arm. This information is suggestive that the papaya genome is more complex and dynamic than we originally hypothesized. Additional research is needed to determine whether the uneven genome composition on chromosome 1 also occurs on other papaya chromosomes.

Finally, the present study gave additional evidence to support the hypothesis that papaya's pair of sex chromosomes is young and evolving. Lack of recombination or crossing over in the small region which carries sex determination gene(s) is an initial step for an autosome to become a sex chromosome. By studying the recombination rate along the papaya sex chromosome, we confirmed that the suppressed recombination was restricted to the MSY region, which occupies only 12% of chromosome length. This study combining the genomic resources with cytological information adds genetic and genomic data that the physical and genetic maps cannot provide. Our presently developed cytogenetic map of the papaya sex chromosome can act as a foundation for a comparative study of sex chromosomes in the closely related *Vasconcellea* species. Once the cytogenetic maps of the sex chromosomes of *Vasconcellea* species are constructed, the suppressed recombination regions can be compared to discover the generality of change in genome structure as sex chromosomes evolve.

Reference

2005 The map-based sequence of the rice genome. Nature **436:** 793-800.

AKHUNOV, E. D., A. W. GOODYEAR, S. GENG, L. L. QI, B. ECHALIER *et al.*, 2003 The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res **13:** 753-763.

ALTINKUT, A., V. KOTSERUBA, V. M. KIRZHNER, E. NEVO, O. RASKINA *et al.*, 2006 Ac-like transposons in populations of wild diploid Triticeae species: comparative analysis of chromosomal distribution. Chromosome Res **14:** 307-317.

AMARILLO, F. I., and H. W. BASS, 2007 A transgenomic cytogenetic sorghum (Sorghum propinquum) bacterial artificial chromosome fluorescence in situ hybridization map of maize (Zea mays L.) pachytene chromosome 9, evidence for regions of genome hyperexpansion. Genetics **177:** 1509-1526.

APPELS, R., R. MORRIS, B. S. GILL and C. E. MAY, 1998 *Chromosome biology*. Kluwer Academic Publisher, Boston.

ARUMUGANATHAN, K., and E. D. EARLE, 1991 Nuclear DNA content of some important plant species. Plant mol bio rep **93:** 208-219.

BAUCOM, R. S., J. C. ESTILL, C. CHAPARRO, N. UPSHAW, A. JOGI *et al.*, 2009 Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. PLoS Genet **5:** e1000732.

BICKMORE, W., and J. CRAIG, 1997 *Chromosome bands: patterns in the genome*. RG Landes Publ, Austin, TX.

BLAS, A. L., Q. YU, C. CHEN, O. VEATCH, P. H. MOORE *et al.*, 2009 Enrichment of a papaya high-density genetic map with AFLP markers. Genome **52:** 716-725.

BUDIMAN, M. A., S. B. CHANG, S. LEE, T. J. YANG, H. B. ZHANG *et al.*, 2004 Localization of jointless-2 gene in the centromeric region of tomato chromosome 12 based on high resolution genetic and physical mapping. Theor Appl Genet **108:** 190-196.

CHAO, S., J. M. GARDINER, S. MELIA-HANCOCK and E. H. COE, JR., 1996 Physical and genetic mapping of chromosome 9S in maize using mutations with terminal deficiencies. Genetics **143:** 1785-1794.

CHARLESWORTH, D., B. CHARLESWORTH and G. MARAIS, 2005 Steps in the evolution of heteromorphic sex chromosomes. Heredity **95:** 118-128.

CHARLESWORTH, D., and J. E. MANK, The birds and the bees and the flowers and the trees: lessons from genetic mapping of sex determination in plants and animals. Genetics **186:** 9-31.

CHARLESWORTH, D., and J. E. MANK, 2010 The birds and the bees and the flowers and the trees: lessons from genetic mapping of sex determination in plants and animals. Genetics **186:** 9-31.

CHEN, C., Q. YU, S. HOU, Y. LI, M. EUSTICE *et al.*, 2007 Construction of a sequence-tagged high-density genetic map of papaya for comparative structural and evolutionary genomics in brassicales. Genetics **177:** 2481-2491.

CHEN, C. C., C. M. CHEN, F. C. HSU, C. J. WANG, J. T. YANG *et al.*, 2000 pachytene chromosomes of maize as revealed by fluorescence in situ hybridization with repetitive DNA sequences. Theor Appl Genet **101:** 30-36.

CHENG, Z., C. R. BUELL, R. A. WING, M. GU and J. JIANG, 2001a Toward a cytological characterization of the rice genome. Genome Res **11:** 2133-2141.

CHENG, Z., C. R. BUELL, R. A. WING and J. JIANG, 2002 Resolution of fluorescence in-situ hybridization mapping on rice mitotic prometaphase chromosomes, meiotic pachytene chromosomes and extended DNA fibers. Chromosome Res **10:** 379-387.

CHENG, Z., G. G. PRESTING, C. R. BUELL, R. A. WING and J. JIANG, 2001b High-resolution pachytene chromosome mapping of bacterial artificial chromosomes anchored by genetic markers reveals the centromere location and the distribution of genetic recombination along chromosome 10 of rice. Genetics **157:** 1749-1757.

CHUNG, M. C., Y. I. LEE, Y. Y. CHENG, Y. J. CHOU and C. F. LU, 2008 Chromosomal polymorphism of ribosomal genes in the genus Oryza. Theor Appl Genet **116:** 745-753.

COOPER, M., F. A. VAN EEUWIJK, G. L. HAMMER, D. W. PODLICH and C. MESSINA, 2009 Modeling QTL for complex traits: detection and context for plant breeding. Curr Opin Plant Biol **12:** 231-240.

CREIGHTON, H. B., and B. MCCLINTOCK, 1931 A Correlation of Cytological and Genetical Crossing-Over in Zea Mays. Proc Natl Acad Sci U S A **17:** 492-497.

CUADRADO, A., M. CARDOSO and N. JOUVE, 2008 Physical organisation of simple sequence repeats (SSRs) in Triticeae: structural, functional and evolutionary implications. Cytogenet Genome Res **120:** 210-219.

DALBO, M. A., G. N. YE, N. F. WEEDEN, H. STEINKELLNER, K. M. SEFC *et al.*, 2000 A gene controlling sex in grapevines placed on a molecular marker-based genetic map. Genome **43:** 333-340.

DANILOVA, T. V., and J. A. BIRCHLER, 2008 Integrated cytogenetic map of mitotic metaphase chromosome 9 of maize: resolution, sensitivity, and banding paint development. Chromosoma **117:** 345-356.

DE CARVALHO, C. R. A. S., L. S., 1997 High-resolution HKG-banding in maize mitotic chromosomes J Plant Res **110:** 417-420.

DE MELO, N. F., and M. GUERRA, 2003 Variability of the 5S and 45S rDNA sites in Passiflora L. species with distinct base chromosome numbers. Ann Bot **92:** 309-316.

DEPUTY, J. C., R. MING, H. MA, Z. LIU, M. M. FITCH *et al.*, 2002 Molecular markers for sex determination in papaya (Carica papaya L.). Theor Appl Genet **106:** 107-111.

DIEGUEZ, M. J., E. ALTIERI, L. R. INGALA, E. PERERA, F. SACCO *et al.*, 2006 Physical and genetic mapping of amplified fragment length polymorphisms and the leaf rust resistance Lr3 gene on chromosome 6BL of wheat. Theor Appl Genet **112:** 251-257.

DONG, F., J. SONG, S. K. NAESS, J. P. HELGESON, C. GEBHARDT *et al.*, 2000 Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. Theor Appl Genet **101:** 1001-1007.

FAOSTAT, 2008, pp.

FENG, Q., Y. ZHANG, P. HAO, S. WANG, G. FU *et al.*, 2002 Sequence and analysis of rice chromosome 4. Nature **420:** 316-320.

FEUILLET, C., J. E. LEACH, J. ROGERS, P. S. SCHNABLE and K. EVERSOLE, 2011 Crop genome sequencing: lessons and rationales. Trends Plant Sci **16:** 77-88.

FIGUEROA, D. M., J. D. DAVIS, C. STROBEL, M. S. CONEJO, K. D. BECKHAM *et al.*, 2011 The selection and use of sorghum (Sorghum propinquum) bacterial artificial chromosomes as cytogenetic FISH probes for maize (Zea mays L.). J Biomed Biotechnol **2011:** 386862.

FITCH, M. M. M., R. M. MANSHARDT, D. GONSALVES and J. L. SLIGHTOM, 1993 Transgenic papaya plants from Agrobacterium-mediated transformation of somatic embryos plant cell reports **12:** 245-249.

FITCH, M. M. M., R. M. MANSHARDT, D. GONSALVES, J. L. SLIGHTOM and J. C. SANFORD, 1990 Stable transformation of papaya via microprojectile bombardment plant cell reports **9:** 189-194.

FRANSZ, P. F., C. ALONSO-BLANCO, T. B. LIHARSKA, A. J. PEETERS, P. ZABEL *et al.*, 1996 High-resolution physical mapping in Arabidopsis thaliana and tomato by fluorescence in situ hybridization to extended DNA fibres. Plant J **9:** 421-430.

FRANSZ, P. F., S. ARMSTRONG, J. H. DE JONG, L. D. PARNELL, C. VAN DRUNEN *et al.*, 2000 Integrated cytogenetic map of chromosome arm 4S of A. thaliana: structural organization of heterochromatic knob and centromere region. Cell **100:** 367-376.

GALL, J. G., and M. L. PARDUE, 1969 FORMATION AND DETECTION OF RNA-DNA HYBRID MOLECULES IN CYTOLOGICAL PREPARATIONS *. proc Natl Acad Sci U S A **63:** 378-383.

GRIFFOR, M. C., L. O. VODKIN, R. J. SINGH and T. HYMOWITZ, 1991 Fluorescent in situ hybridization to soybean metaphase chromosomes. Plant Mol Biol **17:** 101-109.

HAINER, A., and C. U. HESEMANN, 1988 Cytogenetic investigations in rye, wheat and triticale. TAG Theoretical and applied genetics **75:** 305-308.

HEILBORN, O., 1921 Taxonomical and cytological studies on cultivated Ecuodorian species of Carica. Ark Bot **17:** 1-16.

HIZUME, M., F. SHIBATA, Y. MATSUSAKI and Z. GARAJOVA, 2002 Chromosome identification and comparative karyotypic analyses of four Pinus species. Theor Appl Genet **105:** 491-497.

HOFMEYR, J. D. J., 1938 Genetical studies of Carica papaya L. I. The inheritance and relation of sex and certain plant characteristics. II. Sex reversal and sex forms. S Afr Dept Agri and Sci Bul **No. 187:** 64.

HOHMANN, U., T. R. ENDO, K. S. GILL and B. S. GILL, 1994 Comparison of genetic and physical maps of group 7 chromosomes from Triticum aestivum L. Mol Gen Genet **245:** 644-653.

IOVENE, M., S. M. WIELGUS, P. W. SIMON, C. R. BUELL and J. JIANG, 2008 Chromatin structure and physical mapping of chromosome 6 of potato and comparative analyses with tomato. Genetics **180:** 1307-1317.

IRGSP, 2005 The map-based sequence of the rice genome. Nature **436:** 793-800.

ISLAM-FARIDI, M. N., K. L. CHILDS, P. E. KLEIN, G. HODNETT, M. A. MENZ *et al.*, 2002 A molecular cytogenetic map of sorghum chromosome 1. Fluorescence in situ hybridization analysis with mapped bacterial artificial chromosomes. Genetics **161:** 345-353.

ISLAM-FARIDI, M. N., C. D. NELSON and T. L. KUBISIAK, 2007 Reference karyotype and cytomolecular map for loblolly pine (Pinus taeda L.). Genome **50:** 241-251.

JACKSON, S. A., M. L. WANG, H. M. GOODMAN and J. JIANG, 1998 Application of fiber-FISH in physical mapping of Arabidopsis thaliana. Genome **41:** 566-572.

JANSEN, R. C., B. M. TESSON, J. FU, Y. YANG and L. M. MCINTYRE, 2009 Defining gene and QTL networks. Curr Opin Plant Biol **12:** 241-246.

JIANG, J., and B. S. GILL, 2006 Current status and the future of fluorescence in situ hybridization (FISH) in plant genome research. Genome **49:** 1057-1068.

JIANG, J., B. S. GILL, G. L. WANG, P. C. RONALD and D. C. WARD, 1995 Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes. Proc Natl Acad Sci U S A **92:** 4487-4491.

JOHN, H. A., M. L. BIRNSTIEL and K. W. JONES, 1969 RNA-DNA hybrids at the cytological level. Nature **223:** 582-587.

JURKA, J., V. V. KAPITONOV, A. PAVLICEK, P. KLONOWSKI, O. KOHANY *et al.*, 2005 Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res **110:** 462-467.

KAO, F. I., Y. Y. CHENG, T. Y. CHOW, H. H. CHEN, S. M. LIU *et al.*, 2006 An integrated map of Oryza sativa L. chromosome 5. Theor Appl Genet **112:** 891-902.

KATO, A., J. C. LAMB and J. A. BIRCHLER, 2004 Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proc Natl Acad Sci U S A **101:** 13554-13559.

KATO, A., J. M. VEGA, F. HAN, J. C. LAMB and J. A. BIRCHLER, 2005 Advances in plant chromosome identification and cytogenetic techniques. Curr Opin Plant Biol **8:** 148-154.

KEJNOVSKY, E., and B. VYSKOT, Silene latifolia: the classical model to study heteromorphic sex chromosomes. Cytogenet Genome Res **129:** 250-262.

KIM, H., S. R. CHOI, J. BAE, C. P. HONG, S. Y. LEE *et al.*, 2009 Sequenced BAC anchored reference genetic map that reconciles the ten individual chromosomes of Brassica rapa. BMC Genomics **10:** 432.

KIM, J. S., K. L. CHILDS, M. N. ISLAM-FARIDI, M. A. MENZ, R. R. KLEIN *et al.*, 2002 Integrated karyotyping of sorghum by in situ hybridization of landed BACs. Genome **45:** 402-412.

KIM, J. S., M. N. ISLAM-FARIDI, P. E. KLEIN, D. M. STELLY, H. J. PRICE *et al.*, 2005a Comprehensive molecular cytogenetic analysis of sorghum genome architecture: distribution of euchromatin, heterochromatin, genes and recombination in comparison to rice. Genetics **171:** 1963-1976.

KIM, J. S., P. E. KLEIN, R. R. KLEIN, H. J. PRICE, J. E. MULLET *et al.*, 2005b Molecular cytogenetic maps of sorghum linkage groups 2 and 8. Genetics **169:** 955-965.

KLEIN, P. E., R. R. KLEIN, J. VREBALOV and J. E. MULLET, 2003 Sequence-based alignment of sorghum chromosome 3 and rice chromosome 1 reveals extensive conservation of gene order and one major chromosomal rearrangement. Plant J **34:** 605-621.

KOO, D. H., and J. JIANG, 2009 Super-stretched pachytene chromosomes for fluorescence in situ hybridization mapping and immunodetection of DNA methylation. Plant J **59:** 509-516.

KOO, D. H., S. H. JO, J. W. BANG, H. M. PARK, S. LEE *et al.*, 2008 Integration of cytogenetic and genetic linkage maps unveils the physical architecture of tomato chromosome 2. Genetics **179:** 1211-1220.

KOORNNEEF, M., P. FRANSZ and H. DE JONG, 2003 Cytogenetic tools for Arabidopsis thaliana. Chromosome Res **11:** 183-194.

KOUMBARIS, G. L., and H. W. BASS, 2003 A new single-locus cytogenetic mapping system for maize (Zea mays L.): overcoming FISH detection limits with marker-selected sorghum (S. propinquum L.) BAC clones. Plant J **35:** 647-659.

KRISHNAN, P., V. T. SAPRA, K. M. SOLIMAN and A. ZIPF, 2001 FISH mapping of the 5S and 18S-28S rDNA loci in different species of Glycine. J Hered **92:** 295-300.

KUMAR, L., and A. S. V. ABRAHAM, 1945 The cytology of Carica papaya Linn Indian J Agr Sci **15:** 242-253.

KUNZEL, G., L. KORZUN and A. MEISTER, 2000 Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics **154:** 397-412.

LAI, C. W., Q. YU, S. HOU, R. L. SKELTON, M. R. JONES *et al.*, 2006 Analysis of papaya BAC end sequences reveals first insights into the organization of a fruit tree genome. Mol Genet Genomics **276:** 1-12.

LAMB, J. C., T. DANILOVA, M. J. BAUER, J. M. MEYER, J. J. HOLLAND *et al.*, 2007 Single-gene detection and karyotyping using small-target fluorescence in situ hybridization on maize somatic chromosomes. Genetics **175:** 1047-1058.

LEBEL-HARDENACK, S., E. HAUSER, T. F. LAW, J. SCHMID and S. R. GRANT, 2002 Mapping of sex determination loci on the white campion (Silene latifolia) Y chromosome using amplified fragment length polymorphism. Genetics **160:** 717-725.

LENGEROVA, M., E. KEJNOVSKY, R. HOBZA, J. MACAS, S. R. GRANT *et al.*, 2004 Multicolor FISH mapping of the dioecious model plant, Silene latifolia. Theor Appl Genet **108:** 1193-1199.

LIM, K. Y., R. MATYASEK, C. P. LICHTENSTEIN and A. R. LEITCH, 2000 Molecular cytogenetic analyses and phylogenetic studies in the Nicotiana section Tomentosae. Chromosoma **109:** 245-258.

LIN, X., S. KAUL, S. ROUNSLEY, T. P. SHEA, M. I. BENITO *et al.*, 1999 Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana. Nature **402:** 761-768.

LIU, C., J. LIU, H. LI, Z. ZHANG, Y. HAN *et al.*, 2010 Karyotyping in melon (Cucumis melo L.) by cross-species fosmid fluorescence in situ hybridization. Cytogenet Genome Res **129:** 241-249.

LIU, Z., P. H. MOORE, H. MA, C. M. ACKERMAN, M. RAGIBA *et al.*, 2004 A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature **427:** 348-352.

LOU, Q., M. IOVENE, D. M. SPOONER, C. R. BUELL and J. JIANG, 2010 Evolution of chromosome 6 of Solanum species revealed by comparative fluorescence in situ hybridization mapping. Chromosoma **119:** 435-442.

LYONS, E., B. PEDERSEN, J. KANE, M. ALAM, R. MING *et al.*, 2008 Finding and comparing syntenic regions among Arabidopsis and the outgroups papaya, poplar, and grape: CoGe with rosids. Plant Physiol **148:** 1772-1781.

LYSAK, M. A., P. F. FRANSZ, H. B. ALI and I. SCHUBERT, 2001 Chromosome painting in Arabidopsis thaliana. Plant J **28:** 689-697.

LYSAK, M. A., A. PECINKA and I. SCHUBERT, 2003 Recent progress in chromosome painting of Arabidopsis and related species. Chromosome Res **11:** 195-204.

MA, H., P. H. MOORE, Z. LIU, M. S. KIM, Q. YU *et al.*, 2004 High-density linkage mapping revealed suppression of recombination at the sex determination locus in papaya. Genetics **166:** 419-436.

MAYER, K., C. SCHULLER, R. WAMBUTT, G. MURPHY, G. VOLCKAERT *et al.*, 1999 Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana. Nature **402:** 769-777.

MCCLINTOCK, B., 1931 The Order of the Genes C, Sh and Wx in Zea Mays with Reference to a Cytologically Known Point in the Chromosome. Proc Natl Acad Sci U S A **17:** 485-491.

MCCLINTOCK, B., and H. E. HILL, 1931 The Cytological Identification of the Chromosome Associated with the R-G Linkage Group in ZEA MAYS. Genetics **16:** 175-190.

MESSING, J., A. K. BHARTI, W. M. KARLOWSKI, H. GUNDLACH, H. R. KIM *et al.*, 2004 Sequence composition and genome organization of maize. Proc Natl Acad Sci U S A **101:** 14349-14354.

MILLA, R., and J. P. GUSTAFSON, 2001 Genetic and physical characterization of chromosome 4DL in wheat. Genome **44:** 883-892.

MING, R., A. BENDAHMANE and S. S. RENNER, 2011 Sex chromosomes in land plants. Annu Rev Plant Biol **62:** 485-514.

MING, R., S. HOU, Y. FENG, Q. YU, A. DIONNE-LAPORTE *et al.*, 2008 The draft genome of the transgenic tropical fruit tree papaya (Carica papaya Linnaeus). Nature **452:** 991-996.

MING, R., and P. H. MOORE, 2007 Genomics of sex chromosomes. Curr Opin Plant Biol **10:** 123-130.

MING, R., P. H. MOORE, F. T. ZEE, A. C. A., H. MA *et al.*, 2001a Construction and characterization of a papaya BAC library as a foundation for molecular dissection of a tree-fruit genome. Theor Appl Genet **102:** 892-899.

MING, R., P. H. MOORE, F. T. ZEE, C. A. ABBEY, H. MA *et al.*, 2001b Construction and characterization of a papaya BAC library as a foundation for molecular dissection of a tree-fruit genome. Theor Appl Genet **102:** 892-899.

MIURA, A., M. KATO, K. WATANABE, A. KAWABE, H. KOTANI *et al.*, 2004 Genomic localization of endogenous mobile CACTA family transposons in natural variants of Arabidopsis thaliana. Mol Genet Genomics **270:** 524-532.

MOORE, G., K. M. DEVOS, Z. WANG and M. D. GALE, 1995 Cereal genome evolution. Grasses, line up and form a circle. Curr Biol **5:** 737-739.

MURATA, M., J. S. HESLOP-HARRISON and F. MOTOYOSHI, 1997 Physical mapping of the 5S ribosomal RNA genes in Arabidopsis thaliana by multi-color fluorescence in situ hybridization with cosmid clones. Plant J **12:** 31-37.

NAGARAJAN, N., R. NAVAJAS-PÉREZ, M. POP, M. ALAM, R. MING *et al.*, 2008 Genome-Wide Analysis of Repetitive Elements in Papaya Tropical Plant Biol. **1:** 191-201.

NAGARAJAN, N., NAVAJAS-PEREZ, R., NIHAI, P., ALAM, M., MING, R., PATERSON, A. H., SALZBERG, S. L., 2008 Genome-wide analysis of repetitive elements in papaya. Tropical Plant Biol. **1:** 191-202.

OUYANG, S., and C. R. BUELL, 2004 The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. Nucleic Acids Res **32:** D360-363.

PATERSON, A. H., J. E. BOWERS, R. BRUGGMANN, I. DUBCHAK, J. GRIMWOOD *et al.*, 2009 The Sorghum bicolor genome and the diversification of grasses. Nature **457:** 551-556.

PEDROSA-HARAND, A., J. KAMI, P. GEPTS, V. GEFFROY and D. SCHWEIZER, 2009 Cytogenetic mapping of common bean chromosomes reveals a less compartmentalized small-genome plant species. Chromosome Res **17:** 405-417.

PEDROSA, A., C. E. VALLEJOS, A. BACHMAIR and D. SCHWEIZER, 2003 Integration of common bean (Phaseolus vulgaris L.) linkage and chromosomal maps. Theor Appl Genet **106:** 205-212.

PETERS, S. A., E. DATEMA, D. SZINAY, M. J. VAN STAVEREN, E. G. SCHIJLEN *et al.*, 2009 Solanum lycopersicum cv. Heinz 1706 chromosome 6: distribution and abundance of genes and retrotransposable elements. Plant J **58:** 857-869.

PETERSON, D. G., S. M. STACK, H. J. PRICE and J. S. JOHNSTON, 1996 DNA content of heterochromatin and euchromatin in tomato (Lycopersicon esculentum) pachytene chromosomes. Genome **39:** 77-82.

PORTIS, E., G. MAUROMICALE, R. MAURO, A. ACQUADRO, D. SCAGLIONE *et al.*, 2009 Construction of a reference molecular linkage map of globe artichoke (Cynara cardunculus var. scolymus). Theor Appl Genet **120:** 59-70.

PROJECT, I. R. G. S., 2005 The map-based sequence of the rice genome. Nature **436:** 793-800.

REN, Y., Z. ZHANG, J. LIU, J. E. STAUB, Y. HAN *et al.*, 2009 An integrated genetic and cytogenetic map of the cucumber genome. PLoS One **4:** e5795.

ROBLEDO, G., G. I. LAVIA and G. SEIJO, 2009 Species relations among wild Arachis species with the A genome as revealed by FISH mapping of rDNA loci and heterochromatin detection. Theor Appl Genet **118:** 1295-1307.

ROSATO, M., M. CASTRO and J. A. ROSSELLO, 2008 Relationships of the woody Medicago species (section Dendrotelis) assessed by molecular cytogenetic analyses. Ann Bot **102:** 15-22.

ROZEN, S., and H. SKALETSKY, 2000 Primer3 on the WWW for general users and for biologist programmers, pp. 365-386 in *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, edited by KRAWETZ S and M. S. Humana Press, NJ.

SADDER, M. T., and G. WEBER, 2001 karyotype of maize (Zea mays L.)mitotic metaphase chromosome as revealed by fluorescence in situ hybridization (FISH) with cytogenetic DNA markers. Plant mol bio rep **19:** 117-123.

SAINTENAC, C., M. FALQUE, O. C. MARTIN, E. PAUX, C. FEUILLET *et al.*, 2009 Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat (Triticum aestivum L.). Genetics **181:** 393-403.

SALANOUBAT, M., K. LEMCKE, M. RIEGER, W. ANSORGE, M. UNSELD *et al.*, 2000 Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana. Nature **408:** 820-822.

SASAKI, T., T. MATSUMOTO, K. YAMAMOTO, K. SAKATA, T. BABA *et al.*, 2002 The genome sequence and structure of rice chromosome 1. Nature **420:** 312-316.

SCHNABLE, P. S., D. WARE, R. S. FULTON, J. C. STEIN, F. WEI *et al.*, 2009 The B73 maize genome: complexity, diversity, and dynamics. Science **326:** 1112-1115.

SCHUBERT, I., P. F. FRANSZ, J. FUCHS and J. H. DE JONG, 2001 Chromosome painting in plants. Methods Cell Sci **23:** 57-69.

SHERIDAN, W. F., and D. L. AUGER, 2006 Construction and uses of new compound B-A-A maize chromosome translocations. Genetics **174:** 1755-1765.

SHISHIDO, R., Y. SANO and K. FUKUI, 2000 Ribosomal DNAs: an exception to the conservation of gene order in rice genomes. Mol Gen Genet **263:** 586-591.

SONDUR, S. N., R. M. MANSHARDT and J. I. STILES, 1996 A genetic linkage map of papaya based on randomly amplified polymorphic DNA markers Theor Appl Genet **93:** 547-553.

STACK, S. M., S. M. ROYER, L. A. SHEARER, S. B. CHANG, J. J. GIOVANNONI *et al.*, 2009 Role of fluorescence in situ hybridization in sequencing the tomato genome. Cytogenet Genome Res **124:** 339-350.

STOREY, W. B., 1953 Genetics of papaya. J Heredity **44:** 70-78.

SZINAY, D., S. B. CHANG, L. KHRUSTALEVA, S. PETERS, E. SCHIJLEN *et al.*, 2008 High-resolution chromosome mapping of BACs using multi-colour FISH and pooled-BAC FISH as a backbone for sequencing tomato chromosome 6. Plant J **56:** 627-637.

TABATA, S., T. KANEKO, Y. NAKAMURA, H. KOTANI, T. KATO *et al.*, 2000 Sequence and analysis of chromosome 5 of the plant Arabidopsis thaliana. Nature **408:** 823-826.

TAKATA, M., A. KIYOHARA, A. TAKASU, Y. KISHIMA, H. OHTSUBO *et al.*, 2007 Rice transposable elements are characterized by various methylation environments in the genome. BMC Genomics **8:** 469.

TAKETA, S., G. E. HARRISON and J. S. HESLOP-HARRISON, 1999 Comparative physical mapping of the 5S and 18S-25S rDNA in nine wild Hordeum species and cytotypes Theor Appl Genet **98:** 1-9.

TANG, X., J. M. DE BOER, H. J. VAN ECK, C. BACHEM, R. G. VISSER *et al.*, 2009a Assignment of genetic linkage maps to diploid Solanum tuberosum pachytene chromosomes by BAC-FISH technology. Chromosome Res **17:** 899-915.

TANG, X., J. M. DE BOER, H. J. VAN ECK, C. BACHEM, R. G. VISSER *et al.*, 2009b Assignment of genetic linkage maps to diploid Solanum tuberosum pachytene chromosomes by BAC-FISH technology. Chromosome Res.

TANG, X., D. SZINAY, C. LANG, M. S. RAMANNA, E. A. VAN DER VOSSEN *et al.*, 2008 Cross-species bacterial artificial chromosome-fluorescence in situ hybridization painting of the tomato and potato chromosome 6 reveals undescribed chromosomal rearrangements. Genetics **180:** 1319-1328.

THEOLOGIS, A., J. R. ECKER, C. J. PALM, N. A. FEDERSPIEL, S. KAUL *et al.*, 2000 Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana. Nature **408:** 816-820.

TIAN, Z., C. RIZZON, J. DU, L. ZHU, J. L. BENNETZEN *et al.*, 2009a Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons. Genome Res **19:** 2221-2230.

TIAN, Z., C. RIZZON, J. DU, L. ZHU, J. L. BENNETZEN *et al.*, 2009b Do genetic recombination and gene density shape the pattern of DNA elimination in rice long terminal repeat retrotransposons? Genome Res **19:** 2221-2230.

USDA-ARS, 2011, pp. in *USDA National nutrient database*.

VALARIK, M., J. BARTOS, P. KOVAROVA, M. KUBALAKOVA, J. H. DE JONG *et al.*, 2004 High-resolution FISH on super-stretched flow-sorted plant chromosomes. Plant J **37:** 940-950.

WAI, C. M., R. MING, P. H. MOORE, R. E. PAULL and Q. YU, 2010 Development of Chromosome-specific Cytogenetic Markers and Merging of Linkage Fragments in Papaya. Tropical Plant Biology **3:** 171-181.

WALLING, J. G., R. SHOEMAKER, N. YOUNG, J. MUDGE and S. JACKSON, 2006 Chromosome-level homeology in paleopolyploid soybean (Glycine max) revealed through integration of genetic and chromosome maps. Genetics **172:** 1893-1900.

WANG, C. J., L. HARPER and W. Z. CANDE, 2006a High-resolution single-copy gene fluorescence in situ hybridization and its use in the construction of a cytogenetic map of maize chromosome 9. Plant Cell **18:** 529-544.

WANG, K., B. GUAN, W. GUO, B. ZHOU, Y. HU *et al.*, 2008 Completely distinguishing individual A-genome chromosomes and their karyotyping analysis by multiple bacterial artificial chromosome - fluorescence in situ hybridization. Genetics **178:** 1117-1122.

WANG, K., W. GUO and T. ZHANG, 2007 Development of one set of chromosome-specific microsatellite-containing BACs and their physical mapping in Gossypium hirsutum L. Theor Appl Genet **115:** 675-682.

WANG, Y., X. TANG, Z. CHENG, L. MUELLER, J. GIOVANNONI *et al.*, 2006b Euchromatin and pericentromeric heterochromatin: comparative composition in the tomato genome. Genetics **172:** 2529-2540.

WATSON, J. D., and F. H. CRICK, 1953 Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature **171:** 737-738.

WEI, F., J. ZHANG, S. ZHOU, R. HE, M. SCHAEFFER *et al.*, 2009 The physical and genetic framework of the maize B73 genome. PLoS Genet **5:** e1000715.

WILSON, W. A., S. E. HARRINGTON, W. L. WOODMAN, M. LEE, M. E. SORRELLS *et al.*, 1999 Inferences on the genome structure of progenitor maize through comparative analysis of rice, maize and the domesticated panicoids. Genetics **153:** 453-473.

WRIGHT, S. I., N. AGRAWAL and T. E. BUREAU, 2003 Effects of recombination rate and gene density on transposable element distributions in Arabidopsis thaliana. Genome Res **13:** 1897-1903.

WU, C., S. WANG and H. B. ZHANG, 2006 Interactions among genomic structure, function, and evolution revealed by comprehensive analysis of the Arabidopsis thaliana genome. Genomics **88:** 394-406.

YIN, T., S. P. DIFAZIO, L. E. GUNTER, X. ZHANG, M. M. SEWELL *et al.*, 2008 Genome structure and emerging evidence of an incipient sex chromosome in Populus. Genome Res **18:** 422-430.

YU, Q., S. HOU, F. A. FELTUS, M. R. JONES, J. E. MURRAY *et al.*, 2008 Low X/Y divergence in four pairs of papaya sex-linked genes. Plant J **53:** 124-132.

YU, Q., S. HOU, R. HOBZA, F. A. FELTUS, X. WANG *et al.*, 2007a Chromosomal location and gene paucity of the male specific region on papaya Y chromosome. Mol Genet Genomics **278:** 177-185.

YU, Q., E. TONG, R. L. SKELTON, J. E. BOWERS, M. R. JONES *et al.*, 2009 A physical map of the papaya genome with integrated genetic map and genome sequence. BMC Genomics **10:** 371.

YU, W., J. C. LAMB, F. HAN and J. A. BIRCHLER, 2007b Cytological visualization of DNA transposons and their transposition pattern in somatic cells of maize. Genetics **175:** 31-39.

ZHANG, D., Q. YANG, W. BAO, Y. ZHANG, B. HAN *et al.*, 2005 Molecular cytogenetic characterization of the Antirrhinum majus genome. Genetics **169:** 325-335.

ZHANG, W., C. M. WAI, R. MING, Q. YU and J. JIANG, 2010 Integration of Genetic and Cytological Maps and Development of a Pachytene Chromosome-based Karyotype in Papaya. Tropical Plant Biology **3:** 166-170.

ZHANG, W., X. WANG, Q. YU, R. MING and J. JIANG, 2008 DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya. Genome Res **18:** 1938-1943.

ZHONG, X. B., P. F. FRANSZ, J. WENNEKES-EDEN, M. S. RAMANNA, A. VAN KAMMEN *et al.*, 1998 FISH studies reveal the molecular and chromosomal organization of individual telomere domains in tomato. Plant J **13:** 507-517.