

# Chapter 10

## Physical Map of Papaya Genome

Qingyi Yu

### Introduction

The high-throughput and low-cost next-generation sequencing technologies make genome sequencing possible for non-model species. The typical approaches of genome sequencing include either “BAC-by-BAC” or “whole-genome shotgun sequencing” approaches. A high-quality physical map plays essential roles in both sequencing strategies (The Arabidopsis Genome Initiative 2000; Schnable et al. 2009; The Tomato Genome Consortium 2012; Yu et al. 2002). The key step in “BAC-by-BAC” sequencing projects is to identify BAC clones on the “minimal tiling path,” which heavily relies on an accurate BAC-based physical map. In whole-genome shotgun sequencing projects, a physical map, in combination with BAC end sequences (BES), provides anchor points to build scaffolds, especially for large repeat regions. Physical maps provide a valuable genomic resource not only for genome sequence assembly but also for studying genome structure and organization and positional cloning of genes associating with economically and agronomically important traits. In addition, physical maps can be used for comparative genomics studies of closely related species to enhance our understanding of the evolutionary process.

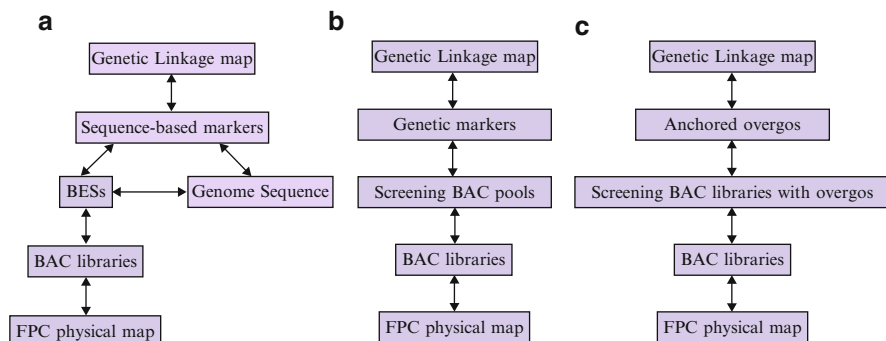
Genetic and physical maps are constructed using independent methods and are not directly related to each other. Various methods have been developed to integrate the genetic and physical maps, including (1) genetically mapping BAC-anchored markers (Yu et al. 2009; Troggio et al. 2007; Mun et al. 2006), (2) BAC pooling with PCR screening (Klein et al. 2000), and (3) hybridization of overgo probes (Yüksel et al. 2005) (Fig. 10.1a–c).

Among these three methods, genetically mapping BAC-anchored markers is the most commonly used. Genetic and physical maps are independent genomic

---

Q. Yu (✉)

Texas A&M AgriLife Research Center at Dallas, Texas A&M University System,  
17360 Coit Road, Dallas, TX 75252, USA  
e-mail: qyu@ag.tamu.edu



**Fig. 10.1** The three major methods for integrating genetic and physical maps. (a) Genetically mapping BAC-anchored markers. (b) BAC pooling with PCR screening. (c) Hybridization of overgo probes

**Table 10.1** Comparison of the three papaya genetic maps

	AFLP-based map	SSR-based map	SSR-AFLP map
Population	Kapoho × SunUp	AU9 × SunUp	AU9 × SunUp
Marker types	AFLP	SSR	SSR and AFLP
No. of markers	1,501	707	990
Length of map (cM)	3,294.2	1,069.9	945.2
No. of linkage groups	12	12	14
Reference	Ma et al. (2004)	Chen et al. (2007)	Blas et al. (2009)

resources and complementary to one another. The integrated map could minimize errors and enhance the quality of these resources. The integrated maps with BES allow assignment of genome sequence contigs on linkage groups. The integrated genomic resources provide a simplified and efficient interface for data mining and bioinformatics analysis.

Significant progress has been made in recent years in developing genomic resources to expedite genetic research for improvement of papaya. Three bacterial artificial chromosome (BAC) libraries, one for each sex type, were constructed, providing the foundation for studying papaya genome structure and organization (Ming et al. 2001; Gschwend et al. 2011). BAC ends of the entire hermaphrodite papaya BAC library were sequenced (Lai et al. 2006; Ming et al. 2008; Yu et al. 2009), providing the first glimpse of the sequence composition of the papaya genome. Three high-density genetic linkage maps have been constructed (Ma et al. 2004; Chen et al. 2007; Blas et al. 2009) (Table 10.1), providing essential tools for comparative genomic analysis, marker-assisted selection, and genomic dissection of complex traits. The first high-density genetic map of papaya was constructed with 1,501 markers, including 1,498 amplified fragment length polymorphism (AFLP) markers, 2 morphological markers, and one transgenic marker (Ma et al. 2004). Due to the anonymous nature of AFLP markers, this map cannot be used to

integrate the genetic and physical maps and is not suitable for aligning papaya genome sequence to linkage groups. To overcome this limitation, highly informative sequence-based simple sequence repeat (SSR) markers were used to construct the second high-density genetic map. This sequence-tagged genetic map contains 707 markers including 706 sequence-based SSR markers and one morphological marker (Chen et al. 2007). The SSR markers were developed from either BES or whole-genome shotgun sequence reads (Chen et al. 2007); and this map has contributed to the integration of genetic and physical maps and the genome sequences (Yu et al. 2009). A BAC-based physical map of papaya was constructed using the high-information-content fingerprinting method (Yu et al. 2009). The draft genome sequence of papaya was published in 2008 and contains 271 Mb of contig sequence and 370 Mb of scaffolds with embedded gaps (Ming et al. 2008). The physical map was integrated with the genetic map and genome sequence through mapping micro-satellite markers derived from BES and the whole-genome shotgun sequences (Yu et al. 2009). This integrated map provides essential information for studying the papaya genome structure and comparative genomics.

## Papaya BAC Libraries

Bacterial artificial chromosome (BAC) libraries, containing large genomic fragments with deep coverage, have served as valuable research tools for physical mapping, positional cloning, and genome sequencing. The first papaya BAC library was constructed from a hermaphrodite plant of inbred cultivar “SunUp” using *Hind*III partial digestion (Ming et al. 2001) (Table 10.2). This BAC library consists of 39,168 clones with an average insert size of 132 kb, providing 13.7× papaya genome equivalents (Ming et al. 2001). This BAC library has been used for cloning flower development (Yu et al. 2005, 2008; Ackerman et al. 2008) and fruit flesh-color genes (Skelton et al. 2006; Blas et al. 2010) and characterization of transgenic insertions in genetically modified papaya (Suzuki et al. 2008). The entire library was end sequenced (Lai et al. 2006; Yu et al. 2009). After trimming the vector sequences and removing the clones with cross-well contamination, a total of 67,179 BES remained with an average length of 666 bp. The total length of the cleaned BES is 44,725,370 bp, accounting for 12 % of the papaya genome. The BES played essential roles in integration of genetic and physical maps and in assembling the genome sequence.

**Table 10.2** The characteristics of the three papaya BAC libraries

	Hermaphrodite BAC lib.	Female BAC lib.	Male BAC lib.
Genotype	SunUp, hermaphrodite	SunUp, female	AU9, male
Restriction enzyme	<i>Hind</i> III	<i>Bst</i> YI	<i>Eco</i> RI
No. of clones	39,168	36,864	55,296
Average insert size (kb)	132	104	101
Genome equivalents	13.7	10.3	15.0

Two additional papaya BAC libraries, one for “SunUp” female and the other for “AU9” male plants (Table 10.2), were constructed to assist physical mapping of the male-specific Y chromosome (MSY) region and its corresponding region on the X chromosome (Gschwend et al. 2011). These two independent BAC libraries used two different restriction enzymes, *EcoRI* and *BstYI*, to increase representative genome coverage. The female BAC library consists of 36,864 clones with an average insert size of 104 kb, providing 10.3× genome equivalents (Gschwend et al. 2011). The male BAC library consists of 55,296 clones with an average insert size of 101 kb, providing 15.0× genome equivalents (Gschwend et al. 2011). The three papaya BAC libraries of three sex types provided essential genomic resources for physical mapping of the hermaphrodite-specific region of the Y<sup>h</sup> chromosome (HSY), the male-specific region of the male Y chromosome (MSY), and the corresponding region on the X chromosome (Na et al. 2012).

## BAC Fingerprinting and Contig Assembly

The basic strategy for construction of a physical map is to identify overlapping clones based on shared restriction fragments or DNA markers. Various methods have been developed to create physical maps. Briefly, these methods can be grouped into either PCR/hybridization-based or DNA fingerprinting-based approaches. Compared with PCR/hybridization-based methods, fingerprinting-based methods are rapid and better suited to map at a whole-genome level. The high-information-content fingerprinting (HICF) method has proved to be the most effective method owing to its high resolution and high throughput (Luo et al. 2003).

The entire library of the hermaphrodite papaya BAC library was fingerprinted using the HICF method. After the 1st round of fingerprinting, the failed clones were repeated twice to obtain successful fingerprints for 38,522 BAC clones (98.4 % of the BAC library). After excluding the clones with no or small inserts (referred as less than 20 true fragments after editing), and cross-contamination (70 % or higher shared fragments for neighboring clones), the remaining 30,824 fingerprints (78.7 % of the BAC library) were subjected to contig assembly using the FPC (FingerPrinted Contigs) program (Soderlund et al. 1997).

A total of 26,466 BAC clones were assembled into 963 contigs, while 4,358 clones remained as singletons. The average number of fragments (bands) of each clone was 69.4. Considering the clones with no or small inserts were excluded from contig assembly and the average insert size of the BAC clones used for contig assembly should be higher than 132 kb, the BAC clones mapped on contigs (26,466 BAC clones) represented at least 9.4× genome equivalents of the papaya genome. On average, each contig contains 27.5 BAC clones. The longest contig contains 1,571 consensus bands, which is about 0.7 % of the total length of the FPC contigs.

## Overgo and Single-Copy Probe Hybridization

Overgo probes consist of two complementary oligonucleotides that can anneal at the 3'-end to generate a double-stranded probe with labeled deoxynucleotides (Ross et al. 1999). Overgo probes are designed to bind to conserved regions between two genomes at regular intervals. The overgo hybridization system was developed for large-scale BAC library screening in a high-throughput format to anchor EST sequences, or genetic markers, on physical maps. Overgo hybridization has been successfully used for integration of genetic and physical maps in cotton, cacao and peanut (Lin et al. 2010; Saski et al. 2011; Yüksel et al. 2005).

Papaya and *Arabidopsis* belong to the order Brassicales and shared the last common ancestor approximately 72 MYA (Wikström et al. 2001). Unlike *Arabidopsis*, papaya has not undergone recent whole-genome duplication events (Ming et al. 2008; Lyons et al. 2008a, b; Tang et al. 2008). To facilitate comparative and evolutionary genomic study within Brassicales, a total of 2,277 overgo probes representing conserved sequences in *Arabidopsis* and genetically mapped Brassica loci were tested against 36,864 papaya BACs. A total of 1,329 overgos (58 % of overgos designed) detected positive BACs in papaya. After eliminating low-quality data, the remaining 1,181 overgos were anchored on the papaya FPC map. Among these overgos, 756 (64.0 %) hit single contigs; the average number of clones per overgo is 6.0.

In addition to overgo probes, 16 probes designed from single-copy loci of papaya were used to screen the papaya BAC library. A total of 153 positive BAC clones were obtained. These 16 single-copy probes were anchored on the FPC map. All the probes except papaya *Pistillata* (*PI*) gene hit single contigs. The probe of papaya *Pistillata* (*PI*) gene hit two contigs, ctg1350 and ctg577. Ctg577 contains only three clones and all of them are PI positive. Thus, ctg577 should be part of the ctg1350, but the high stringency setting failed to merge them into a single contig.

These anchored overgos and single-copy genes further improved the quality of the physical map and provide direct links among papaya, *Arabidopsis*, and *Brassica* genomes for comparative genomic research. These anchored overgos and single-copy genes could help reveal synteny and rearrangements in target regions of these genomes, particularly in these recently duplicated genomes of *Arabidopsis* and *Brassica*.

## Integration of Physical Map, Genetic Map, and Genome Sequence

Physical map, genetic map, and genome sequence are independent genomic resources. Among these resources, physical map forms an intermediate between genetic map and genome sequence. Integration of physical map, genome sequence, and genetic map could help to correct errors and enhance the quality of each of these resources that are complementary to one another. Integration of these three sets of

**Table 10.3** Summary of the integrated genetic and physical map and genome sequence

Number of clones with successful fingerprints	30,824
Number of clones on FPC contigs	26,466
Number of FPC contigs	963
Number of BAC end sequences	67,179
Number of overgos anchored on the FPC contigs	1,181
Mapped SSR markers	706
FPC contigs anchored on the genetic linkage map	535
Scaffolds anchored on the genetic linkage map	255
Genome coverage of the FPC physical map	95.8 %

genomic information provides a valuable resource for comparative genomics, particularly for analyses of Brassicales (Freeling et al. 2008).

The papaya genetic and physical maps were integrated by mapping microsatellite markers derived from BES and whole-genome shotgun sequences (Yu et al. 2009). The sequence-tagged high-density genetic map of papaya was constructed by mapping 706 microsatellite markers (Chen et al. 2007). The resulting genetic map consists of nine major and three minor linkage groups. The three minor linkage groups, LGs 10–12, were merged with major linkage groups, LGs 8, 9, and 7, respectively, based on fluorescence in situ hybridization (FISH) (Wai et al. 2010).

Among the 706 mapped SSR markers, 153 were derived from BES, 466 were designed from shotgun sequence reads, and 87 were developed from assembled shotgun contigs. The BES-derived SSRs directly anchored the associated FPC contigs to the genetic map. Using the 153 SSRs derived from BESs, 122 FPC contigs containing 46,475 consensus bands and 97 shotgun scaffolds covering 132 Mb (35.5 % of the papaya genome) were anchored to the genetic map. The initial integration was extended by aligning BESs to the draft genome sequences. The remaining 553 microsatellites were positioned on shotgun scaffolds by blastn. The BAC clones with paired ends covering the mapped microsatellites were used as bait to search the FPC physical map to find the corresponding FPC contigs. The order and orientation of the FPC contigs were validated by aligning associated BESs on the shotgun scaffolds.

In the final integrated map, 55.6 % FPC contigs (535 FPC contigs) and 81.2 % BAC clones (21,371 BACs) were anchored on the genetic linkage map. A total of 255 shotgun scaffolds covering 233 Mb were anchored to the genetic map. Overall, 63 % of the papaya genome sequences were placed on the genetic map. The summary of the integrated map is shown in Table 10.3.

## Integration of Genetic and Cytogenetic Maps

Papaya chromosomes are relatively small and have similar size and morphology at metaphase stage (Storey 1953). Thus, it is difficult to identify individual papaya chromosomes using traditional karyotyping methods, such as banding pattern, arm ratio, morphology, and size. A set of chromosome-specific cytogenetic markers

**Table 10.4** The physical size of papaya pachytene chromosomes and the length of their corresponding linkage groups

Chromosome	Total length ( $\mu\text{m}$ )	Relative length (%)	Linkage group	Total length (cM)
X/Y	55.70 $\pm$ 8.78	12.30 $\pm$ 2.74	LG1	145.0
2	58.08 $\pm$ 11.84	12.82 $\pm$ 3.70	LG9+LG11	91.3
3	53.49 $\pm$ 8.28	11.81 $\pm$ 2.59	LG6	100.2
4	51.69 $\pm$ 8.74	11.41 $\pm$ 2.73	LG2	138.8
5	51.59 $\pm$ 9.71	11.39 $\pm$ 3.03	LG5	103.6
6	51.31 $\pm$ 10.81	11.33 $\pm$ 3.38	LG3	132.4
7	47.57 $\pm$ 9.11	10.50 $\pm$ 2.85	LG7+LG12	117.8
8	41.78 $\pm$ 8.48	9.23 $\pm$ 2.65	LG4	120.6
9	41.67 $\pm$ 6.34	9.20 $\pm$ 1.98	LG8+LG10	118.9

were developed to assist papaya chromosome identification and cytological analysis (Wai et al. 2010). The chromosome-specific cytogenetic markers were derived from the integrated genetic and physical maps. A total of 104 BAC clones containing SSR markers distributed on 12 linkage groups were tested for signals on papaya chromosomes using FISH. Fifty-four BACs producing repeatable and unambiguous FISH signals were selected as chromosome-specific markers (Wai et al. 2010). This set of chromosome-specific markers served as a foundation for papaya chromosome karyotyping and assigning linkage groups to chromosomes.

The papaya SSR genetic map consists of nine major linkage groups (LGs 1–9) and three minor linkage groups (LGs 10–12) (Chen et al. 2007). The nine major linkage groups were assigned to each individual chromosome by hybridizing LG-specific markers on papaya chromosomes (Wai et al. 2010; Zhang et al. 2010). The three minor linkage groups, LGs 10–12, were assigned to the major linkage groups by simultaneously labeling cytogenetic markers derived from the minor linkage groups with the LG-specific markers developed from the nine major linkage groups. Using this strategy, LGs 10–12 were assigned to the major LGs 8, 9, and 7, respectively (Wai et al. 2010).

A pachytene chromosome-based karyotype of papaya was developed with the assistance of chromosome-specific cytogenetic markers (Zhang et al. 2010). The chromosomes were numbered based on their length except X/Y chromosome pair (Table 10.4). In general, the papaya chromosomes have similar sizes as they were observed at metaphase stage (Table 10.4). The longest chromosome is about 58.08  $\mu\text{m}$  and accounts for 12.8 % of the total length of papaya chromosomes, while the shortest chromosome is about 41.67  $\mu\text{m}$  and accounts for 9.20 % of the total length of papaya chromosomes. The difference between the longest and shortest chromosomes is about 28 %.

The papaya genome is composed largely of euchromatin distributed mainly on chromosome arms (Ming et al. 2008; Zhang et al. 2010). Approximately 83 % of the papaya genome is euchromatic, and 17 % is heterochromatic (Ming et al. 2008). The DAPI brightly stained heterochromatin is concentrated in the centromeric and pericentromeric regions (Zhang et al. 2010), which may explain the recombination suppression found in the middle of most linkage groups (Chen et al. 2007).



The nucleolus organizing regions (NOR), one of the major chromosome structure features, were anchored on the papaya genetic linkage map by FISH (Wai et al. 2010; Zhang et al. 2010). The 5S rDNA loci were located on chromosome 3, 5, 8, 9, and the Y chromosome (Zhang et al. 2010). The strongest signals of 5S rDNA were found at interstitial positions of chromosomes 5 and 8 (Wai et al. 2010; Zhang et al. 2010). The 25S rDNA was located at constriction sites of chromosome 4 (Wai et al. 2010; Zhang et al. 2010).

## Genome Coverage

The genome coverage of the fingerprinted physical map was evaluated by estimating the average consensus band size and excluding the organelle genome contamination. BAC clones containing papaya chloroplast genome sequences were identified by hybridizing sorghum chloroplast *ropB* and *trnK* probes against papaya BAC grids (Yu et al. 2009). By searching the BES using the papaya chloroplast genome sequence as a query, additional BAC clones containing the papaya chloroplast genome were identified. The positive clones were scanned against the papaya FPC map, and FPC contigs 972 and 426 were identified as containing papaya chloroplast genome sequence.

The BAC clones containing papaya mitochondrial genome sequences were identified by searching the papaya BES database. A total of 356 BAC clones were identified with at least one end sharing over 95 % identity with the papaya mitochondrial genome sequence. Among them, 144 BACs were verified with both ends containing the papaya mitochondrial genome sequences. FPC contigs 867, 524, 1,172, 553, and 47 were identified as containing the papaya mitochondrial genome fragments.

Lengths of fingerprinted contigs are analyzed as consensus band (CB) units. To estimate the genome coverage of the fingerprinted physical map, the average band size was determined on 22 randomly selected, non-overlapping contigs (Yu et al. 2009). Based on the length of these 22 FPC contigs of 7,503 CB units and their physical size of 12,256,987 bp in the shotgun assembly, the average band size of the papaya FPC map was estimated at about 1.6 kb.

The total length of the papaya FPC map is 224,354 CB units. After excluding the FPC contigs containing papaya organelle genome fragments, the total length of the remaining contigs was 222,808 CB units and approximately equal to 356.5 Mb. Considering the genome size of papaya at 372 Mb (Arumuganathan and Earle 1991), the papaya physical map covers about 95.8 % of the papaya genome, and 72.4 % of papaya genome (269.15 Mb in total length) was anchored to the genetic map.

## Genetic Recombination

Genetic maps are constructed by using recombination frequencies between markers. Thus, the distance between two loci on a genetic map reflects their recombination frequency, instead of an actual physical distance. On the other hand, a physical map



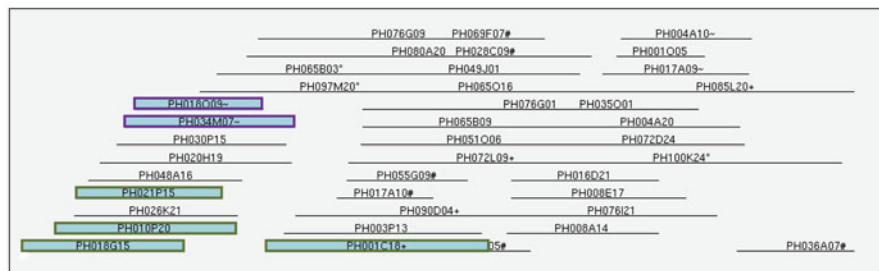
is developed using restriction enzyme fingerprinting data which reflect the relative physical location of overlapped clones. The integrated genetic and physical map makes it possible to reveal the relationship between physical distance and genetic distance at the genome level.

The ratio between genetic and physical distance on the integrated map varies among chromosomes. Most recombination suppressed regions were observed near the center of linkage groups, while most recombination hot spots were found toward the ends of linkage groups. DAPI (4',6-diamidino-2-phenylindole)-stained papaya chromosomes revealed highly condensed heterochromatin knobs located in the centromeric and pericentromeric regions of all nine pairs of chromosomes (Ming et al. 2008). Apparent lengths of the condensed heterochromatin knobs varied among the chromosomes (Ming et al. 2008). Consistent with papaya chromosome structure, recombination suppression regions were found near the center of all the major linkage groups, suggesting that those regions might be centromeric. Higher recombination rates observed toward the telomeric regions could reflect these regions favored by natural selection to contain beneficial alleles.

Extensively suppression of recombination was observed at the male-specific Y chromosome region (MSY) on LG1. The size of the suppressed region was estimated at about 8–9 Mb based on the integrated genetic and physical map. Based on pachytene FISH images, the physical location of the MSY is in the middle of the Y chromosome (Zhang et al. 2008), whereas the MSY was located on the upper half of LG1 on the genetic map (Chen et al. 2007). This discrepancy in locating the MSY is likely caused by the lower recombination rate on one arm of the Y chromosome (Chen et al. 2007; Wai et al. 2012). Also, pachytene FISH images consistently revealed denser heterochromatic regions on one of the Y chromosome arms (Ming et al. 2008). Suppression of recombination has spread to adjacent regions of the MSY, but recombination rates recovered gradually and then increased dramatically at about 10 Mb from the MSY.

## **Map-Based Cloning of the Major Gene Controlling Fruit Flesh Color in Papaya**

Carotenoid antioxidants, which are also called provitamin A, are precursors of vitamin A that carries out a number of functions in the human body including serving as antioxidants, precursors of steroid hormones involved in growth and differentiation, and production of visual proteins (Paiva and Russell 1999; Krinsky 1998; Bartley and Scolnik 1995). Papaya is one of the top fruits that are rich in vitamin A and vitamin C (Liebman 1992). Promoting papaya consumption in underdeveloping and developing tropical and subtropical countries will significantly improve health of poor people as golden rice projected (Dawe et al. 2002). In papaya, the flesh color of the fruit is considered as a quality trait that is often targeted for papaya improvement because it correlates with nutritional value and is linked to shelf life of the fruit (Yamamoto 1964). Two fruit flesh colors, yellow and red, of papaya fruit are present



**Fig. 10.2** FPC Ctg 962 of the papaya BAC-based physical map. The *purple color box* highlights the BAC clones identified by the cDNA probe of the tomato *CYC-B* gene. The *green color box* highlights the BAC clones identified by the genetic marker linked with *flesh color*

in the papaya germplasm. Red-fleshed papaya is preferred by some consumers, but it is associated with a faster rate of softening thus a shorter shelf life.

The fruit flesh color of papaya is controlled by a single gene with yellow color as dominant (Hofmeyr 1938; Storey 1969). It has been previously documented that yellow-fleshed fruits have high amounts of the carotenoid  $\beta$ -cryptoxanthin and  $\zeta$ -carotene but lack lycopene, while red-fleshed fruits have high amounts of lycopene along with  $\beta$ -cryptoxanthin,  $\beta$ -carotene, and  $\zeta$ -carotene (Karrer and Jucker 1950; Yamamoto 1964; Chandrika et al. 2003). The red color of papaya fruit is due to the accumulation of lycopene; the yellow color is the result of converting lycopene to  $\beta$ -carotene and  $\beta$ -cryptoxanthin (Yamamoto 1964).

The high-density AFLP genetic map of papaya was constructed using 54F2 plants derived from cultivars Kapoho and SunUp (Ma et al. 2004). The map consists of 1,501 markers, including 1,498 amplified fragment length polymorphism (AFLP) markers, the papaya ringspot virus coat protein marker, morphological sex type, and fruit flesh color (Ma et al. 2004). The gene controlling the fruit flesh color was placed on linkage group 7 where it is flanked by two AFLP markers at recombination distances 3.4 and 3.7 cM (Ma et al. 2004). These closely linked AFLP markers were converted to sequence-characterized amplified region (SCAR) markers and used as probes to screen the papaya BAC library (Blas et al. 2010). Meanwhile, candidate gene approaches were carried out to clone the gene controlling the fruit flesh color. A tomato *CYC-b* cDNA was used as a probe to screen the papaya BAC library. Five BACs were identified by the SCAR marker and two by the tomato *CYC-b* cDNA probe. These seven positive BACs were used as seed BACs to search into papaya BAC-based physical map. Six of the positive BACs were mapped to a single fingerprint contig FPC-962 of the papaya physical map (Fig. 10.2). Co-localization of the fruit flesh-color-linked SCAR marker and the *CYC-b* cDNA probe on a single contig indicated that the papaya *CYC-b* orthologous gene was a strong candidate for controlling fruit flesh color. The papaya draft genome sequence integrated with FPC-962 was used to identify the papaya homologous gene of tomato *CYC-b* (named as *cp CYC-b*).

The two alleles, the dominant yellow-fleshed allele and recessive red-fleshed allele, were fully sequenced (Blas et al. 2010). A 2 bp insertion was identified within

the red-fleshed *CpCYC-b* coding region, which produces a frameshift mutation resulting in a premature stop codon and a truncated coding region in red-fleshed *CpCYC-b* allele (Blas et al. 2010). The function of *CpCYC-b* alleles was tested by transformation into lycopene-accumulating *E. coli* strain (Cunningham et al. 1994). The gene product of the full-length yellow-fleshed *CpCYC-b* was able to mediate conversion of lycopene (red) to  $\beta$ -carotene (yellow) (Blas et al. 2010). While the red-fleshed *CpCYC-b* lost the lycopene  $\beta$ -cyclase activity as indicated by the lack of color change (Blas et al. 2010).

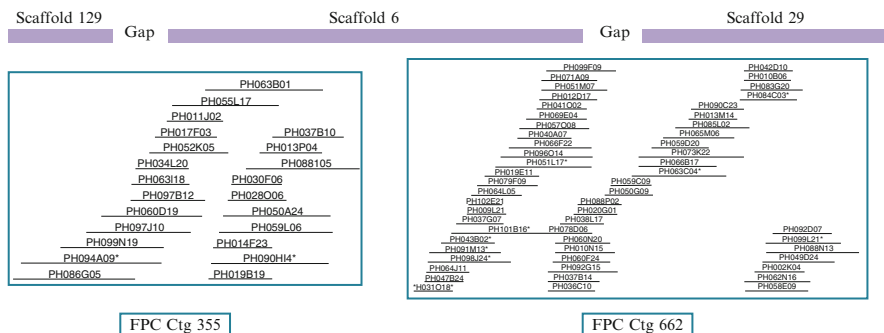
A SCAR marker located at 580 bp away from *CpCYC-b* gene was developed to support marker-assisted selection in papaya breeding programs. Using this marker, a simple PCR-based screening test using agarose gels can identify individuals in a segregating population with the desired fruit flesh color (Blas et al. 2010). Due to the extremely high recombination rate in the surrounding region, this simple PCR test will be able to identify fruit flesh-color genes with approximately 98 % certainty (Blas et al. 2010). A more precise test could be done at a higher cost and lower throughput using acrylamide sequencing gels by targeting the two nucleotides insertion of the red flesh *CpCYC-b* (Blas et al. 2010).

## Perspective

The three major genomic resources, genetic map, physical map, and genome sequence, complement one another and correct errors from each individual source. The integrated information enhances the capacity for map-based cloning, identification of underlying genes controlling quantitative traits in papaya, and comparative genomics studies.

Although the draft genome sequence of papaya covers 92 % of the genes (Ming et al. 2008), most quantitative trait loci (QTLs) controlling economically and agronomically important traits in papaya have not yet been characterized, including those controlling sugar content, fruit size, shape, weight, and complex biotic and abiotic stress reactions. The integrated map in combination with genome sequence will expedite the mapping and cloning of target genes to facilitate papaya breeding through marker-assisted selection.

The current draft sequence of the papaya genome represents about 75 % of the papaya genome (Ming et al. 2008). With rapid progress in next-generation sequencing technology, finishing the papaya genome may be carried out in the foreseeable future. The papaya BAC-based physical map integrated with the genome sequences and genetic map will be an essential resource for closing gaps of any particular genomic region under investigation. The BAC clones located at the ends of the scaffolds can be used as baits to search the FPC physical map. For example, the gap between scaffolds 6 and 129 of the draft genome can be filled by FPC Ctg 355 (Fig. 10.3). The BAC clones on the minimal tilling path can be sequenced to fill the targeted gaps. Moreover, the integrated genetic and physical map could also provide a framework to guide the genome sequencing of related species.



**Fig. 10.3** With assistance from BAC end sequences and BAC-based physical map, the gap between shotgun scaffolds 129 and 6 was filled by FPC Ctg 355, and the gap between shotgun scaffolds 6 and 29 was filled by FPC Ctg 662

The integrated genetic and physical map can serve as a reference to study gene and genome evolution and to reveal the genetic base of unique ecological adaptations. A total of 1,181 overgos representing conserved sequences of *Arabidopsis* and genetically mapped *Brassica* loci were anchored on the integrated genetic and physical map and the draft genome sequence of papaya. These overgos provide direct links among papaya, *Arabidopsis*, and *Brassica* genomes for comparative genomic research among species within the order Brassicales. The overgos were designed from single-copy genes and sequences of *Arabidopsis* and *Brassica*. Along with the FPC contigs and WGS sequences, overgo markers could help identify synteny and rearrangements in target regions of these genomes, particularly in these recently duplicated genomes of *Arabidopsis* and *Brassica*.

## References

- Ackerman CM, Yu Q, Kim S, Paull RE, Moore PH, Ming R (2008) Differential expression of B-class MADS-box genes in trioecious papaya flowers. *Planta* 227:741–753
- Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9:208–218
- Bartley GE, Scolnik PA (1995) Plant carotenoids: pigments for photoprotection, visual attraction, and human health. *Plant Cell* 7:1027–1038
- Blas AL, Yu Q, Chen C, Veatch O, Moore PH, Paull RE, Ming R (2009) Enrichment of a papaya high-density genetic map with AFLP markers. *Genome* 52:716–725
- Blas AL, Ming R, Liu Z, Veatch OJ, Paull RE, Moore PH, Yu Q (2010) Cloning of papaya chloroplast specific lycopene  $\beta$ -cyclase, CpCYC-b, controlling fruit flesh color reveals conserved microsynteny and a recombination hotspot. *Plant Physiol* 152:2013–2022
- Chandrika UG, Jansz ER, Wickramasinghe SMDN, Warnasuriya ND (2003) Carotenoids in yellow- and red-fleshed papaya (*Carica papaya* L.). *J Sci Food Agric* 83:1279–1282
- Chen C, Yu Q, Hou S, Li Y, Eustice M, Skelton RL, Veatch O, Herds R, Diebold L, Saw J, Feng Y, Bynum L, Wang L, Moore PH, Paull RE, Alam M, Ming R (2007) Construction of a sequence-tagged high-density genetic map of papaya for comparative structural and evolutionary genomics in brassicales. *Genetics* 177:2481–2491

- Cunningham FX Jr, Sun Z, Chamovitz D, Hirschberg J, Gantt E (1994) Molecular structure and enzymatic function of lycopene cyclase from the cyanobacterium *Synechococcus* sp strain PCC7942. *Plant Cell* 6:1107–1121
- Dawe D, Robertson R, Unnevehr L (2002) Golden rice: what role could it play in alleviation of vitamin A deficiency? *Food Policy* 27:541–560
- Freeling M, Lyons E, Pedersen B, Alam M, Ming R, Lisch D (2008) Many or most genes in *Arabidopsis* transposed after the origin of the order Brassicales. *Genome Res* 18:1924–1937
- Gschwend AR, Yu Q, Moore P, Sasaki C, Chen C, Wang J, Na J-K, Ming R (2011) Construction of papaya male and female BAC libraries and application in physical mapping of the sex chromosomes. *J Biomed Biotechnol* 2011:929472
- Hofmeyr JDJ (1938) Genetical studies of *Carica papaya* L.I. The inheritance and relation of sex and certain plant characteristics. II Sex reversal and sex forms. South African Department of Agriculture and Science Bulletin, no. 187
- Initiative TAG (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
- Karrer P, Jucker E (1950) Carotenoids. Elsevier, New York
- Klein PE, Klein RR, Cartinhour SW, Ulanich PE, Dong J, Obert JA, Morishige DT, Schlueter SD, Childs KL, Ale M, Mullet JE (2000) A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. *Genome Res* 10:789–807
- Krinsky NI (1998) Overview of lycopene, carotenoids, and disease prevention. *Exp Biol Med* 218:95–97
- Lai CWJ, Yu Q, Hou S, Skelton RL, Jones MR, Lewis KLT, Murray J, Eustice M, Guan P, Agbayani R, Moore PH, Ming R, Presting GG (2006) Analysis of papaya BAC end sequences reveals first insights into the organization of a fruit tree genome. *Mol Genet Genomics* 276:1–12
- Liebman B (1992) Nutritional aspects of fruit. *Nutr Action Newslett* (May 10–11)
- Lin L, Pierce GJ, Bowers JE, Estill JC, Compton RO, Rainville LK, Kim C, Lemke C, Rong J, Tang H, Wang X, Braidotti M, Chen AH, Chicola K, Collura K, Epps E, Golser W, Grover C, Ingles J, Karunakaran S, Kudrna D, Olive J, Tabassum N, Um E, Wissotski M, Yu Y, Zuccolo A, ur Rahman M, Peterson DG, Wing RA, Wendel JF, Paterson AH (2010) A draft physical map of a D-genome cotton species (*Gossypium raimondii*). *BMC Genomics* 11:395
- Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput finger-printing of bacterial artificial chromosomes using the SNaP-shot labeling kit and sizing of restriction fragments by capillary electrophoresis. *Genomics* 82:378–389
- Lyons E, Pedersen B, Kane J, Alam M, Ming R, Tang H, Wang X, Bowers J, Paterson A, Lisch D, Freeling M (2008a) Finding and comparing syntenic regions among *Arabidopsis* and the outgroups papaya, poplar, and grape: CoGe with Rosids. *Plant Physiol* 148:1772–1781
- Lyons E, Pedersen B, Kane J, Freeling M (2008b) The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the Rosids. *Trop Plant Biol* 1:181–190
- Ma H, Moore PH, Liu Z, Kim MS, Yu Q, Fitch MMM, Sekioka T, Paterson AH, Ming R (2004) High-density linkage mapping revealed suppression of recombination at the sex determination locus in papaya. *Genetics* 166:419–436
- Ming R, Moore PH, Zee F, Abbey CA, Ma H, Paterson AH (2001) 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, Hou S, Feng Y, Yu Q, Dionne-Laporte A, Saw JH, Senin P, Wang W, Ly BV, Lewis KLT, Salzberg SL, Feng L, Jones MR, Skelton RL, Murray JE, Chen C, Qian W, Shen J, Du P, Eustice M, Tong E, Tang H, Lyons E, Paull RE, Michael TP, Wall K, Rice D, Albert H, Wang M-L, Zhu YJ, Schatz M, Nagarajan N, Agbayani R, Guan P, Blas A, Wai CM, Ackerman CM, Ren Y, Liu C, Wang J, Wang J, Na J-K, Shakirov EV, Haas B, Thimmapuram J, Nelson D, Wang X, Bowers JE, Gschwend AR, Delcher AL, Singh R, Suzuki JY, Tripathi S, Neupane K, Wei H, Irikura B, Paidi M, Jiang N, Zhang W, Presting G, Windsor A, Navajas-Pérez R, Torres MJ, Alex FF, Porter B, Li Y, Burroughs AM, Luo M-C, Liu L, Christopher DA, Mount SM,

- Moore PH, Sugimura T, Jiang J, Schuler MA, Friedman V, Mitchell-Olds T, Shippen DE, de Pamphilis CW, Palmer JD, Freeling M, Paterson AH, Gonsalves D, Wang L, Alam M (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452:991–996
- Mun J-H, Kim D-J, Choi H-K, Gish J, Debellé F, Mudge J, Denny R, Endré G, Saurat O, Dudge A, Kiss GB, Roe B, Young ND, Cook DR (2006) Distribution of microsatellites in the genome of *Medicago truncatula*: a resource of genetic markers that integrate genetic and physical maps. *Genetics* 172:2541–2555
- Na J-K, Wang J, Murray JE, Gschwend AR, Zhang W, Yu Q, Navajas-Pérez R, Feltus AF, Chen C, Kubat Z, Moore PH, Jiang J, Paterson AH, Ming R (2012) Construction of physical maps for the sex-specific regions of papaya sex chromosomes. *BMC Genomics* 13:176
- Paiva SAR, Russell RM (1999)  $\beta$ -Carotene and other carotenoids as antioxidants. *J Am Coll Nutr* 18:426–433
- Ross M, LaBrie T, McPherson S, Stanton VP (1999) Screening large-insert libraries by hybridization. In: Haines JL, Korf BR, Morton CC, Seidman JG, Smith DR (eds) *Current protocols in human genetics*. Wiley, New York, pp 5.6.1–5.6.32
- Saski CA, Feltus FA, Staton ME, Blackmon BP, Ficklin SP, Kuhn DN, Schnell RJ, Shapiro H, Motamayor JC (2011) A genetically anchored physical framework for *Theobroma cacao* cv. Matina 1–6. *BMC Genomics* 12:413
- Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reilly AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddelloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326:1112–1115
- Skelton RL, Yu Q, Srinivasan R, Manshardt R, Moore PH, Ming R (2006) The expression of the gene for Lycopene  $\beta$ -cyclase is elevated in leaves and flowers and down-regulated in both yellow- and red-fleshed papaya fruits. *Cell Res* 16:731–739
- Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. *Comput Appl Biosci* 13:523–535
- Storey WB (1953) Genetics of papaya. *J Hered* 44:70–78
- Storey WB (1969) Papaya. In: Ferwerda FP, Wit F (eds) *Outlines of perennial crop breeding in the tropics*. Veenman, Wageningen, pp 389–408
- Suzuki JY, Tripathi S, Fermín GA, Jan F-J, Hou S, Saw JH, Ackerman CM, Yu Q, Schatz MC, Pitz KY, Yépes M, Fitch MMM, Manshardt RM, Slightom JL, Ferreira SA, Salzberg SL, Alam M, Ming R, Moore PH, Gonsalves D (2008) Characterization of insertion sites in rainbow papaya, the first commercialized transgenic fruit crop. *Trop Plant Biol* 1:293–309
- Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH (2008) Synteny and collinearity in plant genomes. *Science* 320:486–488



- The Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. *Nature* 485:635–641
- Troggio M, Malacarne G, Coppola G, Segala C, Cartwright D, Pindo M, Stefanini M, Mank R, Moroldo M (2007) A dense SNP-based genetic linkage map of grapevine (*Vitis vinifera* L.) anchoring Pinot Noir BAC contigs. *Genetic* 176:2637–2650
- Wai CM, Moore PH, Paull RE, Ming R, Yu Q (2012) An integrated cytogenetic and physical map reveals unevenly distributed recombination spots along the papaya sex chromosomes. *Chromosome Research* 20:753–767
- Wai CM, Ming R, Moore PH, Paull RE, Yu Q (2010) Development of chromosome-specific cytogenetic markers and merging of linkage fragments in papaya. *Trop Plant Biol* 3:171–181
- Wikström N, Savolainen V, Chase MW (2001) Evolution of the angiosperms: calibrating the family tree. *Proc R Soc Lond B* 268:2211–2220
- Yamamoto HY (1964) Comparison of the carotenoids in yellow- and red-fleshed carica papayas. *Nature* 201:1049–1050
- Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L, Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X, Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D, Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z, Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* 296:79–92
- Yu Q, Moore PH, Albert HH, Roder AHK, Ming R (2005) Cloning and characterization of a *FLORICAULA/LEAFY* ortholog, *PFL*, in polygamous papaya. *Cell Res* 15:576–584
- Yu Q, Steiger D, Kramer EM, Moore PH, Ming R (2008) Floral MADS-box genes in trioecious papaya: characterization of *AG* and *API* subfamily genes revealed a sex-type-specific gene. *Trop Plant Biol* 1:97–107
- Yu Q, Tong E, Skelton RL, Bowers JE, Jones MR, Murray JE, Hou S, Guan P, Acob RA, Luo MC, Moore PH, Alam M, Paterson AH, Ming R (2009) A physical map of the papaya genome with integrated genetic map and genome sequence. *BMC Genomics* 10:371
- Yüksel B, Bowers J, Estill J, Goff L, Lemke C, Paterson A (2005) Exploratory integration of peanut genetic and physical maps and possible contributions from *Arabidopsis*. *Theor Appl Genet* 111:87–94
- Zhang W, Wang X, Yu Q, Ming R, Jiang J (2008) DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya. *Genome Res* 18:1938–1943
- Zhang W, Wai CM, Ming R, Yu Q, Jiang J (2010) Integration of genetic and cytological maps and development of a pachytene chromosome-based karyotype in papaya. *Trop Plant Biol* 3:166–170