Tropical Plant Biol. (2008) 1:49–57
DOI 10.1007/s12042-007-9005-7

Recent Origin of Dioecious and Gynodioecious Y Chromosomes in Papaya

Qingyi Yu · Rafael Navajas-Pérez · Eric Tong · Jon Robertson · Paul H. Moore · Andrew H. Paterson · Ray Ming

Received: 6 November 2007 / Accepted: 7 December 2007 / Published online: 31 January 2008

Abstract Sex of dioecious and gynodioecious papayas is controlled by two slightly different Y chromosomes, Y for males and Y^h for hermaphrodites. All combinations of the Y and/or Y^h chromosomes are lethal. We investigated the features of paired dioecious X- and Y-specific bacterial artificial chromosomes (BACs) and compared their sequences to corresponding gynodioecious X- and Y-specific BACs. Numerous chromosomal rearrangements were detected between the X- and Y-specific BACs, including inversions, deletions, insertions, and duplications. DNA sequence expansion was documented on the Y BAC. Dioecious and gynodioecious X-specific BACs were virtually identical. The Y- and Y^h-specific BACs shared high degree of DNA sequence identity, but local chromosomal rearrangements were detected, as the consequence of suppression of recombination in the male specific region and the isolation of Y and Y^h chromosomes enforced by the lethal effect. Analysis of sequence divergence between three dioecious X and Y gene pairs resulted in the estimated ages of divergence from 0.6 to 2.5 million years, reinforcing the hypothesis of a recent origin of the papaya sex chromosomes. The estimated age of divergence between Y and Y^h chromosomes was approximately 73,000 years for Gene 5. Our findings indicate that Y and Y^h chromosomes evolved from a common ancestral Y chromosome, possibly prior to the origin of agriculture. The existence of a hermaphrodite Y^h chromosome is less likely to have resulted from human selection as once suggested.

Keywords *Carica papaya* · Chromosomal rearrangements · Molecular evolution · MSY male specific region of the Y chromosome · Sex chromosomes

Communicated by

Electronic supplementary material The online version of this article (doi:10.1007/s12042-007-9005-7) contains supplementary material, which is available to authorized users.

Q. Yu · E. Tong
Hawaii Agriculture Research Center,
Aiea, HI 96701, USA

R. Navajas-Pérez · J. Robertson · A. H. Paterson
Plant Genome Mapping Laboratory, University of Georgia,
Athens, GA 30602, USA

P. H. Moore
USDA-ARS, Pacific Basin Agricultural Research Center,
Hilo, HI 96720, USA

R. Ming (✉)
Department of Plant Biology,
University of Illinois at Urbana–Champaign,
Urbana, IL 61801, USA
e-mail: rming@life.uiuc.edu

Introduction

Papaya (*Carica papaya* L.) belongs to a small family Caricaceae with 32 dioecious species, two trioecious species, and one monoecious species [5, 19]. It is the only species in the genus *Carica* and its closest related genus is *Vasconcellea* with 21 species [4, 6]. Papaya is a productive tropical fruit crop cultivated in tropical and subtropical regions worldwide.

Papaya is trioecious with three sex types: male, female, and hermaphrodite. In any given breeding system of papaya, it is either dioecious with male and female or gynodioecious with hermaphrodite and female. Sex determination in papaya is controlled by primitive sex chromosomes with a small male specific region about 10–15% of the Y chromosome [17, 29]. There are two slightly different Y chromosomes, Y controlling male and Y^h controlling

hermaphrodite [18]. All combinations of the Y and/or Y^h chromosomes are lethal, resulting in no pairing and recombination between YY, YY^h, and Y^hY^h chromosomes. For this reason, there is no breeding system in papaya that produce all three sex types. However, males can pollinate gynodioecious females as easily as hermaphrodites pollinate dioecious females, so that the X chromosomes in dioecious and gynodioecious chromosomes are not biologically isolated.

The sequence features of Y^h chromosome were investigated by complete sequencing of selected bacterial artificial chromosomes (BACs) in the male specific region of the Y^h chromosome (MSY^h) [30]. Abundant transposable elements and local sequence duplications were detected in the MSY^h that is suppressed for recombination. No gene was detected in the first five sequenced MSY^h BACs, suggesting extreme gene paucity in this region. The low gene density might be partly explained by the chromosomal location of the MSY^h that is near the centromere as demonstrated by *in situ* hybridization of two sequenced MSY^h BACs [30].

Sequencing two pairs of homologous X- and Y^h-BACs further improved our understanding about sequence features and the evolutionary history of the MSY^h [29]. Within the two large MSY^h BACs having a combined length of 440 Kb, all possible chromosomal rearrangements were detected, including inversions, deletions, insertions, duplications, and translocations, which collectively demonstrate the dynamic evolutionary process on the MSY^h after chromosomal recombination in this region ceased. DNA sequence expansion on the Y^h chromosome was documented in the two MSY^h BACs when compared with their homologous X chromosome counterparts. The estimated time of divergence ranged from 0.5 to 2.2 million years between four X and Y^h gene pairs on the two pairs of X- and Y^h-BACs that are 4–5 Mb apart. Three of the four pairs of genes were on a single BAC with estimated age of divergence ranging from 0.5–1.8 million years within a confined region of about 70 Kb. These are the first estimates of chromosome divergence time and support the hypothesis of a recent origin of the sex chromosomes in papaya. These findings suggested that sex chromosomes did not evolve at the family level and most likely evolved at the genus or species level, because Caricaceae diverged from its closest family Moringaceae about 60 MYA [28]. The divergence time between *Carica* and its closest genus *Vasconcelleae* is not yet available.

Sequence comparison between X- and Y^h-specific BACs provided evidence for the antiquity of the papaya sex chromosome and the nature of sequence variation after the suppression of recombination that triggered the evolution of the primitive sex chromosomes. The rational of this project is to utilize available genomic resources to test the hypothesis of the recent divergence of the dioecious Y and the gynodioecious Y^h chromosomes from an ancestral Y chromosome by sequencing a pair of homologous dioecious X- and Y-specific BACs. The objectives are: (1) to estimate the time of divergence of homologous genes on dioecious X and Y chromosomes to verify the conclusions of low X/Y divergence of sex linked genes [29] (2) to estimate the time of divergence of homologous genes on Y and Y^h chromosomes to test the hypothesis of their recent separation after the X and Y diverged; (3) to compare the dioecious X-specific BAC sequence with its homologous gynodioecious X BAC sequence to confirm that there is no biological barrier between X chromosomes of these two breeding systems; and (4) to analyze contiguous genomic sequences of dioecious X- and Y-BACs to validate the chromosomal rearrangements observed on MSY^h BACs.

Results

Identification of Male X- and Y-specific BAC Clones

Previously identified hermaphrodite papaya X- and Y^h-specific BAC clones, PH61H02 and PH95B12, contain three genes. Three pairs of primers were designed based on the cDNA sequences of the three genes and used to amplify the cDNA from leaf and young flower tissues of hermaphrodite papaya plants. The amplified cDNA fragments were used as probes to screen the male papaya BAC library. Ten positive BAC clones were identified and confirmed by PCR. The PCR products were sequenced and compared with gynodioecious X-specific BAC PH61H02 and Y^h-specific BAC PH95B12. BAC ends of the ten positive BAC clones were sequenced and aligned with BACs PH61H02 and PH95B12. The dioecious X-specific BAC clone DM10G24 and Y-specific BAC clone DM62K05 were identified corresponding to gynodioecious BACs PH61H02 and PH95B12, respectively.

Sequence Comparison of Dioecious and Gynodioecious X, Y, and Y^h BACs

Sequence comparison of dioecious X- and Y-specific BACs can reveal chromosomal rearrangements and spontaneous mutations of the MSY after the suppression of recombination. The alignment of dioecious Y-specific BAC DM10G24 and X-specific BAC DM62K05 is shown in Fig. 1a, based on the direction of alignment of their gynodioecious counterparts [29]. The entire aligned region appeared to be involved in an inversion, from 19.5–91 Kb on the X-specific BAC DM10G24 and 0 to 87 Kb on the MSY BAC DM62K05. There was about 15.5 Kb (21.7%) sequence expansion on the MSY BAC. Closer examination of the aligned sequences revealed a separation of three regions with distinctive sequence homology. Region 1 (from

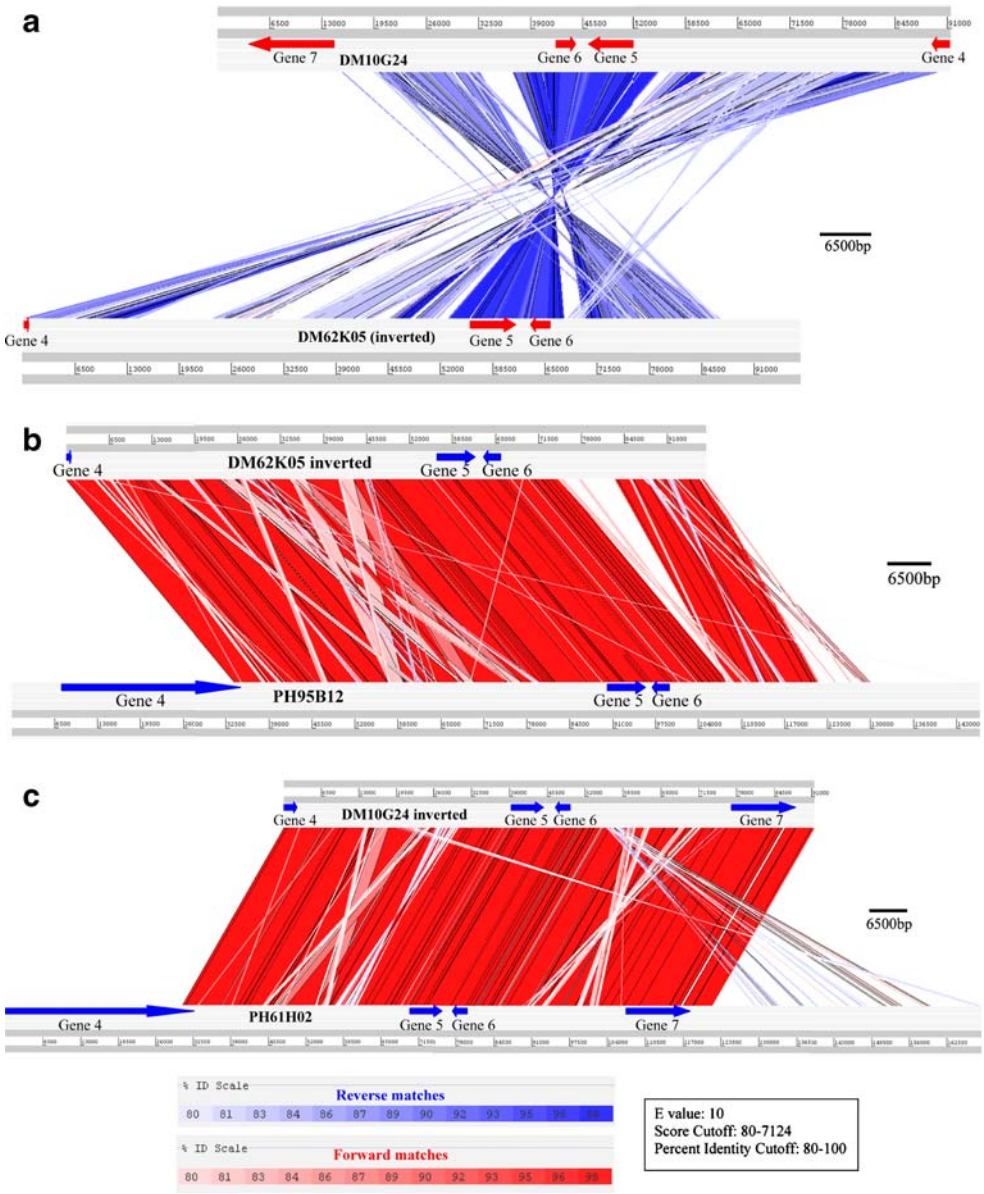

Fig. 1 DNA sequence comparison between paired BACs of dioecious X and Y, dioecious X and gynodioecious X, and dioecious Y and gynodioecious Y^h. The connected lines showing lower percentage of homology and local rearrangements consist largely of repetitive sequences. **a** The male X BAC DM10G24 was compared with its corresponding male Y BAC DM62K05. *Red arrows* highlight the genes' locations and orientations. Gene 6 is located on the complementary strand. **b** The male Y BAC DM62K05 was compared with its corresponding hermaphrodite Y^h BAC PH95B12. **c** The male X BAC DM10G24 was compared with its corresponding hermaphrodite X BAC PH61H02

0 to 52 Kb on the MSY BAC DM62K05) shared the lowest sequence homology with its X counterpart with about 81–87% sequence identity. Region 2 (from 52 to 72.5 Kb, including Genes 5 and 6) shared the highest sequence homology with its X counterpart showing about 96–98% sequence identity. Region 3 (from 72.5 to 87 Kb) shared moderate sequence homology with its X counterpart at 87–90% sequence identity.

Insertions, deletions, and local small scale inversions were also detected in the aligned MSY sequences. There were five visible insertions in region 1 with the lowest sequence identity and the largest insertion about 9.5 Kb at 13.5–23 Kb on MSY BAC DM62K05. One insertion in region 2 located a short distance (about 2 Kb) away from Gene 6. Small scale insertions were seen as tiny gaps in the aligned sequence of DM62K05. A major deletion on MSY appeared to have happened near or at the border of regions 1 and 2, corresponding to 32.5–36 Kb of the X BAC DM10G24. Small scale deletions on the MSY were seen as gaps in the aligned region on the X-specific BAC DM10G24. A local inversion within this major inversion appears to have occurred involving the border sequences between regions 1 and 2 at 45.5–51.5 Kb on the MSY BAC DM62K05. The different levels of sequence identity are distinguishable in this local inversion.

Sequence comparison of the dioecious Y and gynodioecious Y^h BACs revealed the level of DNA sequence homology and the extent of chromosomal rearrangements between long contiguous genomic sequence of Y and Y^h chromosomes. Sequence alignment of MSY BAC DM62K05 and MSY^h BAC PH95B12 revealed high degree of sequence homology of the entire aligned region at above

98% sequence identity (Fig. 1b). The cross lines of lower sequence identity between these two homologous BACs were likely noise caused by repetitive sequences. Detailed sequence comparison revealed a large insertion/deletion (indel) of 8398 bp, shown as the large gap from 73.5–82 Kb on DM62K05. Besides this indel, 36 small indels and 231 single nucleotide polymorphisms (SNPs) were observed in the aligned 88 Kb sequence between DM62K05 and PH95B12.

Sequence comparison between dioecious and gynodioecious X chromosomes should provide evidence to test whether there is biological isolation between these two breeding systems. Direct alignment of dioecious X BAC DM10G24 and gynodioecious X BAC PH61H02 revealed nearly 100% sequence identity. Detailed sequence comparison revealed five single nucleotide indels, three transitions (1 g–a, 1 a–g, 1 t–c), and four transversions (3 a–t, 1 g–t) in the 91 Kb aligned sequence between the two X-specific BACs.

Further analysis of gapless alignment of the XY, XX, and YY^h BAC pairs using BlastZ provided a quantitative assessment of homologous DNA sequences between different sex chromosomes. Between homologous dioecious X- and Y-BACs, 417 gapless segments were aligned, indicating 416 rearrangements (indels, inversions, duplications) in the homologous region (Table 1). The average sequence identity was 84.4% in the aligned 62 Kb region. Six gapless segments were aligned between paired dioecious X- and gynodioecious X^h-BACs, separated by the five single nucleotide indels (The designation of X^h is purely for the convenience of data presentation and there is no real difference between X and X^h chromosomes. The X chromosome is used for both X and X^h chromosomes in general statements). The average sequence identity was 100% in the aligned 91 Kb region. Thirty nine gapless segments were aligned between homologous dioecious Y- and gynodioecious Y^h-BACs, separated by the three large indels and 35 single nucleotide indels. The average sequence identity was 98.6% in the aligned 88 Kb region.

Sequence Divergence of the Three X and Y^h Gene Pairs

Based on the previously annotated genes on the MSY^h [29], four genes were found in X BAC DM10G24, including full length Genes 5, 6, and 7, and partial Gene 4; and three genes on Y BAC DM62K05, including full length Genes 5 and 6, and partial Gene 4 (Table 2). The gene structure was conserved among alleles on X, X^h, Y, and Y^h chromosomes. DNA sequence identity was 100% between X and X^h alleles for all exons and introns except Gene 7 Intron 9 that differed by one nucleotide indel. DNA sequence identity was 100% for all exons and 11 of the 17 introns of between Y and Y^h alleles of Genes 4, 5, and 6, but six of the 17 introns accumulated indels and SNPs. In contrast, only five of the 19 exons shared 100% sequences identity between X^h and Y^h, and X and Y alleles of Genes 4, 5, and 6, and the other 14 exons and all 17 introns showed various levels of sequence divergence. The average sequence identity between X^h/Y^h and X/Y gene pairs ranged from 94.35 to 98.7%.

The Exon 6 of Y and Y^h alleles of Gene 6 shared 95.7% sequence identity with their X counterparts. More importantly, there is a premature stop codon in Exon 6 of Y and Y^h alleles that would produce a truncated protein without the last six amino acids (Supplementary Figure 1).

The coding sequences of the three X/Y gene pairs (Genes 4, 5, and 6) were analyzed to determine the degree of synonymous (K_s) and nonsynonymous (K_a) divergence between them (Tables 3 and 4). The X and Y pair of Gene 5 has K_a/K_s ratios that are considerably less than 1 at 0.04, and Gene 6 at 0.64, suggesting their divergence has been functionally constrained. The X and Y pair of the partial Gene 4 showed no nonsynonymous mutation, thus there is no estimate of K_a/K_s ratio. The Y and Y^h pairs of Genes 4, 5, and 6 showed either no synonymous or nonsynonymous mutation or both, preventing the calculation of K_a/K_s ratios.

The degree of silent site nucleotide divergence (K_{sil}) reflects the underlying neutral mutation rate, and can be used to estimate the time of divergence between the X and

Table 1 Summary of gapless comparison between dioecious X- and Y- and gynodioecious X^h- and Y^h-specific BACs by BLASTZ

	DM10G24(X)/DM62K05(Y)	DM10G24(X)/PH61H02(X^h)	DM62K05(Y)/PH95B12(Y^h)
Length of BAC (bp)	91,377/96,769	91,377/168,440	96,769/146,496
Aligned sequence (bp)	60,853	91,373	88,341
Span of aligned BAC (bp)	61,782/61,947	91,377/91,374	88,418/88,394
% Aligned covered	98.5/98.2	99.9/99.9	99.9/99.9
Total contigs	417	6	39
Average % identity	84.35	100	98.64
Average contig length (bp)	146	15,228.83	2,265.154
Longest contig (bp)	4,011	39,638	6,999

Table 2 Detailed sequence comparison of four genes on dioecious and gynodioecious sex chromosomes

Gene ID	Gene structure	Intron/exon sizes (bp)				% Identity			
		X^h	Y^h	X	Y	X/X^h	Y/Y^h	X^h/Y^h	X/Y
Gene 4	Intron 16 (Partial)	300	313	300	313	100.0	100.0	88.8	88.8
	Exon 17	63	63	63	63	100.0	100.0	96.8	96.8
	Intron 17	84	84	84	84	100.0	100.0	94.0	94.0
	Exon 18	132	132	132	132	100.0	100.0	97.7	97.7
	Average %identity					100.0	100.0	94.3	94.3
Gene 5	Exon 1	85	85	85	85	100.0	100.0	98.8	98.8
	Intron 1	928	927	928	927	100.0	99.6	98.9	98.7
	Exon 2	133	133	133	133	100.0	100.0	100.0	100.0
	Intron 2	102	102	102	102	100.0	100.0	98.0	98.0
	Exon 3	72	72	72	72	100.0	100.0	100.0	100.0
	Intron 3	1,342	1,328	1,342	1,328	100.0	99.9	95.3	95.4
	Exon 4	144	144	144	144	100.0	100.0	97.2	97.2
	Intron 4	366	364	366	364	100.0	100.0	96.7	96.7
	Exon 5	72	72	72	72	100.0	100.0	95.8	95.8
	Intron 5	104	92	104	92	100.0	100.0	87.5	87.5
	Exon 6	72	72	72	72	100.0	100.0	98.6	98.6
	Intron 6	83	83	83	83	100.0	100.0	95.2	95.2
	Exon 7	113	113	113	113	100.0	100.0	100.0	100.0
	Intron 7	361	358	361	358	100.0	99.4	93.6	93.6
	Exon 8	132	132	132	132	100.0	100.0	98.5	98.5
	Intron 8	76	76	76	76	100.0	100.0	97.4	97.4
	Exon 9	342	342	342	342	100.0	99.7	99.1	99.4
	Intron 9	165	165	165	165	100.0	100.0	97.6	97.6
	Exon 10	395	395	395	395	100.0	100.0	98.2	98.2
	Intron 10	207	207	207	207	100.0	100.0	94.2	94.2
	Exon 11	315	315	315	315	100.0	100.0	98.1	98.1
	Average %identity					100.0	99.9	97.1	97.1
Gene 6	Exon 1	50	50	50	50	100.0	100.0	100.0	100.0
	Intron 1	495	493	495	495	100.0	99.6	99.2	99.6
	Exon 2	133	133	133	133	100.0	100.0	100.0	100.0
	Intron 2	70	70	70	70	100.0	100.0	97.1	97.1
	Exon 3	104	104	104	104	100.0	100.0	99.0	99.0
	Intron 3	116	116	116	116	100.0	100.0	99.1	99.1
	Exon 4	172	172	172	172	100.0	99.4	98.8	99.4
	Intron 4	115	115	115	115	100.0	100.0	99.1	99.1
	Exon 5	141	141	141	141	100.0	100.0	99.3	99.3
	Intron 5	76	76	76	76	100.0	100.0	97.3	97.3
	Exon 6	1,002	987	1,002	987	100.0	100.0	95.7	95.7
	Average %identity					100.0	99.9	98.6	98.7
Gene 7	Exon 1	91		91		100.0			
	Intron 1	90		90		100.0			
	Exon 2	48		48		100.0			
	Intron 2	245		245		100.0			
	Exon 3	42		42		100.0			
	Intron 3	711		711		100.0			
	Exon 4	120		120		100.0			
	Intron 4	205		205		100.0			
	Exon 5	146		146		100.0			
	Intron 5	78		78		100.0			
	Exon 6	148		148		100.0			
	Intron 6	127		127		100.0			
	Exon 7	84		84		100.0			
	Intron 7	359		359		100.0			
	Exon 8	116		116		100.0			

Table 2 (continued)

Gene ID	Gene structure	Intron/exon sizes (bp)				% Identity			
		X^h	Y^h	X	Y	X/X^h	Y/Y^h	X^h/Y^h	X/Y
	Intron 8	916		916		100.0			
	Exon 9	64		64		100.0			
	Intron 9	6,980		6,979		99.9			
	Exon 10	91		91		100.0			
	Intron 10	181		181		100.0			
	Exon 11	277		277		100.0			

Y, and Y and Y^h gene pairs. The degree of silent site divergence was low between three X/Y gene pairs, ranging between 0.018 and 0.074, and extremely low between Y/Y^h gene pairs, ranging from 0 to 0.0022 (Tables 5 and 6). The estimate time of divergence ranges from approximately 0.6 to 2.5 million years between X/Y gene pairs, and approximately 73,000 years between Y/Y^h pair Gene 5.

Discussion

Because dioecy is prevalent in the small family Caricaceae (32 of the 35 species), it was thought that dioecy might be the ancestral state of the family and that sex chromosomes could have been originated at the family level. Our estimates of 0.6 to 2.5 million years of divergence between three X/Y gene pairs were similar to the previous estimates of 0.5–2.2 million years of divergence in four pairs of X^h/Y^h genes [29]. These results reinforce the conclusion that sex chromosomes in papaya likely evolved at the genus or species level, long after the divergence of Caricaceae from its closest related family Moringaceae about 60 MYA [28]. Investigation of sequence divergence between X/Y gene pairs from several species of the genus *Vasconcellea*, formerly a section of the genus *Carica* [5], is under way and will further clarify the origin of sex chromosomes in Caricaceae.

The prevalence of dioecious species in Caricaceae also led to the suggestion that hermaphrodites could have been resulted from human selection for improving the efficiency of fruit production [26], as every hermaphrodite tree produces fruits comparing with the two-fold cost of fruit production in dioecious system (male trees will not produce fruits). The isolation of X/Y and Y/Y^h gene pairs made it possible to study the evolutionary history of sex chromosomes in papaya, as demonstrated by the X/Y gene pairs of the plant sex chromosome model species *Silene latifolia* [1, 7, 10, 11, 21, 24]. Because only two exons in Gene 4 is covered by dioecious X and Y BACs and Gene 6 is a small gene with a slow molecular clock as seen in the X/Y and X/Y^h gene pair [29], no silent mutations were observed in their Y/Y^h gene pairs. However, there are silent mutations of Gene 5 between its Y/Y^h pair and the estimated time of divergence of about 73,000 year would set the minimum amount of time since the Y and Y^h chromosomes diverged, as the upper range of 2.5 million years sets the minimum time of divergence between X and Y chromosomes. Our results would place the separation of the dioecious and gynodioecious breeding systems before the origin of agriculture about 10,000 years ago [12], and make it unlikely that the appearance of hermaphrodite papaya is due to human selection.

The homologous X chromosomes in female plants are still recombining during meiosis in both dioecious and gynodioecious systems. Our results confirm that there is no biological barrier between the dioecious and gynodioecious breeding systems in papaya, as showed by the nearly 100% genomic sequence identity of the 91 Kb aligned region of the X and X^h BACs and the 100% identity of exon and intron sequences of the four genes (except for the one

Table 3 Estimates of synonymous and nonsynonymous nucleotide divergence of dioecious X- and Y-linked gene sequences

Gene ID	No. of sites (bp)				No. of mutations		Sequence divergence		K_a/K_s
	Total sites	Total Coding sites	Syn. sites	Nonsyn. sites	Syn. mutations	Nonsyn. mutations	Syn. sites (K_s)	Nonsyn. sites (K_a)	
4	579	192	46.17	145.83	5	0	0.12	0.0	–
5	5,572	1,872	452.42	1,419.58	23	3	0.053	0.002	0.04
6	2,474	1,599	366.42	1,232.58	14.5	31.5	0.041	0.026	0.64

Syn. Synonymous; *Nonsyn*. nonsynonomous

Table 4 Estimates of synonymous and nonsynonymous nucleotide divergence of dioecious Y- and gynodioecious Y^h-linked gene sequences

Gene ID	No. of sites (bp)				No. of mutations		Sequence divergence		K_a/K_s
	Total sites	Total Coding sites	Syn. sites	Nonsyn. sites	Syn. mutations	Nonsyn. mutations	Syn. sites (K_s)	Nonsyn. sites (K_a)	
4	592	192	46.17	145.83	0	0	0.0	0.0	–
5	5577	1,872	451.17	1,420.83	1	0	0.0022	0.0	–
6	2,472	1,599	366	1,233	0	1	0	0.0008	–

nucleotide indel in the 6980 kb intron 9 of Gene 7; Tables 1 and 2). The single nucleotide polymorphism rate between the X and X^h chromosomes is 0.0077% based on their 91 Kb homologous sequences. This rate is an order of magnitude lower than the 0.03% SNP between the Columbia and Landsberg *erecta* ecotypes of *Arabidopsis* and two orders of magnitude lower than the 0.53–0.78% SNP polymorphism rates between the rice subspecies *Oryza japonica* and *O. indica* [14, 27]. In contrast, the SNP rate between recently diverged Y and Y^h chromosomes is 0.261% based on the 88 Kb gapless alignment, 34 times of that between X and X^h chromosomes from the same papaya varieties, demonstrating the effect of Muller's ratchet after the suppression of recombination in the MSY region [22]. No SNP estimate in autosomal regions is available between papaya varieties AU9 and SunUp, the two genotypes used in this study of dioecious and gynodioecious sex chromosomes. It is not clear whether X chromosomes have more or less SNP due to the fact that only two of the three X chromosomes in the parental male and female plants recombine in the region corresponding to the MSY in each generation, while every autosome will recombine in every generation.

Sequence analysis of seven MSY^h BACs provided evidence of gene paucity in the MSY^h region [29, 30]. It is not yet clear whether this low gene density is due to the degeneration process of Y chromosome evolution or low gene density of the pericentromeric region that the MSY^h appeared to have derived from [30]. However, a premature stop codon was found in Y and Y^h alleles of Gene 6, suggesting that this gene might have been disrupted (Supplementary Figure 1). This is the first evidence that a Y and Y^h copy of the gene might have been degenerated. Accelerated rate of degeneration was documented in a recently evolved neo-Y chromosome in *Drosophila miranda* in merely one million years [2, 3].

The three regions with different levels of sequence homology in the 87 Kb inverted region of MSY BAC DM62K05 is another example of the dynamic evolutionary process of the Y chromosome. These three regions were from a single inversion, not three inversion events as the distinctively different sequence homology would suggest, since three separate inversions would have resulted in the exchange of chromosomal location of region 1 (0–52 Kb) and region 3 (72.5–87 Kb) from their current location. Region 1 started from the end of Gene 4 continued to a large region without functional genes. Higher level of sequence conservation from the end of Gene 4 to 13 Kb was observed, likely due to the selective constraint of Gene 4. The remaining 39 Kb include five noticeable insertions and low level of sequence conservation. The highly conserved region 2 (52–72.5 Kb) is also likely due to the selective constraint of Genes 5 and 6. However, Gene 6 appeared to have started the degeneration process and the rate of sequence divergence in Gene 6 and its neighboring region 3 would be accelerated.

The recent divergence of Y and Y^h chromosome from an ancestral Y (or Y^h) chromosome is also supported by the much higher sequence identity between Y/Y^h (98.8%) than between X/Y (84%) in the gapless comparison of homologous regions (Table 1). We are curious whether dioecy or gynodioecy is the ancestral state of *Carica papaya* although it is not the objective of this project. If considering X chromosome represent the ancestral state of the autosomes

Table 5 Estimated age of divergence of dioecious X- and Y-linked genes

Gene ID	Silent sites	Silent mutations	Silent site divergence (Ksil)	Estimated Age (MYA)[a]
4	433.17	32	0.074	2.47
5	4,152.42	128	0.031	1.03
6	1,241.42	22.5	0.018	0.6

[a] Based on synonymous substitution rate of 1.5×10^{-8} substitutions/synonymous site/year [15]

Table 6 Estimated age of divergence of dioecious Y- and gynodioecious Y^h-linked genes

Gene ID	Silent sites	Silent mutations	Silent site divergence (K_{sil})	Estimated Age (MYA)
4	446.17	0	0.0	0
5	4,156.17	9	0.0022	0.073
6	1,239	0	0.0	0

from which the sex chromosomes evolved from, our preliminary evidence pointed to the dioecious Y chromosome as ancestral, because Y-specific BAC shared slightly higher DNA sequence homology with its X counterpart than Y^h-specific BAC. Moreover, among the 37 indels, more deletions occurred on Y^h, including the large deletion of 8,398 bp. On the other hand, sequence divergence of the Y and Y^h chromosomes occurred after their divergence, and the sequence homology with their X counterparts provide no valid evidence whether the Y or Y^h chromosome is the ancestral one. It has been hypothesized that sex chromosome evolution involves two tightly linked sex determination genes, one controlling male sterile and the other female sterile [9]. If this hypothesis proved to be true, it would favor dioecy (the Y chromosome) as the ancestral state of *Carica papaya*. Identification of a large deletion (about 1 Mb) in two male to female sex reversal mutants generated by irradiation appeared to support the above hypothesis (Q. Yu, P. H. Moore, R. Ming, unpublished data), because a male to female sex reversal requires the removal of the gene controlling carpel abortion (male sex determination) and the gene controlling stamen development (female sex determination). The ultimate proof of the ancestral state of *Carica papaya* sex chromosomes would come from the identification of both male and female sex determination genes and detailed analysis of their evolutionary history in dioecious and gynodioecious breeding systems from a wide collection of papaya germplasm.

Methods

Plant Material

Young leaf tissue from male plants of Australian dioecious variety AU9 was used for nuclei DNA isolation. The leaf and young flower buds from hermaphrodite plants of SunUp were used for RT-PCR amplification of target cDNA sequences to be used as probes to screen the male BAC library.

Construction of a Male Papaya BAC Library

The nuclei DNA was isolated from young leaf tissue of a male tree of variety AU9 as described by [20]. The nuclei DNA was embedded in agarose and partially digested by *Eco*RI. The fraction around 100 kb was recovered and ligated into pIndigoBAC536. A total of 55,296 BAC clones were picked and achieved in 384-well plates with freezing medium. BAC clones were gridded onto Hybond N membrane (Genetix, Hampshire, UK) by Q-Pix 2 (Genetix). Thirty sets of high density BAC filters have been made containing the entire library for screening target MSY and dioecious X-specific BACs (Q. Yu, C. Saski, P. H. Moore, R. Ming, unpublished data).

Sequencing Male X and Y BACs and Sequence Assembly

The papaya BAC clones were sequenced using the shotgun approach with at least 10X coverage. BAC DNAs were isolated using Qiagen-Plasmid Midi Kit (Qiagen, Valencia, CA, USA) and randomly sheared using Hydroshear (Genomic Solutions, Ann Arbor, MI, USA) to generate ~3 kb insert fragments. The sheared fragments were size-selected on an agarose gel, purified with a Gel Extract and Purification Kit (Qiagen), end-repaired by DNA Terminator End Repair Kit (Lucigen, Middleton, WI, USA), and ligated into the pSMART-HCKan vector (Lucigen). DNAs from the 3-kb libraries were cycle-sequenced with ABI BigDye Terminator v3.1 and analyzed on a 3730XL DNA Analyzer (Applied Biosystems, Foster City, CA, USA).

Phred/Phrap/Consed and CAP3 packages were used for sequence assembly. Gaps in assembly and regions of low-quality were resolved by resequencing subclones identified by Autofinish, sequencing PCR products, and/or additional random subclone sequencing. All BAC clones were manually examined for signs of misassembly. Suspect regions were clarified either by ambiguous read removal, PCR amplification and sequencing, and/or alignment with a neighboring BAC. A BAC was not considered complete until all inconsistent read pairs were resolved and Consed reported an error rate of less than 1/10,000 bases.

RT-PCR

At least one intron was covered by primers designed for RT-PCR experiments to control genomic DNA contamination. Total RNA was extracted from young flower buds (0.4–0.7 cm, before meiosis) and leaf tissues. Approximately 2 μg of total RNA was treated with RNase-free DNase I (Promega, WI, USA) and reverse transcribed using RETROscript kit (Ambion, CA, USA). The synthesized cDNAs served as templates for RT-PCR.

Sequence Divergence of X and Y^h Gene Pairs

Exon and intron regions of the X and Y^h gene pairs that corresponded to EST-supported sequences were manually aligned using BioEdit [13]. The numbers of synonymous substitutions per synonymous site (K_s), nonsynonymous substitutions per nonsynonymous site (K_a), and synonymous and noncoding (silent) substitutions per silent site (K_{sil}) were estimated according to the method of [23] and implemented in DnaSP 4.0 [25]. Divergence times for the paired X and Y^h alleles were determined using K_{sil} and the

methods described in [16] using the synonymous substitution rate for dicot nuclear genes estimated by [15].

X and Y BAC Sequence Comparison Analysis

The X and Y BAC sequence comparison study was conducted using the Artemis Comparison Tool [8].

Acknowledgments We thank Chris Saski for constructing the papaya male BAC library. This work was supported by a grant from NSF to R.M., Q.Y., P.H.M., J.J., and A.H.P. (DBI-0553417) and a USDA-ARS Cooperative Agreement (CA 58-3020-8-134) with the Hawaii Agriculture Research Center.

References

1. Atanassov I, Delichère C, Filatov DA, Charlesworth D, Negrutiu I, Monéger F (2001) Analysis and evolution of two functional Y-linked loci in a plant sex chromosome system. Mol Biol Evol 18:2162–2168
2. Bachtrog D (2006) Expression profile of a degenerating neo-Y chromosome in *Drosophila*. Curr Biol 16:1694–1699
3. Bachtrog D (2005) Sex chromosome evolution: molecular aspects of Y-chromosome degeneration in *Drosophila*. Genome Res 15: 1393–1401
4. Badillo VM (2000) *Carica* L. vs. *Vasconcella.* St. Hil. (Caricaceae): con la rehabilitación de este último. Ernstia 10:74–79
5. Badillo VM (1971) Monografia de la familia Caricaceae. Publicada por la Associacion de Profesores, Venezuela, Univ. Centr. Venez, p 220
6. Badillo VM (2001) Nota correctiva *Vasconcellea* St. Hil y no *Vasconcella* (Caricaceae). Ernstia 11:75–76
7. Bergero R, Forrest A, Kamau E, Charlesworth D (2007) Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. Genetics 175:1945–1954
8. Carver TJ, Rutherford KM, Berriman M, Rajandream M-A, Barrell BG, Parkhill J (2005) ACT: the Artemis Comparison Tool. Bioinformatics 21:3422–3423
9. Charlesworth B, Charlesworth D (1978) A model for the evolution of dioecy and gynodioecy. Am Nat 112:975–997
10. Delichère C, Veuskens J, Hernould M, Barbacar N, Mouras A, Negrutiu I, Monéger F (1999) *SlY1*, the first active gene cloned from a plant Y chromosome, encodes a WD-repeat protein. EMBO J 18:4169–4179
11. Filatov DA (2005) Substitution rates in a new *Silene latifolia* sex linked gene, *SlssX/Y.* Mol Biol Evol 22:402–408
12. Gupta AK (2004) Origin of agriculture and domestication of plants and animals linked to early Holocene climate amelioration. Current Sci 87:54–59
13. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids SYp Ser 41:95–98
14. International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800
15. Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). Mol Biol Evol 17:1483–1498
16. Li W-H (1997) Molecular evolution. Sinauer, Sunderland
17. Liu Z, Moore PH, Ma H, Ackerman CM, Ragiba M, Yu Q, Pearl HM, Kim MS, Charlton JW, Stiles JI, Zee FT, Paterson AH, Ming R (2004) A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature 427:348–352
18. Ming R, Yu Q, Moore PH (2007) Sex determination in papaya. Semin Cell Dev Biol 18:401–408
19. Ming R, Van Droogenbroeck B, Moore PH, Zee FT, Kyndt T, Scheldeman X, Sekioka T, Gheysen G (2005) Molecular diversity of *Carica papaya* and related species. In: Sharma AK, Sharma A (eds) Plant genome: biodiversity and evolution, vol. vol 1B. Science Publishers, New Hampshire, pp 229–254
20. 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
21. Moore RC, Kozyreva O, Lebel-Hardenack S, Siroky J, Hobza R, Vyskot B, Grant SR (2003) Genetic and functional analysis of *DD44*, a sex-linked gene from the dioecious plant *Silene latifolia* provides clues to early events in sex chromosome evolution. Genetics 163:321–334
22. Muller HJ (1964) The relation of recombination to mutational advance. Mutat Res 106:2–9
23. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426
24. Nicolas M, Marais G, Hykelova V, Janousek B, Laporte V, Vyskot B, Mouchiroud D, Neqrutiu I, Charlesworth D, Moneqer F (2005) A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. PLoS Biology 3:47–56
25. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497
26. Storey WB (1976) Papaya. In: Simmonds NW (ed) Evolution of crop plants. Longman, London, pp 21–24
27. The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796–815
28. Wikström N, Savolainen V, Chase MW (2001) Evolution of the angiosperm: calibrating the family tree. Proc R Soc Lond B 268: 2211–2220
29. Yu Q, Hou S, Feltus FA, Jones MR, Murray J, Veatch O, Lemke C, Saw JH, Moore RC, Thimmapuram J, Liu L, Moore PH, Alam M, Jiang J, Paterson AH, Ming R (2008) Low X/Y divergence in four pairs of papaya sex-liked genes. Plant J 53:124–132
30. Yu Q, Hou S, Hobza R, Feltus FA, Wang X, Jin W, Skelton RL, Blas A, Lemke C, Saw JH, Moore PH, Alam M, Jiang J, Paterson AH, Vyskot B, Ming R (2007) Chromosomal location and gene paucity of the male specific region on papaya Y chromosome. Mol Genet Genomics 278:177–185