Doctoral Dissertation (Shinshu University)

Structural and functional analysis of sex chromosomes in papaya (*Carica papaya*)

March 2015
11ST106A

Hiroki Ueno

Summary

Sexual reproduction in higher plants is highly divergent. The majority of higher plants are hermaphrodite species, and approximately 6% of flowering plants are dioecious. Of these, 48 species have been identified to have sex chromosomes, which are responsible for their sex determinations. According to the previous studies, those dimorphic sex chromosomes have derived from ancestral autosomes.

Papaya (*Carica papaya*), which belongs to the Caricaceae family, is known to be a species harboring sex chromosomes. Particularly, papaya is a trioecious plant, and its sex types are determined by the pairing of three sex chromosomes, XX, XY and XY^h for females, males and hermaphrodites, respectively. Phylogenic analysis presumed that hermaphrodite of papaya has derived from a natural mutation from male plants. Draft genome sequence was accessible in papaya and partial sequences of X and Y^h chromosomes are also available as BAC (Bacterial Artificial Chromosome) clone sequences. On the other hand, Y chromosome sequence information is limitedly analyzed. In spite of these studies, any genes for papaya sex determination have not been found in these sex chromosomes. Therefore, the present study aimed to identify candidate genes for sex determination in papaya by comparing structure and function of genes in the three sex chromosomes.

For identifying candidate genes for female/hermaphrodite or male determination, genome-wide gene expression was analyzed using High-throughput SuperSAGE in flower buds in all three sex types of papaya. Of the expressed genes in these tissues, expressed genes mapped in the sex chromosome sequences were focused. This analysis allowed to identifying 47 tags (genes) in the sex chromosomes, showing sex

type dependent expression. Among these genes, Cp2671, uniquely mapped on Y^h chromosome, encoded a MADS-box protein, which is a homologue of *SHORT VEGETATIVE PHASE* (*SVP*) of *Arabidopsis thaliana*. Since allele of this gene was absent in the X chromosome, it was suggested that this MADS-box gene was responsible for determination between female and other sex types.

For elucidating details of genome structures and functions of papaya sex chromosomes, reference sequences of X and Y^h chromosomes were constructed by own genome sequence analysis and assembly. Whole genomes of female and hermaphrodite plants were sequenced by next generation sequencer, and obtained sequence reads were re-assembled with publicly available BAC clone sequences of sex chromosomes. By comparison of predicted genes and genome sequences between the X and Y^h chromosomes, rearrangements process could be predicted during sex chromosome evolution. Transcriptome analysis of 99 unique genes on the X chromosomes, using RNA-seq revealed that no significant difference in expression level was found in the leaves between female and hermaphrodite plants, regardless of difference of their gene dosage, suggesting the presences of their compensation mechanism.

For identifying genes for male/hermaphrodite determination, genome structure of Y and Y^h chromosomes were compared by reference mapping the male genome sequence reads to the Y^h reference sequence. In total 15,493 polymorphisms were found between the Y and Y^h chromosome, together with three highly diverged regions. Of these polymorphisms between the Y^h and Y chromosomes, specific insertion of a putative transposon in the *SVP-like* gene was identified in the Y^h chromosome. Sequence of its cDNA demonstrated that truncation of MADS-box

domain was found in the transcripts from the Y^h chromosome, suggesting its concern to male/hermaphrodite determination.

These results in the present genomic studies must be greatly helpful for elucidating sex chromosome evolution in papaya. Particularly, the *SVP-like* gene was a candidate for sex determination (female-hermaphrodite, male-hermaphrodite). Further functional studies for this gene should shed light on sex determination mechanisms in papaya.

Preface

Sexual reproduction systems in higher plants are highly diverged, depending on the plant adaptation to environments. A majority of higher plants are hermaphrodite (bisexual) species, and approximately 6% of flowering plants are dioecious species having separate male and female individuals (Ming et al. 2011). In some plant species, hermaphrodite individual was also observed as an additional sex type to male and female, which is designated as 'trioecious'. Additionally, monoecy in which a plant carries both unisexual flowers (male and female in a single plant), was observed in Cucurbitaceae species like melon or cucumber, and other various patterns of sexual reproduction systems are present in flowering plants. These sex determination systems are one of the most important characteristics for understanding evolution or diversity of higher plants, and have great influences on fruit or seed production in crops. Nonetheless, their molecular or physiological mechanisms of sex determination were mostly unknown. As sex determination genes in higher plants, only in cucumber (Bai and Xu 2013) and melon (Martin et al. 2009), ACC synthase gene, encoding an enzyme for ethylene biosynthesis, and *Wip1* gene, encoding a DNA binding protein, were identified as genes for male determination and female determination, respectively. However, these genes or ethylene signaling are not always conserved in sex determination in other plant species.

According to classical genetic or cytogenetic studies, it was shown that sex chromosomes determined sex types in several dioecious or trioecious species. So far, 48 plant species are known to have sex chromosomes (Ming et al. 2011). Similar to

sex determination in mammals, X and Y chromosome are present in 44 of these plants, so that XX and XY represent female and male, respectively (Ming et al. 2011).

An evolutionary model of these sex chromosomes in higher plants was suggested, and it consisted of several stages as follows (Figure 0.1). Dioecious species carrying sex chromosomes were supposed to be derived from hermaphrodite plants, and their X and Y chromosomes were also presumed to be originated from a pair of ancestral chromosome. It was hypothesized that, in the ancestral plant species, a recessive male sterile mutation was introduced in one of the paired chromosomes, and subsequently, at closely linked region to male sterile locus, a dominant female sterility mutation was occurred in the homologous chromosome (stage 1). These mutations inevitably led recombination suppression between paired homologous chromosomes, where two sterility loci were located, because it was impossible to maintain germline carrying both male and female sterile loci by recombination (stage 2). This primary recombination suppression could promote accumulation of mutations and transposon insertions in the corresponding region, which also led genomic rearrangements and degeneration of functional genes. This gene degeneration by frequent mutations, was supposed to be predominant in the chromosome carrying the dominant sterile gene, and accumulated mutations could extend recombination suppression region (stage 3,4). These processes separated X chromosome, carrying functional genes, and Y chromosome (stage 5,6). Although, in mammalian sex chromosomes, recombination was suppressed over the entire chromosome, it was frequently seen in higher plants that limited section of sex chromosome showed recombination suppression. These were possibly intermediates of sex chromosome development as described above. This recombination suppressed region

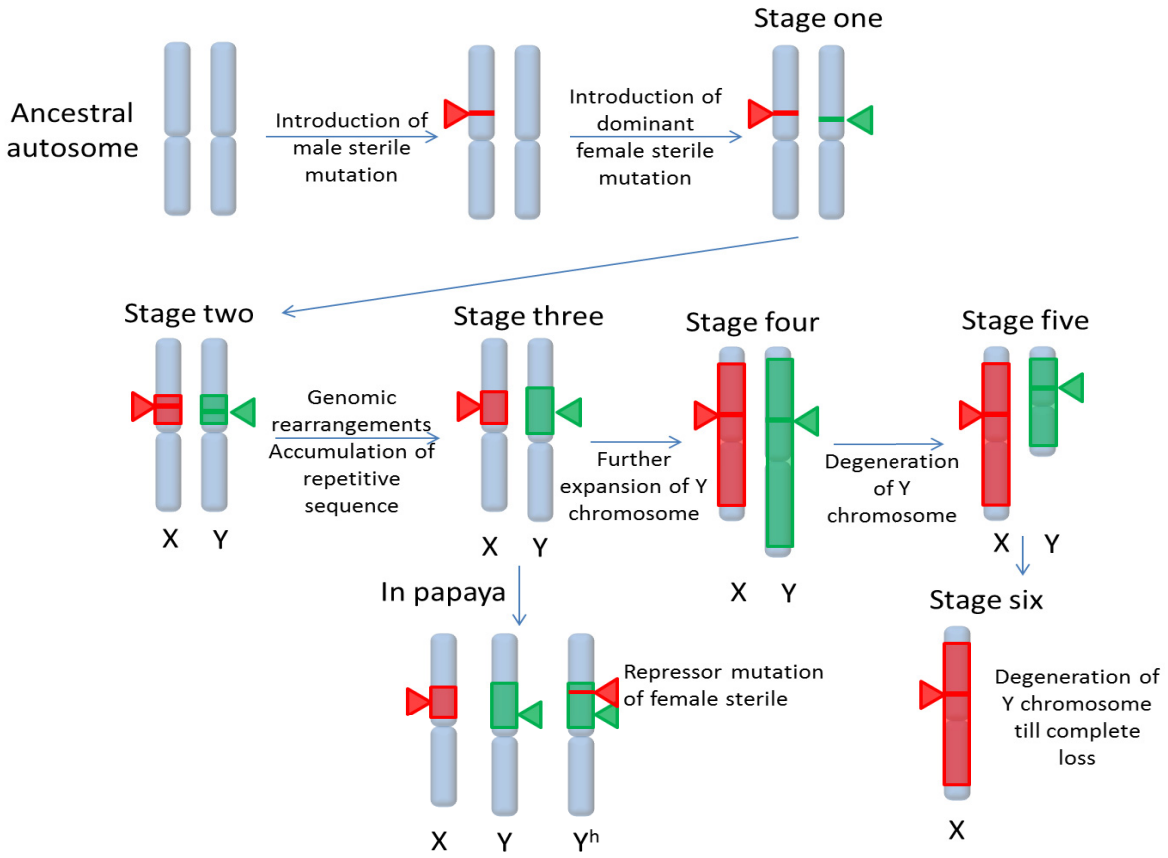

Figure 0.1 Evolutional scheme of sex chromosome development

Stage one: Firstly, a male sterile mutation was introduced, and subsequently, at closely linked region in the homologous chromosome, a dominant female sterile mutation occurred.

Stage two: Recombination suppression occurred due to the two sterile mutations between two chromosomes. A recombinant individual was not most likely to produce offspring. The most primitive sex chromosomes were formed.

Stage three, four: The recombination suppression promotes accumulation of transposons and repetitive sequences, leading to expansion of non-recombining region. Degeneration of genes will occur in the harboring dominant mutation.

Stage five, six: Severe degeneration will shrink the Y chromosome, until the complete loss of Y chromosome.

In papaya, it is hypothesized that the sex chromosomes are in stage three, and the Y^h chromosome derived from the Y chromosome due to a repressor mutation of female sterile mutation in Y chromosome.

in the Y chromosome was designated as males-specific region in Y (MSY).

Among plant species carrying sex chromosomes, dioecious *Silene latifolia* is well studied. Y chromosome of *S.latifolia* has diverged largely from the ancestral chromosome and the non-recombination area is thought to cover most of the chromosome. Despite this genomic complexity, several Y chromosome linked genes, *SlY*, *SlssY*, *DD44Y* and *SlAP3Y*, were successfully cloned from transcripts in male flowers, suggesting a model for the molecular mechanism of its sex determination (Delichere et al. 1999, Nicolas et al. 2005, Filatov 2005).

Sex chromosomes were also found in well-known crop species, including papaya, asparagus, spinach and kiwifruit. Among them, sex chromosome in papaya (*Carica papaya*) has been extensively studied. Papaya belongs to Caricaceae family, which is comprised of 35 species in six genera (Carvalho and Renner 2012, Gschwend et al. 2013). Members of this family are known to be sexually dimorphic; 32 species are dioecious, one is monoecious, and two species are trioecious. Papaya (*C.papaya*) is a trioecious species, which have three sex types, female, male and hermaphrodite (Ming et al. 2007). These three sex types of papaya are genetically determined by the pairing of sex chromosomes (XX, XY, and XY^h for females, males, and hermaphrodites, respectively). Phylogenetic analysis revealed that dioecy is the ancestral state of the family, and the hermaphrodite papaya plants evolved as a natural mutation from male plants and was most likely selected by humans for its pear shaped fruits that are commercially preferred (Carvalho and Renner 2012).

In agricultural production of papaya, the fruits from hermaphrodite plants were commercially preferred due to its pear shaped fruit. Nonetheless, it is impossible to fix hermaphrodite phenotype genetically, because of the heterogametic pair (YY,

YY^h, Y^hY^h are lethal) of sex chromosomes (XY^h). Therefore, elucidation of sex determination mechanisms in papaya will be quite applicable for developing stable hermaphrodite production.

In papaya, its draft genome sequence was already analyzed as BAC (Bacterial Artificial Chromosome) clone sequences in hermaphrodite cultivar, and its genome size was suggested to be 372 Mbp (Ming et al. 2008). These draft genome sequence also include genome sequences of the X and Y^h chromosome. For Y chromosome in male, its partial genome sequence information was limitedly analyzed. Comparative studies of these genome sequences suggested that the recombination suppression between X and Y chromosomes has occurred around 2-3 MYA (Million Years Ago), and the Y^h chromosome has diverged from the Y chromosome around 73,000 years ago (Yu et al. 2007, 2008a, b). Regardless of these genomic studies of sex chromosomes in papaya, any genes for sex determination, including male/female or male/hermaphrodite determination have not yet identified.

Except for genome structure, sex chromosome often shows unique functional regulation as gene dosage compensation in mammals or insect species (Cheng and Disteche 2006, Mank et al. 2009, 2011). In those organisms, female individual carries two X chromosomes (XX), while only a single X chromosome was present in male (XY), indicating that gene dosage of the X chromosome in female was twice of that in male. Therefore, expression level of these genes was expected to be higher in female than that in male individual. However, their gene expression was actually almost equal in both sex types, due to gene dosage compensation mechanisms. Its regulatory mechanism was well-studied, and it was known that one of the X chromosome in mammalian female was totally inactivated by

heterochromatinization (Johnson et al. 1982), while transcription activity in the X chromosome in male (XY) was elevated in *Drosophila* species. In the X chromosome of higher plants, study of gene dosage effect was limited (Muyle et al. 2012). In animal species, male and female individuals show distinctive phenotype other than sexual organs, which is known as sexual dimorphism. On the other hand, phenotypic difference between male and female plants was not so apparent except for flowers. Therefore, analysis of gene dosage effect of X chromosome in papaya might be helpful for understanding indistinguishable sexual dimorphism in higher plants.

This study aimed to identify genes for sex determination in papaya sex chromosomes. In the first chapter, based on transcriptome analysis of the flowers, sex chromosomal genes showing sex-dependent expression pattern were explored. In chapter 2, genome sequences of the papaya X and Y^h chromosome were analyzed for constructing reference sequences, and gene models were predicted. Additionally, gene dosage effect of the X chromosome was evaluated by comparing transcriptome data between female and hermaphrodite. Finally in chapter 3, genome sequences of Y and Y^h chromosomes were compared by resequencing analysis for finding candidate genes for male-hermaphrodite determination. In every study, next generation sequencing technologies were employed, which allowed large scale genomic or transcriptomic analysis.

These studies focused on the genome sequence of non-recombined region in each sex chromosome involving sex determination gene, and this region in Y chromosome was usually called as MSY (male specific region in Y) as described above. However, for convenience, MSY was designated as 'Y chromosome', and corresponding regions

in other sex chromosomes were also designated as 'Y^h chromosome' and 'X chromosome' in this dissertation.

Chapter 1 Transcriptome analysis in papaya flowers for exploring genes for female-hermaphrodite determination

1-1. Introduction

Transcriptome analysis is one of useful tools for identifying candidate gene or genes for particular traits in papaya sex determination, it was expected that this approach was effective for finding responsible genes, because genetic mapping strategy was not applicable to identify the gene on the sex chromosomes, due to their recombination suppression (Ming et al. 2008). Nonetheless, global gene expression analysis (transcriptome analysis) has not yet been achieved in papaya.

As a transcriptome analysis tool, microarray (DNA chip) has been a standard in model-organisms. However, it requires labor and cost for designing and constructing arrays from cDNA or genome sequence data. Recently, emerging next generation sequencing technologies, sequence-based transcriptome analysis come to be a major tool, since its analytical cost was discounted and it was applicable to any life organisms independent of known cDNA or genome sequence data. In fact, DNA sequencing has been employed in transcriptome analysis for fifteen years, as SAGE (serial analysis of gene expression). In this method, a portion of the cDNA was extracted as tag (13bp in original SAGE protocol), and by sequencing sufficient amount of the pooled tags, expression levels of the genes could be estimated (Velculescu et al. 1995). Later, SuperSAGE has been developed (Matsumura et al. 2003, 2005), which could extract a 26 bp long tag. In SuperSAGE, 26 bp long tag sequence was able to accurately determine the origin of tags, making it perfect for transcriptome analysis in non-model plants. After that, High-throughput (Ht-)

SuperSAGE (Figure 1.1) was introduced by combining the SuperSAGE with the next-generation massive parallel sequencing platform (Matsumura et al. 2010, 2011).

Using the Ht-SuperSAGE method, Urasaki and colleague in Okinawa prefectural agriculture research center tried to transcriptome analysis of papaya flower buds for identifying any genes for sex determination. RNAs were extracted from flower buds at two developmental stages of female, hermaphrodite and male papaya plants (Figure 1.2). SuperSAGE library for each RNA sample was prepared and sequenced by Applied Biosystems SOLiD DNA sequencer. Consequently, 1,545,624 tags (26-bp)on average were obtained in each sample (Table 1.1). In the present study, these SuperSAGE tags were annotated by mapping to BAC clone sequences of papaya sex chromosomes (Ming et al. 2008, Yu et al. 2007) for speculating their functions in sex determination.

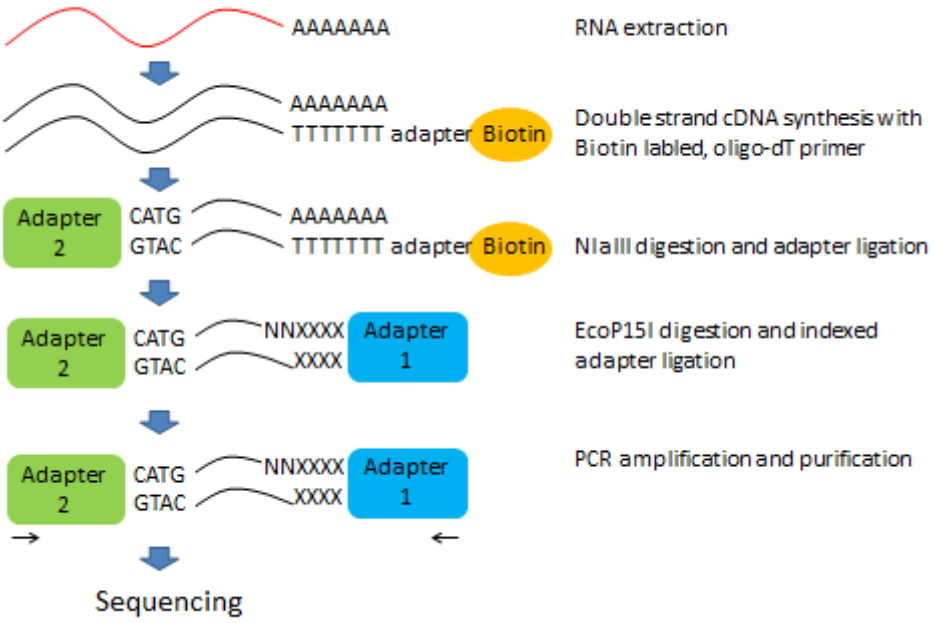

Figure 1.1 Scheme of High-throughput SuperSAGE

Double-stranded cDNA was synthesized using the biotinylated adapter-oligodT primer. Purified cDNA was digested with NlaIII. The resulting fragments were bound to streptavidin-coated beads, and non-biotinylated cDNA fragments were removed by washing. Adapter-2 was ligated to cDNA fragments on the beads and was digested with EcoP15I after being washed. EcoP15I-digested fragments were ligated to adapter-1 with defined index sequences for sample identification. Tags sandwiched between two adapters were amplified by PCR, and applied to sequencing.

Table 1.1 Summary of Ht-SuperSAGE analysis

Sample code	Flower buds analyzed	Index sequence	Number of total tags	Number of unique tags	Number of non-singleton tags
P1	Male early-stage	ACACAA	1,886,181	190,431	43,441
P2	Male late-stage	ATAGAG	1,585,205	200,775	42,921
P3	Female early-stage	AGAGTG	1,857,405	390,915	54,694
P4	Female late-stage	AGGCTG	1,410,279	179,994	40,297
P5	Hermaphrodite early-stage	CCGAAG	1,130,499	113,254	32,657
P6	Hermaphrodite late-stage	CGGATG	1,404,175	139,087	39,505
	(Total)		9,273,744	1,214,456	253,515

1-2. Materials and Methods

Plant materials

The plant materials used in this study are the Hawaiian papaya cultivar 'Sunrise solo' and breeding material TM1 derived from a cross between cv. 'Wonder frea' and Okinawan land race 'IG4', which is locally grown line in Okinawa.

RNA preparation

The female and hermaphrodite flower buds were harvested from the Sunrise solo and the male flower buds were from the TM1. Early-stage (no longer than 7 mm) and late-stage (approximately 20 mm in length) flowers were harvested from male, female and hermaphrodite plants (samples P1-P6, Figure 1.2). Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen). For cDNA synthesis, 10 µg total RNA per sample was used.

Data analysis

C. papaya whole-genome sequences were downloaded from GenBank (ftp://asgpb.hawaii.edu/papaya/assembly/supercontigs.filtered_012808.fasta). Furthermore, BAC clone sequences of the sex chromosomes were also downloaded from the NCBI website following previous studies of papaya. Tag sequences appearing more than 10 times in all the samples were applied to BLASTN searching as a query against BAC clone sequences for sex chromosomes (Table 1.2) or papaya whole genome sequences to investigate redundancy. Only the tags showing a perfect match were regarded as SC-tags (sex chromosome tags). For annotation of genes corresponding to the tags, a 1 kbp genome sequence upstream

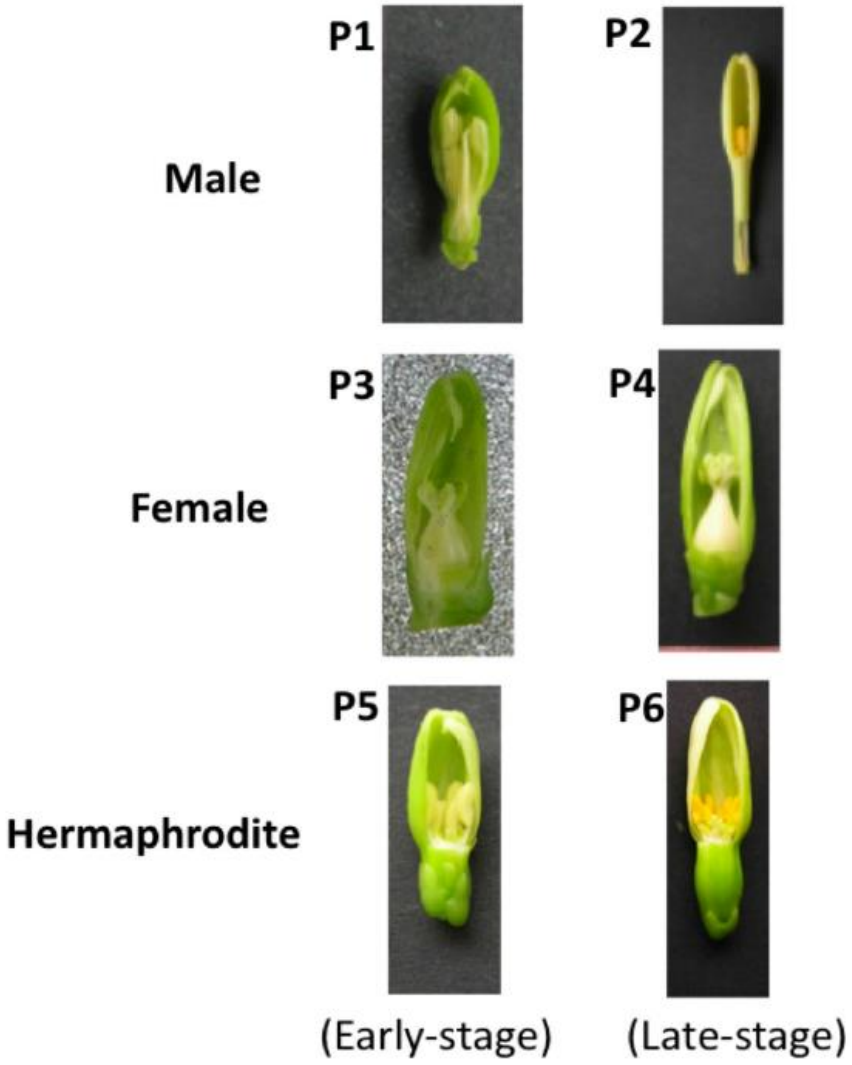

Figure 1.2 Flower samples for the Ht-SuperSAGE analysis.
The papaya flowers used for the RNA extraction and Ht-SuperSAGE analysis (P1: male 7-mm flower; P2: male 20-mm flower; P3: female 7-mm flower; P4: female 20-mm flower; P5: hermaphrodite 7-mm flower; and P6: 20-mm hermaphrodite flower).

Table 1.2 List of papaya sex chromosome derived BAC clones used in the study

Sex chromosome	BAC clones (Accession No.)
Y^h chromosome	53E10(FJ429366), 66N15(FJ429369), 75H06(FJ429370), 90D06(FJ429371), 99O03(FJ429372), 102N12(FJ429373), PH41F24(EF625817), PH42B05(EF625818), PH54H01(EF625819), PH76M08(EF625820), PH85B24(EF661024), PH94E22(EF625821), PH95B12(EF661025), 17J04(AC238761), 66N15(AC238768), 72J22(AC238620), 89I15(AC238633), 27G03(AC239146), 69C01(AC238618), PH76M08(EF625820), 83L14(CG680755 - CG680757, CG680765 - CG680777), 71E16(AC239163), 67C23(CG680534 - CG680559), 83D03(CG680732 - CG680754), 79C23(AC238625), 31E20(CG680477 - CG680485), 95F17(CG680952 - CG680975), PH89M06(AM778096), 89A05(CG680805 - CG680827), 74J19(CG680614 - CG680627), 80L24(AC239164), 93J16(CG680850 - CG680871), 96C10(CG680994 - CG681019), DM62K05(EU369761), 95D22(CG680927 - CG680951), 93O24(CG680872 - CG680899), 82H21(AC238627.1), 52B03(CG680503 - CG680515), 33L10(CG680486 - CG680502), 52H15(AC238762),
X chromosome	PH53E18(EF661026), PH61H02(EF661023), DM10G24(EU369760), PH12G20(GU799580), PH95I24(GU799579), PH89P09(GU799578), 99B01(AC238637), 87M18(AC238632), 85C03(AC238631), 84J07(AC238629), 83B06(AC238628), 80F18(AC238626), 54M13(AC238609), 46O19(AC238599), 40J04(AC238598), 39A12(AC238597), 31E12(AC238595), 12I03(AC238590), 51A03(AC238603), 23D07(AC238592), 08K16(AC239251), 65C15(AC239160), 30B21(AC239147), 23D15(AC239144), 10J17(AC239138), 79M13(AC239204), 54M22(AC239202), 50J21(AC239200), 49A07(AC239150), 06M23(AC239136), 93K15(AC239170), 84M10(AC239166), 69D05(AC239161), 65C06(AC239159), 49N10(AC239151), 38K02(AC239149), 24L08(AC239145), 22E07(AC239143), 136D11(AC239253) 54A04(AC239201)*, 69D24(AC239162), 54A09(AC239152)*

* This clones were regarded as BAC clones on the X chromosome

of 5'-CATG in each individual tag sequence was applied to the BLASTX search as a query against all registered protein sequences.

PCR amplification of genes corresponding to the tags

Genomic DNA was extracted from mature leaf for each papaya plant using the DNeasy Plant Mini Kit (Qiagen). Single-stranded cDNA was synthesized from the total RNA of samples P1 to P6 in by reverse-transcription using adapter-dT primer or oligodT primer. These cDNAs were employed for 3'-RACE and RT-PCR as template, respectively.

1-3. Result

Brief summary of Ht-SuperSAGE analysis of papaya flowers

Ht-SuperSAGE analysis from RNA preparation to 26-bp tag sequencing was carried out by Urasaki and colleague in Okinawa prefectural agriculture research center. Brief summary of the analysis was as follows.

Total RNA was extracted from two developmental stages in each male, female and hermaphrodite plant, and the RNA samples were designated as samples P1 to P6 (Figure 1.2). Using these six RNA samples from papaya flowers, double-stranded cDNA was synthesized and applied to 26-bp tag extraction from NlaIII sites in the cDNA using EcoP15I restriction enzyme. Extracted 26-bp tags with adapter DNA were once PCR amplified and applied to DNA sequencing using SOLiD sequencer.

In total, 9,273,744 reads were selected as SuperSAGE tag-encoding sequence reads, and 26-bp tag sequences were extracted from NlaIII sites, and their counts were calculated to obtain an expression (tag) profile for each sample. The total tag counts and the number of unique tags in each sample are shown in Table 1.1. From each of the samples, more than one million tags (with an average of 1,545,624 tags per sample) were obtained.

Annotation of SuperSAGE tags

By comparing SuperSAGE tag profiles (count of each tag) among the six flower samples (P1-P6), many differentially expressed tags were found. These tags may include many expressed genes involved in the development of male or female floral organs or crucial genes for sex differentiation or determination. To identify key expressed genes for sex determination from these data, tags were mapped on the

sex chromosomes in papaya (Figure 1.3). Tags occurring more than 10 counts in all the samples combined were selected for mapping to BAC clone sequences of the sex chromosomes (Table 1.2). BAC clones of X and Y^h chromosome genome sequences were derived from GenBank database. Against these sequences as databases, selected 26-bp tag sequences as query were applied to BLAST searching. In this analysis, no mismatch in 26 nucleotides with BAC clone sequences was accepted (only the completely matched sequences were allowed). In total, 456 unique tag sequences showed a perfect match to the BAC clone sequences of the sex chromosomes (X or Y^h). These tags are designated as SC-tags (sex chromosome tags). Of these tags, 312 were uniquely mapped on the sex chromosomes. The remaining SC-tags showed a perfect match to nuclear gene sequences not found on sex chromosomes or to chloroplast or mitochondrial gene sequences in papaya. Among the 312 sex chromosome specific tags, 31 unique tags were mapped on both the X and the Y^h chromosomes. Most of the tags mapped to only the Y^h chromosome were absent in female flower samples as expected, but 12 tags were unexpectedly observed in female samples, despite the lack of a Y^h chromosome. Tags located on only the X chromosome were usually expressed in all samples, and their tag counts were mostly equal among sex types. However, some tags on the X chromosome showed sex-dependent expression patterns as described later. To identify putative functions of the corresponding genes to the SC-tags, homologous proteins were investigated approximately 1 kbp upstream of all 456 SC-tags using BLASTX searching because the tags were located at the closest NlaIII site to the polyA tail of cDNA. All the SC-tags were categorized based on their estimated function of annotated corresponding genes (Figure 1.4). More than 20% of the SC-tags were

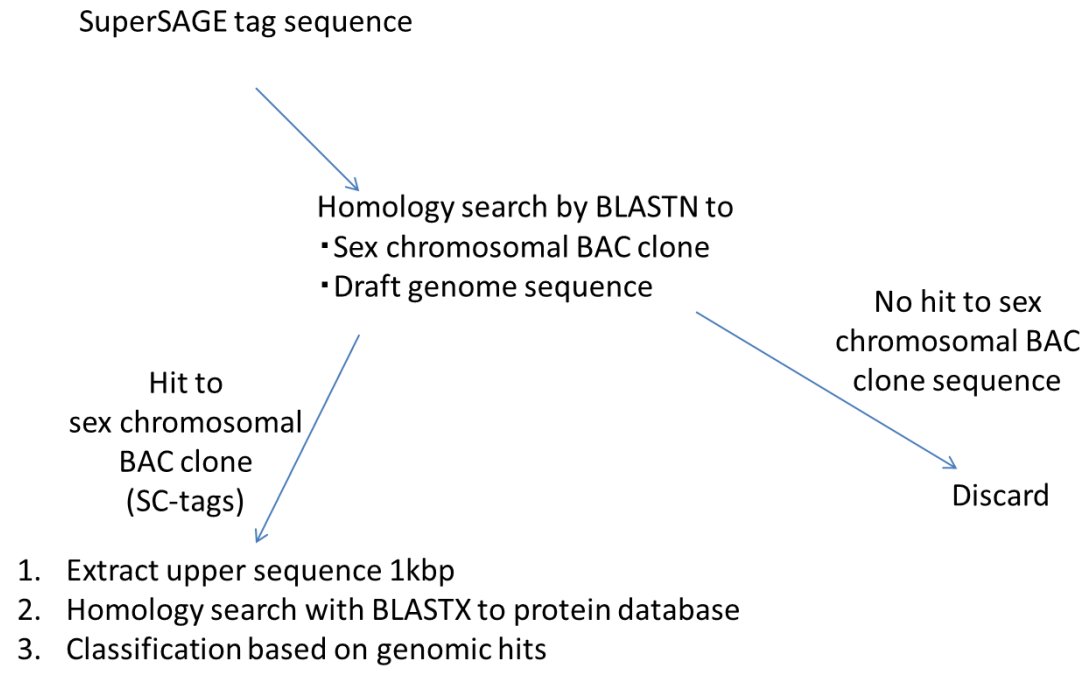

Figure 1.3 Flow of SuperSAGE tag annotation

Homology of the tag sequence was searched against the sex chromosomal BAC clone sequences indicated in Table 1.2, and draft genome sequence of papaya. Tags showing complete match to the sex chromosomal BAC sequences were selected, and DNA sequence 1 kbp upstream of the tag was extracted, and used for BLASTX searching against all registered protein in Genbank database.

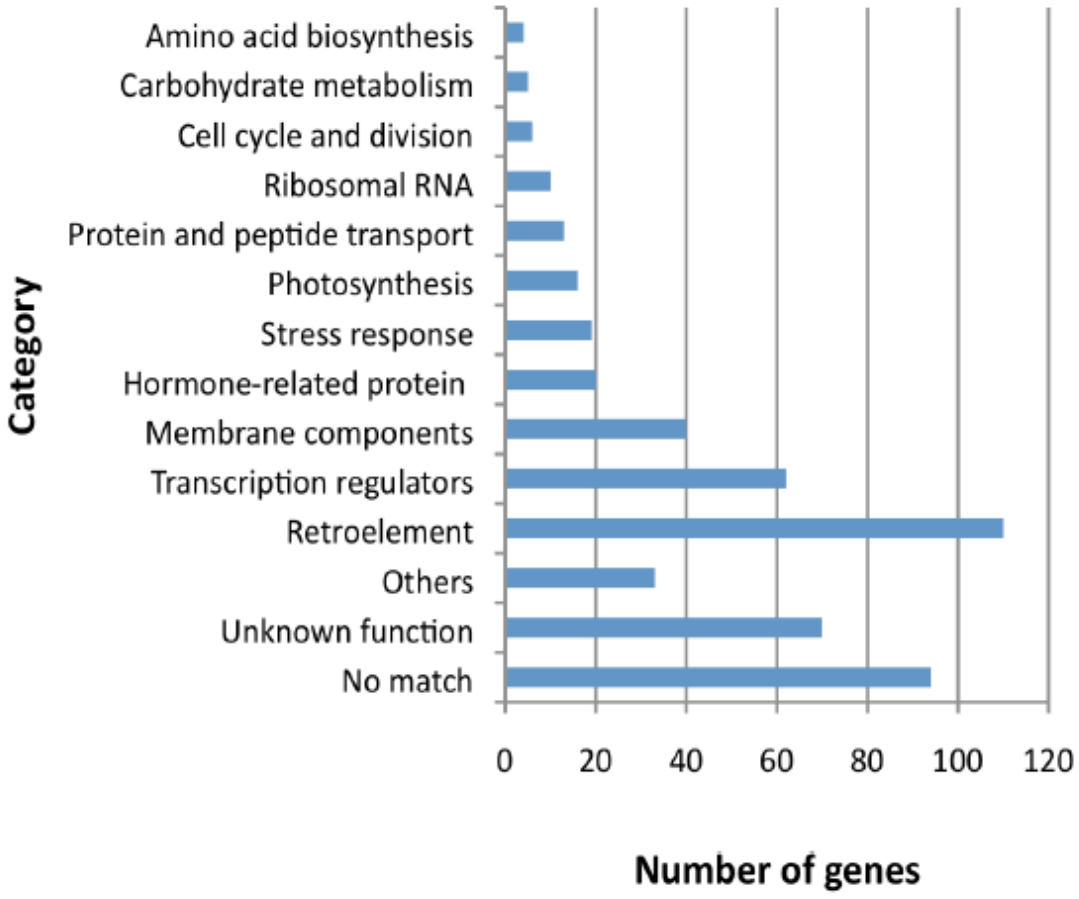

Figure 1.4 Functional categories of genes annotated from SC-tags

Genes for 456 SC-tags categorized by function, based on their annotation as described in the methods. "No match" indicates tags having no significant homologies with ESTs and non-redundant protein sequences in databases. "Unknown function" indicates genes for which the functions of their encoded proteins could not be identified. SC-tags having different annotations were grouped into the corresponding categories.

derived from retroelements (including retrotransposons, which are known to occupy sex chromosomes, Zhang et al. 2008). They were found throughout the sex chromosomes and were frequently seen on other chromosomes. It is unique that transcriptional factors were a major group among the SC-tags. The SC-tags also corresponded to various other categories of genes. Nevertheless, many genes corresponding to SC-tags were functionally unknown or did not show significant homology to registered genes at the amino acid level.

Differentially expressed genes in the sex chromosome genes

Of the 456 SC-tags, 47 tags (Table 1.3) were selected that were specifically expressed in one or two sex types. These tags were designated as sex-dependent SC-tags. Count of these tags in each sample was less than 100, showing a relatively low level of expression. Most of these tags showed specific expression in males and hermaphrodites or in females and hermaphrodites, except for a few tags that were limited to male or hermaphroditic flowers. However, no female-specific tags were observed in this study. As described above, most of the tags mapped on Y^h were seen in hermaphrodites or in both males and hermaphrodites. As a validation of these results from SuperSAGE, PCR analysis was carried out for a few selected genes to sex-dependent SC-tags. RT-PCR analysis demonstrated that expression of genes for the Cp2671 tag (encoding a MADS-box protein) was male and hermaphrodite specific, and its absence in females was confirmed by genomic PCR amplification using several different primer sets (Figure 1.5). Several genes on the X chromosome also showed sex-dependent expression, even though all plants should have an X chromosome. Using RT-PCR and sequencing of amplified fragments, it was shown

Table 1.3 SC tags showing sex specific expression patterns

Tag ID	Tag sequence	BAC clone	Chr.	Tag count of samples						Accession no., deduced protein*
				P1 (XY)	P2(XY)	P3(XX)	P4(XX)	P5(XY^h)	P6(XY^h)	
Cp2671	catgtttgcaagtacagagtttgcca	71E16	Y^h	261	159	0	0	18	41	ACY82403, MADS-box protein
Cp4997	catgcgagagtgtagcgcattgtata	PH41F24	Y^h	87	65	0	0	34	46	No significant homologies found.
Cp7929	catgtatccacggcaccacttcagtt	71E16	Y^h	0	122	0	0	0	0	ACY72569.1, retrotransposon
Cp8602	catgtttctcccaacgaacaaactcc	PH85B24	Y^h	29	15	0	0	26	38	XP_002263526.1, hypothetical protein
Cp11430	catgtcgtcaaggacaactgcaggct	90D06	Y^h	45	20	0	0	4	1	No significant homologies found.
Cp12204	catgttgtagcattgctcgtgctagg	90D06	Y^h	0	0	0	0	22	42	XP_002521454.1, monodehydroascorbate reductase
Cp13833	catgcacaggtgggttgtttgcgcgt	89I15	Y^h	0	0	0	0	42	10	No significant homologies found.
Cp14501	catgaaacgcaaagctggaaatagaa	96C10	Y^h	33	8	0	0	4	2	No significant homologies found.
Cp15892	catgactcttcccgcactctgatgaa	PH42B05	Y^h	0	0	6	11	2	22	1510387A, retrotransposon del1-46
Cp18724	catgaaggaaaagggaaagagggtgg	PH94E22	Y^h	14	6	0	0	6	4	No significant homologies found.
Cp21244	catgctttgtcctgcactctgatgag	99O03	Y^h	16	7	0	0	1	0	AAL79340, Ty3-Gypsy type retrotransposon
Cp25137	catgtgatccaaatccgaggggttg	PH85B24	Y^h	5	5	0	0	4	4	ABS32231, phosphoglucosamine mutase
Cp27852	catgcgtgacgacaaattggccgtga	72J22	Y^h	4	1	0	0	9	1	No significant homologies found.
Cp28108	catgtcgtgctccatagagggagtga	72J22	Y^h	2	6	0	0	2	5	No significant homologies found.
Cp29229	catgctagtcagaagcacagaagtaa	90D06	Y^h	2	8	0	0	3	1	XP_002521454, monodehydroascorbate reductase
Cp19081	catgcatataatttgaacatcgtgct	PH95I24	X	0	0	7	3	5	15	XP_002513875, conserved hypothetical protein
Cp19647	catgatggataaatgttctgttgctt	54A04	X	5	2	15	6	0	0	XP_002514770, inositol 5-phosphatase

Cp30711	catgtacagaaaggttgctgtaaatt	54A09	X	1	0	8	4	0	0	AAG25966, cytokinin-regulated kinase 1
Cp35914	catgggagcagctgatgctgttcgtc	54A04	X	0	1	9	0	0	0	XP_002514770.1, inositol 5-phosphatase
Cp11249	catgataatatgcagtttgtgtttat	69D05	X	0	0	37	24	2	10	NP_177343.2, zinc finger protein
Cp18004	catgtcaccaagaaatgaaagagata	3.1E+13	X	0	33	0	0	0	0	AAD51623, seed maturation protein PM35
Cp18118	catgagaactcgtgtggtactccggt	85C03	X	9	21	2	0	0	0	No significant homologies found.
Cp22551	catggtggaggaaggtggcgttgccc	50J21	X	4	1	10	7	0	0	AAF79262, dihydrolipoamide succinyltransferase
Cp22813	catgtttacagagataaaaataaaca	49N10	X	0	0	12	8	1	1	NP_198850, PWWP domain-containing protein
Cp24896	catggtctgtaattgatggtagaggt	99B01	X	0	0	14	2	3	0	No significant homologies found.
Cp25457	catgatttggaaagcaaaggcctcac	54M13	X	2	1	10	5	0	0	CAN78776, hypothetical protein
Cp25520	catgtatatgtaaattgtaaagaagg	54M22	X	1	11	4	2	0	0	XP_002874566, DNAJ heat shock protein
Cp25578	catgtaaattgtaaagaaggtttctg	65C06	X	0	0	8	0	4	6	XP_002874566, DNAJ heat shock protein
Cp25721	catgttatttgtgttgggttggtgtg	PH95I24	X	0	0	8	2	6	2	NP_173630, structural constituent of ribosome
Cp26599	catgagctgaagctcgcaactggaga	PH95I24	X	0	0	1	7	5	4	XP_002513881, Homomeric Acetyl-CoA Carboxylase
Cp26618	catgcaaagtgaacaggtggaggtcc	PH95I24	X	0	17	0	0	0	0	XP_002513872, MADS box protein
Cp27160	catgcgcctccgtgacccggtctacg	PH53E18	X	2	0	0	0	14	0	ABS32237, asymmetric leaves 2
Cp29051	catggcgcggttgcggactagacaag	PH95I24	X	3	8	0	0	0	3	ABB59581, putative sulfate transporter
Cp29228	catgtactcataaatacagaaatacc	PH61H02	X	2	2	6	4	0	0	ABY60778, exocyst complex subunit SEC6
Cp30075	catggatgtcgtagtacgaccaagga	PH53E18	X	4	6	2	1	0	0	ABS32235, protein kinase
Cp30376	atgttagagaacagaaagagagaag	65C06	X	2	4	5	2	0	0	NP_173675, C2 domain-containing protein
Cp31170	catgctaacaccaaacaaagagcatc	84J07	X	6	0	5	1	0	0	XP_002300037, predicted protein
Cp31460	catgtaggtgtaccaaaatccgtggt	30B21	X	3	6	1	2	0	0	CBI26540, unnamed protein product

Cp31776	catgacaggatggcttcattctctta	65C06	X	2	3	5	2	0	0	XP_002521816, radical SAM protein
Cp31823	catgtaaactaagcacaaggtcttga	23D07	X	2	0	5	5	0	0	No significant homologies found.
Cp32380	catggcactgagtcagcgtgtcttgc	80F18	X	0	5	0	7	0	0	XP_002271693, retrotransposon protein Ty3-gypsy
Cp33154	catgtagttgaaggagataatgtaga	69D05	X	2	5	3	1	0	0	XP_002526800, alpha-(1,4)-fucosyltransferase
Cp33648	catgtcattactggctggcgctttat	93K15	X	1	3	7	0	0	0	XP_002887402, hypothetical protein
Cp34050	catgccttttccattcctcatacaca	06M23	X	0	0	7	1	1	2	AAO23078, polyprotein
Cp34585	catggtgtgcttgtaaatgctcacta	PH95I24	X	4	1	3	2	0	0	XP_002513881, Homomeric Acetyl-CoA Carboxylase
Cp35918	catgagaaagctagacagcatatcca	06M23	X	0	0	8	1	0	1	AAM94350.1, gag-pol polyprotein
Cp35942	catgagcatattgttatatatgatta	50J21	X	0	4	5	1	0	0	XP_002529005.1, membrane protein

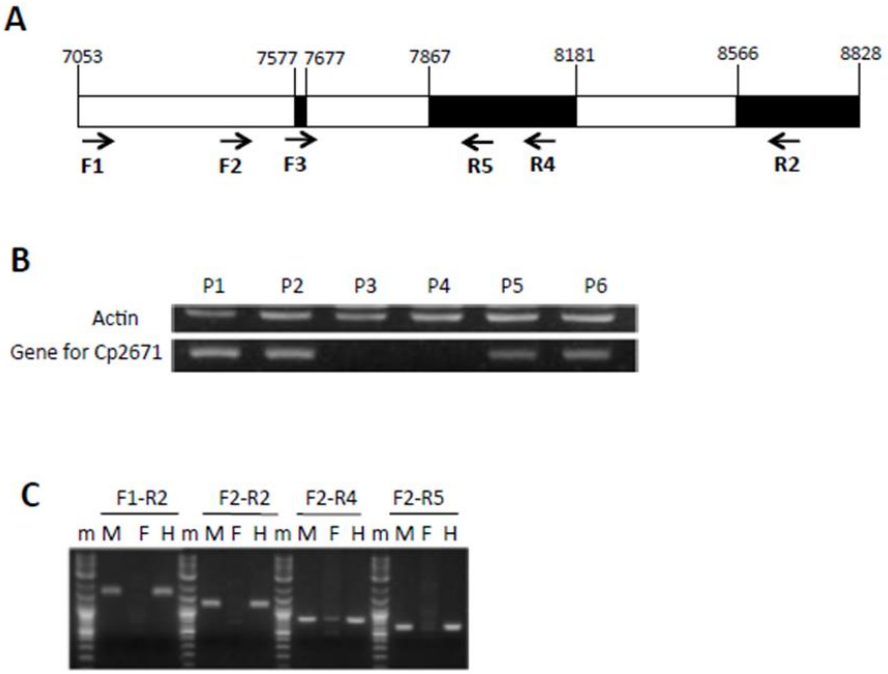

Figure 1.5 Confirmation of sex-dependent gene expression for Cp2671, which encodes a putative MADS box protein.

A: Structure of the gene corresponding to Cp2671 tags on BAC clone 71E16. Arrows indicate the locations of PCR primers. Black regions represent putative exons.
B: RT-PCR with primers Cp2671-F3 and Cp2671-R2 using RNA from the same tissue samples as for Ht-SuperSAGE (Figure 1.2). An actin gene was used as a control constitutively expressed gene.
C: Genomic PCR amplification of the gene for Cp2671. In the gel picture, m indicates a lane of DNA size markers, 2-Log DNA Ladder (New England Biolabs). M, F and H represent male, female and hermaphrodite, respectively

that differential expression of the Cp11249 tag on the X chromosome was caused by sequence polymorphisms between male plants and the other sexes (Figure 1.6).

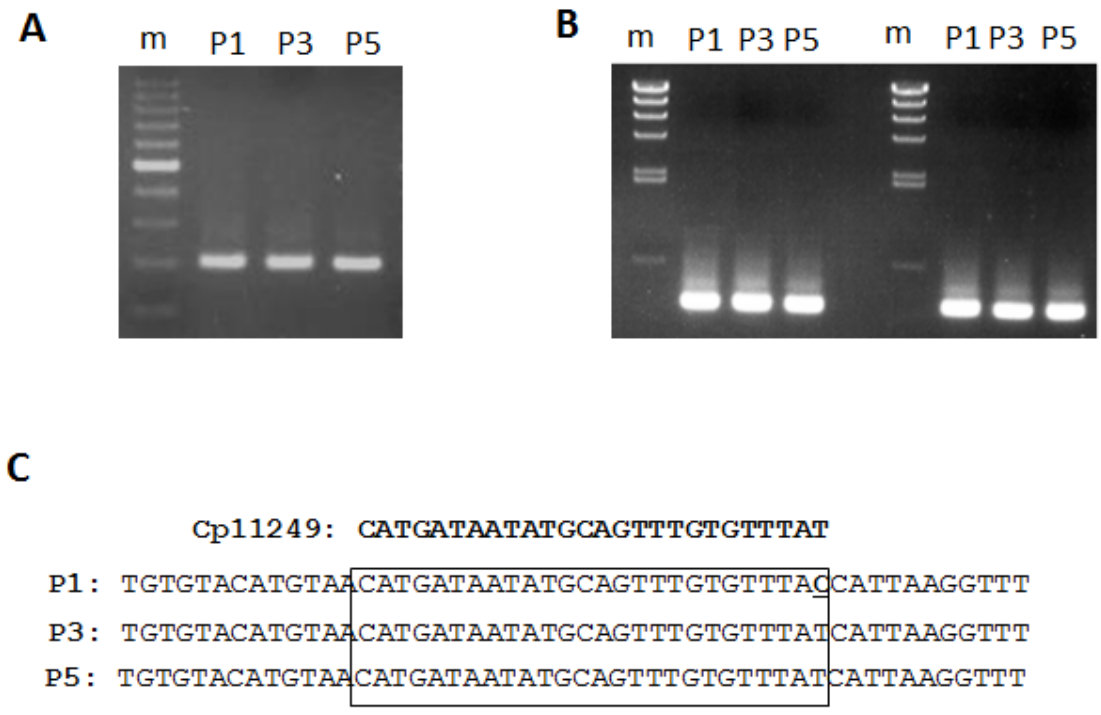

Figure 1.6 Amplification of the Cp11249 tag

SuperSAGE tag Cp11249 showing sex specific expression and locating on X chromosome was amplified by RT-PCR. A: RT-PCR of Cp11249 gene on P1, P3 and P5 sample described in Figure 1.2.

B: Replicate of the RT-PCR conducted in Figure1.6A.

C: DNA sequence of the amplified RT-PCR product. Underlined nucleotide shows the detected sequence polymorphism.

1-4 Discussion

In papaya, although genetic and structural genomic studies of sex determination have been undertaken, information about gene expression was limited. In the present study, combining the genomic data as BAC clone sequences of the sex chromosomes with transcriptome data as SuperSAGE tags was succeeded in narrowing the target genes for further analysis from thousands of tags by mapping tag sequences to the primitive sex chromosomes. Finally, 47 differentially expressed genes on the sex chromosomes (sex-dependent SC-tags) were selected as candidates for elucidating sex determination in papaya.

According to observations of papaya flower development, immature male flowers were morphologically similar to those of hermaphrodites, whereas female flowers were distinctive (Figure 1.2). This suggests that there are two important sex determination points during flower development. The first is the determination of female or other sexes, and the second is the differentiation of males and hermaphrodites.

The absence of female-specific expressed genes among sex-dependent SC-tags implies that genes for female organ development are necessary in any sex type, at least in the early stages of flower development, and that they are located on the X chromosome or autosomes. Genes on the Y or Y^h chromosomes would direct the transition from female to other sexes. Flower development in papaya was expected to follow the ABC model, a well-known genetic regulation model in higher plants because homologs of B-class MADS-box genes in this model were isolated in papaya (Ackerman et al. 2008, Alvarez-Buylla et al. 2010). Although it could be supposed that the Y and Y^h chromosomes carry genes directly related to stamen development,

the annotation of 47 genes for sex-dependent SC-tags did not reveal any homologs to known genes of the ABC model, predicting that sex determination genes are unique in papaya. Therefore, genes for primary sex determination would have regulatory functions of initiating differentiation between male or hermaphrodite and female, and they are expected to control the expression of other genes for stamen development in transformation. A similar sex determination system is known in *S. latifolia*. In this species, the presence of a Y chromosome determines sex differentiation by controlling floral organ development (Zluvova et al. 2005). It was suggested that a putative *SPF* (*stamen-promoting factor*) gene on the Y chromosome repressed expression of the *Superman-like* (*SISUP*) gene for female organ development (Kazama et al.2009), which was also a suppressor of the gene for stamen elongation (*SLM2*). Among the genes corresponding to sex-dependent SC-tags in papaya, several genes, including functionally unknown genes, can be regarded as candidate genes for regulation. In particular, a gene corresponding to Cp2671 tags on the Y^h chromosome encoded a putative MADS-box domain protein and showed male- and hermaphrodite-specific expression. According to its similarity to known genes, the gene for Cp2671 showed homology to the SHORT VEGETATIVE PHASE (SVP) protein gene (Hartman et al. 2000), which transcriptionally regulates the expression of the gene for *FT* (Flowering locus T, Jang et al. 2009). Other than this putative gene, a Y^h -specific regulatory gene must be a candidate gene for the conversion of females to other sexes.

As a secondary step for papaya sex determination between hermaphrodites and males, Y-specific or Y^h -specific genes were expected to be involved. The structural differences between the Y and Y^h chromosomes are probably not very distinctive,

since SuperSAGE tags on Y^h were frequently observed in male samples. The developmental shift from young hermaphroditic flowers to mature male flowers would be accomplished by suppression of female organ development. It was presumed that candidate regulatory genes for female organ suppression were specifically expressed in later stages of male flower development. The Cp7929 tag, corresponding to a putative retrotransposon gene, was the only tag showing a unique expression in later stages of male plants, although it is impossible that it could play a positive role in male flower development. On the contrary, several Y^h chromosome genes expressed only in hermaphrodites were observed in the present SuperSAGE data, indicating that the repression of their expression might contribute to the retardation of female floral organs in mature male flowers. In the present gene expression analysis, candidate genes for secondary sex determination were actually limited. Deciphering Y chromosome sequences is a prerequisite for elucidating the mechanisms that differentiate hermaphrodites from males.

Among the transcripts from sex chromosome regions in flowers, retroelements were the most frequently observed category of genes, in accordance with their high frequency in the genome. In 2008, Zhang et al. reported the observation of four knob-like heterochromatin structures specific to the Y chromosome and revealed that the DNA sequences associated with the heterochromatic knobs were heavily methylated compared with the sequences in the corresponding X chromosome domains. Therefore, retroelements in the Y^h chromosome are thought to be inactivated by DNA methylation. The results of the present study demonstrate that some of these retroelements genes, specifically those located on sex chromosomes, were at least transcriptionally active, although it is unknown whether they

underwent *de novo* transposition in the papaya genome. It is not possible for the retroelements themselves to have any positive functions in the process of sex determination. Instead, their insertion would knock out genes that might be responsible for sex determination.

In the present study, 456 SC-tags were the first information of expressed genes on papaya sex chromosomes with quantitative data on transcription levels, implying characteristics of structure and gene expression regulation in sex chromosomes.

Approximately 30% of unique tags were redundantly present in the papaya genome on non-sex chromosomes; tags primarily encoded retroelements. Notably, similarities to chloroplast or mitochondria genome sequences were also found on the Y^h chromosome. Although further experimental analysis is required to define the corresponding genes, we have to carefully confirm which gene the tag was derived from. Tags specifically mapped on the Y^h chromosome were not expected to appear in female samples. However, several tags on the Y^h chromosome showed inconsistent expression in females (Urasaki et al. 2012). For this inconsistency, several causal possibilities could be considered, such as possible redundancy of these tags on other undetermined genome sequences. Also, some genes on BAC clones 54A09 and 54A04 were observed in female samples. We deduced these genes to be BAC clones on the X chromosome, judging from their sequence similarities to other clones on sex chromosomes, although they are annotated as Y chromosome clones.

For analysis of selected genes for sex-dependent SC-tags in the present study, differential expression patterns were derived from structural variation among sex

chromosomes, such as sequence polymorphism, in/del of genes or retroelement insertion. It was obvious that Y- or Y^h-specific genes were not present in females owing to a lack of Y or Y^h chromosome as described above. Even some genes on the X chromosome showed sequence polymorphisms between males and non-males, representing sex-dependent expression of the SuperSAGE tags of these X-chromosome genes (Figure 1.6). This was the result of different genetic backgrounds in our male plant material compared with the other sexes. The female and hermaphrodite materials were offspring from the same hermaphrodite plant and thus shared an identical genetic background other than the Y^h chromosome. Still, many tags on the X chromosome did not appear in either females or hermaphrodites. These tags are expected to be derived from transcripts that were differentially regulated at the expression level.

In the present study, it was succeeded to find several genes, which might be concerned to sex determination in papaya. However, genome structure of papaya sex chromosomes was complicated and further comparative analysis of sequences of genes among sex chromosomes was necessary for identifying genes for sex determination. For these studies, available BAC clones sequences of the X and Y^h chromosomes were insufficient, and more accurate genome sequences, including the Y chromosome, were required.

Chapter 2 Structural and functional genomic study of X and Y^h chromosome

2-1. Introduction

Evolutionary model of sex chromosomes predicted that the suppression of recombination in sex determination region between the chromosomes will undergo genetic degeneration in either chromosome, due to accumulation of deleterious mutations and transposable elements (Charlesworth and Charlesworth 2000). In mammalian sex chromosomes, which evolve about 166 million years ago, 95% of the Y chromosome is occupied by the non-recombining region (Veryrunes et al. 2008). In the non-recombining region in human sex chromosome, only 78 protein coding genes were found, while 1,098 genes were located in the corresponding region of the X chromosome. These 78 genes encode 27 different kinds of proteins of which 16 genes have X-linked homologs or alleles (Bhowmick et.al 2007). Still, composition of genes on the papaya sex chromosomes was unspecified, but comparative studies of these genes between the sex chromosomes will be greatly helpful for understanding evolution and sex determination mechanisms.

In the first chapter, expressed genes on the Y^h and corresponding region in the X chromosome were identified by mapping 26-bp SuperSAGE tags to BAC clone sequences. Although the sequence information of papaya sex chromosomes were available as BAC clone sequences in the public database, it was difficult to elucidate gene structures thoroughly, because there were frequent gaps among these disordered BAC clones, indicating that genome sequences of papaya sex chromosomes remain uncompleted.

The draft genome sequence of papaya was constituted of the BAC clones (Ming

et al. 2008), which were sequenced by traditional sanger sequencing method with low coverage (6x). Recently, next generation sequencing (NGS) technologies provide massive sequence information with high quality. They allowed to produce draft genome sequence of many plant species with deep coverage of sequence data. (van Bakel 2011, Wang et al. 2011, Wu et al. 2014). Therefore, NGS technology must be greatly helpful for reconstructing more precise genome sequences of papaya sex chromosomes for further structural and functional studies. Additionally, since draft genome sequence of papaya was constructed from a cultivar, which was different from that in the present study, genome sequence data of own material was necessary as a reference.

Then, pseudo-molecules of the sex determination regions on X and Y^h chromosomes as reference sequences, which could be a standard for further studies, were newly constructed. For this purpose, whole genome sequences of female and hermaphrodite plants were analyzed using illumina sequencing technology. Based on these newly developed reference sequences of papaya sex chromosomes, predicted genes were defined for identifying structural difference between X and Y^h chromosomes. As shown in sex chromosomes of animals, structural differences between X and Y or Y^h chromosomes cause difference of gene dosage on the X chromosome between female and other sex-types (Disteche 2012, Julian et al. 2012). In mammalian females (XX), genes on one of the X chromosomes are totally repressed by heterochromatinization of whole chromosome (Figure 2.1), leading that their gene dosage was adjusted to the equal level of that in male (XY) individuals (Johnson 1982, Xiong et al. 2010). Gene dosage compensation on the X chromosome was also observed in *Drosophila* (Mank 2009), but genes on the X

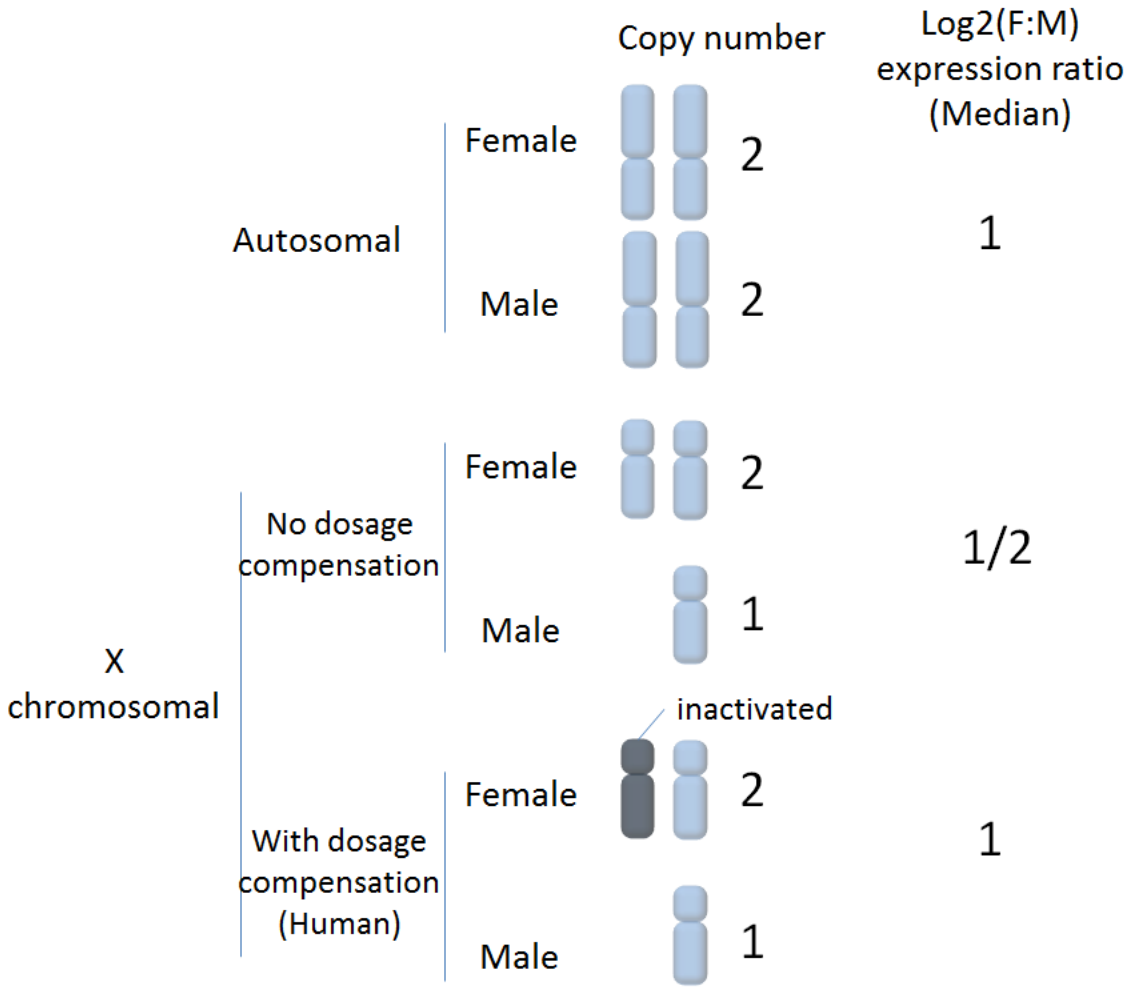

Figure 2.1 Model of gene dosage compensation in sex chromosomes

In autosomal genes, there are two alleles (maternal and paternal alleles). Therefore, their gene dosages are equal in both male and female individuals.

In the X chromosomal genes (XY system in this case), two alleles were present in female (XX), but male (XY) has only a single allele, indicating that its gene dosage was different between female and male. Its gene expression in female should be twice of that in male, if gene expression level was proportional to gene dosage. When gene dosage was compensated, expression level was equalized in the X chromosome. In mammals, including human, either of X chromosome in female cell was totally inactivated.

chromosome in male (XY) showed elevated expression. In sex chromosomes in higher plants, presence or absence of gene dosage effect was unclear. For revealing any regulation at gene expression level by gene dosage in papaya, we compared transcriptome of the genes on the X chromosome between female and hermaphrodite plants.

2-2. Material and Methods

Plant materials

Plant material used for whole genome sequencing and RNA-seq were derived from Hawaiian papaya cultivar 'Sunrise Solo'. Mature leaves of papaya were harvested from greenhouse-grown plants.

Extraction of genomic DNA and RNA

Genomic DNA was extracted from leaves using DNeasy Plant Mini Kit (Qiagen). Total RNA was also extracted from mature leaves of papaya plants grown in the greenhouse, using RNeasy Plant Mini Kit (Qiagen).

Whole genome sequencing and *de novo* assembly

From genomic DNA of female and hermaphrodite plants, paired-end sequencing libraries were prepared using TruSeq DNA Sample Prep Kits (Illumina). Briefly, 1 μg genomic DNA was fragmented by sonication and approximately 300-500 bp fragments were prepared. Ends of fragmented DNA was repaired by T4 DNA polymerase and T4 DNA kinase, followed by addition of single adenine nucleotide to the 5' ends. Specific adapter DNA for Illumina sequencing was ligated to both ends of prepared DNA fragments. Adapter-ligated DNA was PCR amplified using primer sets for adapter sequences. Amplified DNA was applied to sequencing by an Illumina Genome Analyzer llx (110 bp paired-end read) as template libraries. Sequencing analysis was carried out in Iwate Biotechnology Research Center. Sequencing reads of fastq format files from female and hermaphrodite genomic DNA were applied to *de novo* assembly by CLC Assembly Cell (CLC Bio), and

assembled sequences were obtained as scaffold sequences.

Construction of reference sequences of X and Y^h chromosome

For construction of the "reference sequences" of the sex determination region, available BAC clone sequences (Ming et al. 2008, Na et al. 2012) corresponding to the X and Y^h chromosome were retrieved from a database (Table 2.1) and assembled using ATGC software (GENETYX) respectively. In the first step, BAC clone sequences annotated as "unordered" were separated into shorter fragments by deleting long ambiguous sequences (N_{100}). Subsequently, these fragmented BAC clone sequences were re-assembled with the scaffolds of *de novo* assembled hermaphrodite genomic DNA, as described above, together with the "complete" or "ordered" BAC clone sequences for the Y^h chromosome. These reference sequences were ordered and joined to form a pseudomolecule, according to the BAC order reported previously using fingerprinting technique (Na et al. 2012, Yu et al. 2009).

Gene prediction on the reference sequence

Gene prediction on the reference sequence was conducted by geneid web server (http://genome.crg.es/geneid.html). Additionally, putative transposons in the Y^h-reference sequences were predicted through TransposonPSI program. Both annotation data was integrated into CLC genomics workbench program (CLC Bio) by importing the GFF3 output format. Overlap of the two predictions was compared using track tools, and predicted genes not overlapping with transposons (in other words, predicted genes that are not transposons) were used for further study. The step of reference sequence construction and annotation are described in Figure 2.2.

Table 2.1 Accession list of BAC clones used for assembly

Sex chromosome	Accession No
Y^h chromosome	AC239156, AC238630, FJ429372, AC238635, AC238627, AC238772, AC238615, AC239165, AC238596, AC238767, AC238602, AC238616, AC238624, AC238619, AC238620, AC238589, AC238618, AC239142, AC238761, EF625821, AC239146, AC238633, AC239169, AC239135, AC238621, AC238611, FJ429373, AC239158, AC238769, AC238762, AC238766, AC238636, AC239198, AM778096, AC238764, AC238770, EF625817, AC238765, AC238759, EF625819, AC238768, AC238600, AC238594, AC238601, EF661025, AC238607, AC238593, AC238771, AC239153, JN088492, AC239163, AC238625, AC238608, AC239157, EF661024, AC239154, AC239203, AC239168, AC239167, EF625820, AC238760, AC239141, AC238763
X chromosome	AC238609, AC239251, AC239149, AC238599, FJ429368, AC238592, AC239145, AC238626, EF661023, AC238595, AC239143, AC239204, AC238628, AC238603, AC239162, AC239253, AC239202, AC239161, AC239170, AC238597, AC238632, AC239150, AC239147, AC238631, AC239152, AC239166, AC238764, AC239136, AC238637, AC239144, FJ429367, AC239160, AC238590, AC239151

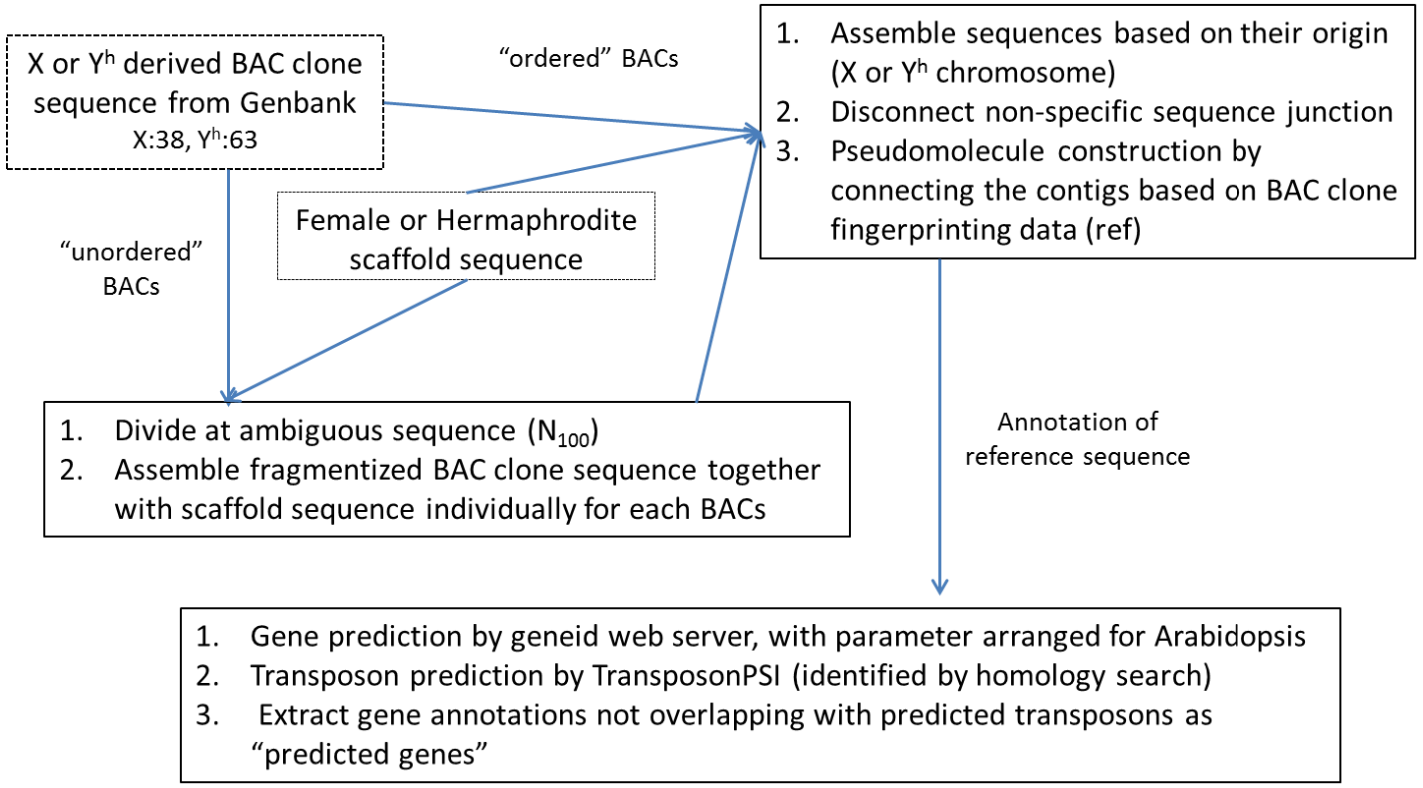

Figure 2.2 Flow of the construction of X and Y^h chromosome reference sequences

Available BAC clone sequences of papaya sex chromosomes were obtained from GenBank, and divided into two groups, "ordered" or "unordered", based on their annotation. Firstly, "unordered" BAC clone sequences were divided at ambiguous sequences ('N'), representing gaps. Then, separated BAC clone sequences were assembled together with the scaffold sequences. Obtained contigs were subsequently assembled with the "ordered" BAC clone sequences. Assembled contigs were ordered and pseudomolecules, "X reference" and "Y^h reference" sequences, were formed. Genes and transposons on these reference sequences were predicted by geneid web server, and TransposonPSI, respectively. Finally, genes except for putative transposons were defined as "Predicted genes".

Comparative analysis between X and Y^h chromosome by predicted genes

Coding DNA sequences (CDS) were compared between predicted gene on the X and Y^h chromosomes, using TBLASTX reciprocally. According to the BLAST search results, putative homologous or allelic genes between X and Y^h chromosomes were defined. Location of homologous or allelic genes in the X and Y^h chromosome was visualized by plotting.

RNA-seq analysis

Ribosomal RNA was removed from total RNA using Ribo-Zero Magnetic Kit (epicenter). Libraries for Illumina sequencing was prepared using NEBNext mRNA Library Prep Reagent Set for Illumina (New England Biolabs), following the manufactures instructions. Briefly, mRNA was fragmented and double-stranded cDNA was synthesized after reverse transcription using random hexamer. End of double stranded cDNA fragments was once dA-tailed and specific adapter DNA was ligated. Size-selected adapter ligated DNA was amplified using primers for adapter sequences. Amplified DNA from the prepared library was applied to 75 bp paired-end sequencing using Illumina Miseq (Illumina). Sequencing reads from each sample were *de novo* assembled using Trinity (Grabherr et al. 2011) program with Trimmomatic option for trimming and discarding low quality reads. Sequence reads were mapped to the assembled cDNA contigs for quantifying expression level by read count using CLC Genomics workbench 6.5 as RNA-seq analysis. RPKM (read per kilobase per million) was calculated as expression value of each gene (contig).

Identification of sex chromosomal genes

Genes located on the sex chromosome were identified by mapping the assembled RNA-seq contigs to reference sequence for X and Y^h chromosomes using Splign (Kapustin et al. 2008) in the NCBI genomics workbench. Additionally, contig sequences were applied to BLAST searching against scaffold sequences of whole genome of female for finding paralogous genes and defining unique genes on sex chromosomes. Based on these BLAST searching results, autosomal genes, X chromosome specific (X1Y0) genes and genes on the both sex chromosome (X1Y1) were classified (Figure 2.3).

Evaluation of gene expression ratio between female and hermaphrodite

Based on RPKM of each gene, expression ratio was calculated in each category of genes. Logarithm of expression ratio (female/hermaphrodite) was calculated. Statistic difference of distribution of gene expression ratio was evaluated by Kolmogorov-Smirnov test, which was written in R script.

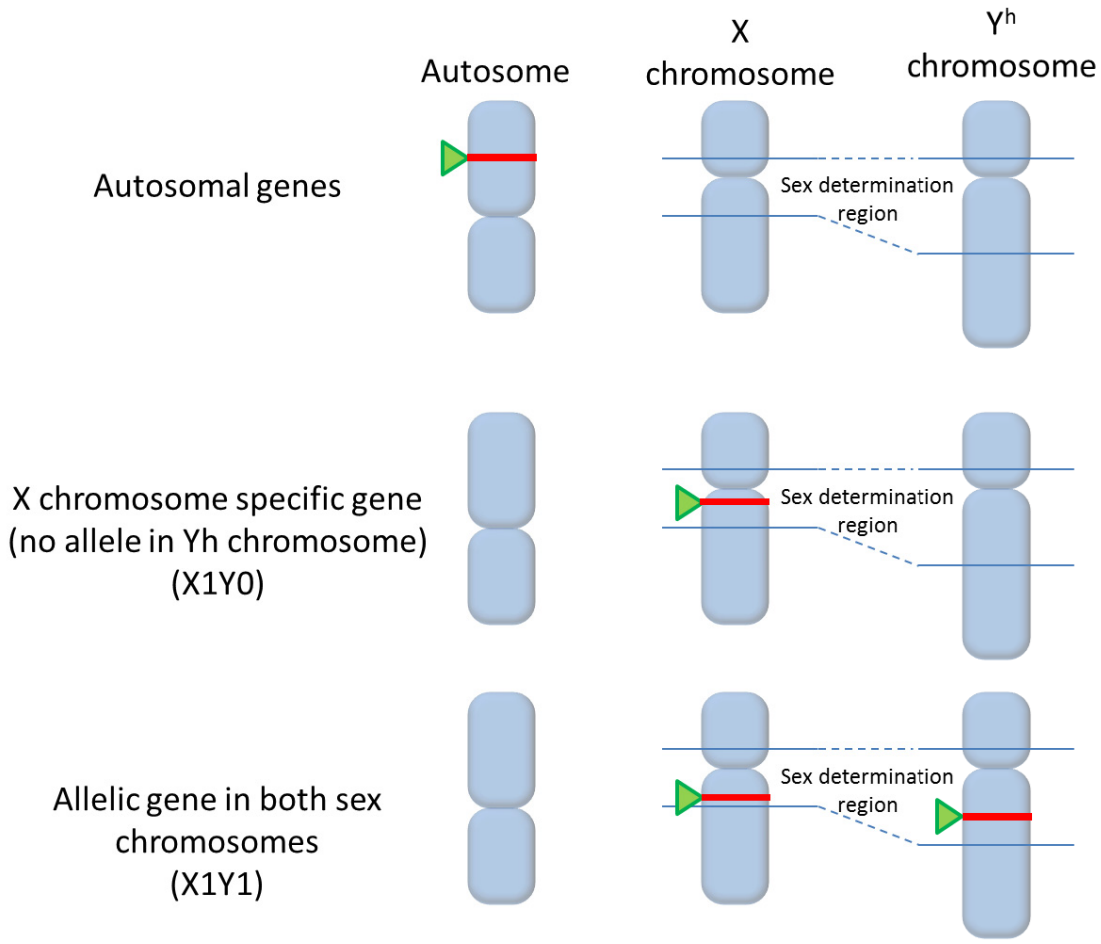

Figure 2.3 Scheme of the four categories of genes in autosome and sex chromosomes

Genes that are located only on the autosome were defined as autosomal genes. Unique genes in the X chromosome were designated as 'X1Y0'. Genes showing alleles in both X and Y^h chromosomes were represented as 'X1Y1'.

2-3. Result

For constructing reference sequences of sex chromosomes, newly analyzed whole genome sequences of female and hermaphrodite were merged with known BAC clone sequences for X and Y^h chromosomes.

Genome sequencing of the X and Y^h chromosomes

Firstly, the whole genome of female and hermaphrodite plants was sequenced using next generation sequencing technology. Genomic DNA was extracted from female and hermaphrodite plants 'Sunrise Solo', and fragmented by sonication. Fragmented DNA was then end repaired followed by dA-tailing, and adapter ligation. Constructed libraries were sequenced by Illumina Genome Analyzer llx (Table 2.2). Sequencing analysis was carried out in Iwate Biotechnology Research Center. In this sequencing analysis, sequences of the libraries were read from both ends, as paired-end sequencing. Consequently, 188,872,150 and 188,442,304 reads of 110 bp fragments were obtained from females and hermaphrodites, representing 20.5 and 20.6 Gbp in total (Table 2.2). According to the predicted genome size of papaya (422.5 Mbp), their approximate coverages were 48.5 fold in female and 48.8 fold in hermaphrodite. These short read sequences were subjected to *de novo* assembly for constructing scaffolds using CLC Assembly Cell, resulting in 88,615 scaffolds (275.5 Mbp) for females and 89,275 scaffolds (282.7 Mbp) for hermaphrodites. Their N50 value, which was similar to median of scaffold length for evaluating efficiency of the assembly, were 22,988 bp and 19,960 bp in female and hermaphrodite, respectively (Table 2.2). These assembled genomic DNA sequences were designated as "Female scaffolds" and "Hermaphrodite scaffolds" in

Table 2.2 Summary of sequencing analysis and *de novo* assembly of female and hermaphrodite genomic DNA

Sample name	Read Summary			*De novo* assembly		
	Total length (bp)	Number of reads	Read depth (x)	Number of contigs	Total length (bp)	N50 (bp)
Hermaphrodite	20,587,064,350	188,872,150	46.5	89,275	282,701,114	19,960
Female	20,540,211,136	188,442,304	46.4	88,615	275,491,175	22,988

this study.

Construction of sex chromosome sequences as reference sequences

By BLAST searching Female scaffolds against 38 BAC clone sequence for the X chromosome, corresponding scaffolds to the X chromosome were extracted. These sequences were assembled with 38 BAC clone sequences using the ATGC program to construct reference sequences for the X chromosome. In total, 12 contigs (designated as XSR_Ref1 to XSR_Ref12. Table 2.3), ranging from 75,834 bp to 699,029 bp, were obtained and their total lengths reached 4.3 Mbp, representing 80% of the predicted sex determination region in the X chromosome (Gschwend et al. 2013). Similarly, reference sequences for the Y^h chromosome were constructed by assembling the selected scaffolds for the Y^h chromosome and 63 Y^h originated BAC clone sequences. Contigs ranging from 64,215 bp to 2,543,606 bp, were obtained (designated as HSY_Ref1 to HSY_Ref12, Table 2.3), and their total lengths reached 7.7 Mbp, which covered approximate 95% of the predicted Y^h chromosome (Gschwend et al. 2013).

Gene prediction in the reference sequences of the X and Y^h chromosome

Genes in the X and Y^h reference sequences were predicted by *ab initio* procedure using geneid program in web server. Consequently, 867 and 1,559 genes were predicted in the reference sequences of X and Y^h chromosomes, respectively. According to the previous study (Yu et al. 2007), it was known that transposable elements were abundant in the papaya sex chromosomes. Then, transposons were also surveyed in these sex chromosome sequences using TransposonPSI program.

Table 2.3 Detail of the constructed X- and Y^h-reference sequences

Chromosomal location	contig ID	length (bp)	No. of constituent BAC clones	No. of constituent scaffolds
X chromosome	XSR_Ref1	699,029	6	2
	XSR_Ref2	626,277	8	3
	XSR_Ref3	523,661	4	1
	XSR_Ref4	382,457	3	2
	XSR_Ref5	376,585	3	2
	XSR_Ref6	345,528	4	3
	XSR_Ref7	318,340	2	2
	XSR_Ref8	269,713	2	3
	XSR_Ref9	268,773	3	1
	XSR_Ref10	228,632	1	2
	XSR_Ref11	176,819	1	1
	XSR_Ref12	75,834	1	1
	Total	4,291,648	38	23
Y^h chromosome	HSY_Ref1	2,543,606	23	2
	HSY_Ref2	2,028,808	18	0
	HSY_Ref3	736,124	5	2
	HSY_Ref4	637,403	5	2
	HSY_Ref5	557,735	5	2
	HSY_Ref6	336,922	2	1
	HSY_Ref7	245,996	2	1
	HSY_Ref8	209,725	2	1
	HSY_Ref9	155,172	1	1
	HSY_Ref10	110,712	1	1
	HSY_Ref11	86,650	1	1
	HSY_Ref12	64,215	1	1
	Total	7,713,068	66	15

This analysis showed presence of 1,014 and 2,283 putative transposable elements in the X and Y^h chromosomes, respectively. Combining gene and transposon prediction results, 553 and 929 predicted genes were found to encode non-transposable elements in the X and Y^h chromosomes, respectively.

Comparative analysis of predicted genes between the X and Y^h chromosomes

According to the hypothesis of sex chromosome evolution, it was suggested that X and Y^h (Y) chromosomes were derived from the same parental autosome. Therefore, alleles of some loci could be present in both X and Y^h chromosomes. By sequence comparison of the predicted genes using TBLASTX analysis, 120 gene pairs showed sequence similarities between the X and Y^h chromosomes. Relative position of these homologous genes in the X and Y^h chromosomes were compared by scatter plotting (Figure 2.4). In this comparison, two regions showed quite similar gene orders between two sex chromosomes (S1, S2), harboring 20 and 35 genes respectively. Although a few weakly correlated regions of gene location were found (P1-P3), order of other genes were not always conserved between X and Y^h chromosomes possibly due to rearrangements, like duplication, conversion or insertion.

Evaluation of dosage effect of genes on the X chromosome

As demonstrated in the above study, allelic genes were limited between the X and Y^h chromosomes, indicating that unique genes on the X chromosomes were assumed to be abundant. Therefore, dosage of most genes on the papaya X chromosome could be different between female (XX) and hermaphrodite (XY^h), as

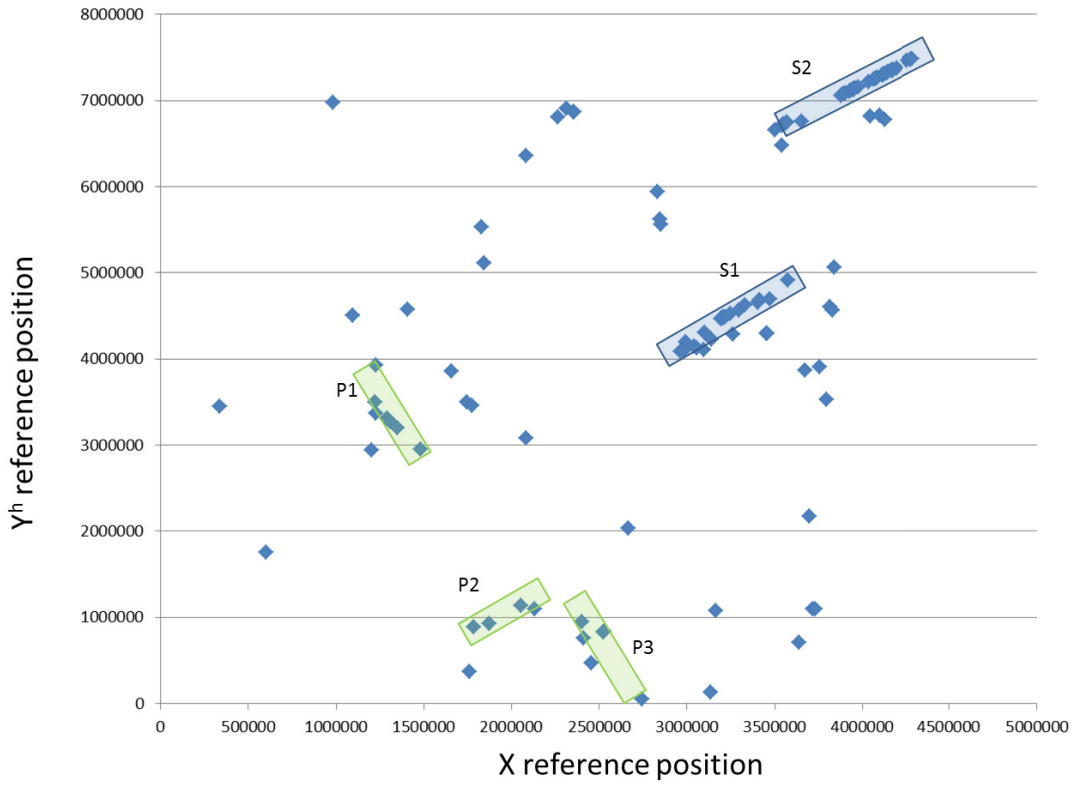

Figure 2.4 Chromosomal locations of alleles in the X and Yʰ chromosomes

Positions of putative allelic genes were plotted in the X and Yʰ chromosome. Vertical axis indicated the Yʰ chromosome, and horizontal axis represented the X chromosome. Putatively strongly conserved regions of gene order between two chromosomes were highlighted in blue (S1, S2), and regions showing weak correlation is highlighted in green (P1-P3).

seen in mammalian X chromosome. Then, expression of genes on the X chromosome was compared between female and hermaphrodite plants, for revealing gene dosage effect. For this purpose, transcriptome of mature leaves in female and hermaphrodite plants were analyzed using RNA-seq method.

Total RNAs were extracted from mature leaves of female and hermaphrodite plants (cv. Sunrise solo), carrying identical genetic background except for the sex chromosomes. After eliminating ribosomal RNA, mRNA was once fragmented and double-stranded cDNA was synthesized, followed by adapter ligation to the cDNA fragments. Prepared adapter-ligated fragments (libraries) were sequenced by Illumina Miseq sequencer. Obtained sequence reads were once assembled using Trinity (Grabher et al. 2011) software for constructing contigs of cDNA. It resulted in 48,195 contigs with an N50 of 1,128 bp (Table 2.4). Each contig was assumed to be equivalent to gene. By BLAST searching these gene (contig) sequences against papaya genome sequences, it was shown that 173 and 46 genes were located uniquely on the X chromosome and Y^h chromosome, respectively. Another 151 genes were mapped on both sex chromosomes. The rest of 47,722 contigs were present in the autosome, By mapping the obtained short reads in each RNA sample to these genes and counting mapped reads, expression level for each gene was represented by RPKM (reads per kb per million) as normalized gene expression value. In this analysis, genes showing only single read mapped were not involved in further analysis, resulting that expression data of 32,104 genes were employed.

Based on these transcriptome data, expression ratio of individual genes was calculated between female and hermaphrodite leaf samples (female/hermaphrodite), and converted to log2 value. The expression ratios of four categories of genes, all the

Table 2.4 Summary of sequencing analysis and *de novo* assembly of female and hermaphrodite RNA

Sample name	Read Summary		*De novo* assembled contigs		
	Total length (bp)	Number of reads	Number of contigs	Total length (bp)	N50 (bp)
female	784,263,423	10,528,988	48,195	33.634.135	1,128
hermaphrodite	827,051,795	11,106,268			

expressed genes (all), autosomal genes (autosomal), X chromosome specific genes (X1Y0) and genes on both sex chromosomes (X1Y1), are shown as histograms in Figure 2.5. Number of compared expressed genes was 26,945 for all, 26,685 for autosomal, 99 for X1Y0 and 82 for X1Y1 (Table 2.5). Significant difference of distribution of gene expression ratio was evaluated between 'autosomal' and 'X1Y0' or 'X1Y1' by Kolmogorov-Smirnov test. As results, there were no significant differences in distribution of gene expression ratio between autosomal genes and X chromosome specific genes (p-value=0.3447) or genes located on the X and Y^h chromosome (p-value=0.4642). Since median of expression ratio was usually employed as a criterion of dosage effect in transcriptome data (Julien et al. 2012), the values were calculated in these four categories, which were shown to be -0.0218 to 0.1909 (Table 2.6). When gene expression was correlated with gene dosage, its expression level carrying two alleles should be twice of that carrying only a single allele. Median and histogram distribution in X1Y0 (X chromosome specific genes) showed that these genes were most likely to express equally both in female and hermaphrodite. Nonetheless, 20 unique genes on the X chromosome were abundantly expressed in female leaves with more than two-fold. Of these genes, two genes were highly expressed in female at significant level ($P<0.05$). In summary, gene dosage could be compensated in many genes on the X chromosome between female and hermaphrodite, but several genes were more abundantly expressed in female than hermaphrodite.

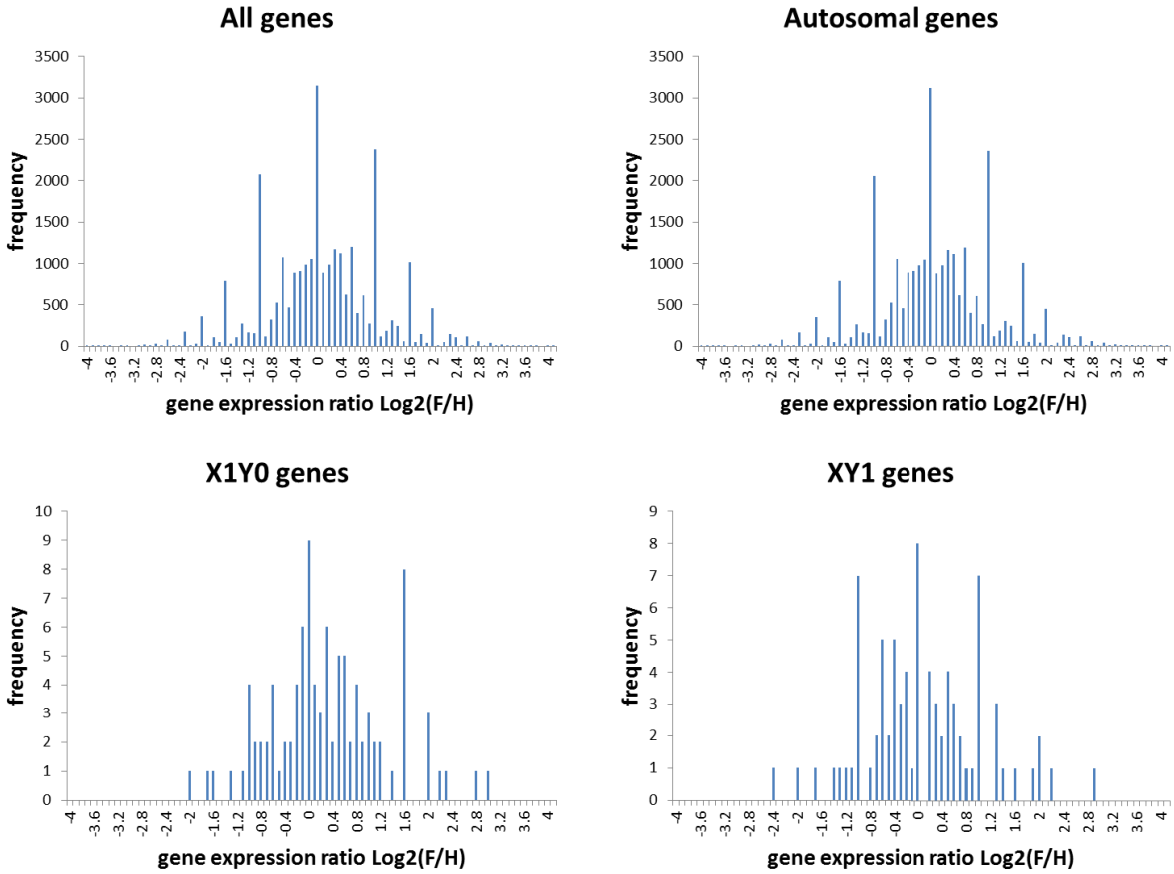

Figure 2.5 Distribution of gene expression ratio between female and hermaphrodite

Based on RNA-seq data, histogram of gene expression ratio between female and hermaphrodite leaves (log2 value) illustrated in four categories of genes. Horizontal axis represents the expression ratio, and vertical axis represent gene number.

Table 2.5 Number of mapped expressed genes (assembled contigs of RNA-seq) on the X, Y^h chromosomes, and autosomes

Classification based on gene location	Number of compared genes for evaluating expression ratio
All	26,945
Autosomes	26,685
X1Y0	99
X1Y1	82

Table 2.6 Median and P-value of Kolmogorov-Smirnov test (ks test) of expression ratio of genes locating on X, Y^h chromosomes, and autosomes between female and hermaphrodite leaves

Classification of gene location	expression ratio median*	P-value of ks test *
Autosomes	-0.0218	-
X1Y0	0.1909	0.3447
X1Y1	-0.0222	0.4642

* Calculation of these analyses was described in the method section.

2-4. Discussion

By constructing reference sequences of X and Y^h chromosomes, more contiguous (longer) genome sequences were obtained than previously analyzed BAC clone sequences. It proved that combination of data from NGS and traditional BAC clone sequencing were quite useful for constructing complicated genome sequences. As shown in the present study, these reference sequences of sex chromosomes allowed to defining gene models and structural differences. These structural analysis data is greatly helpful for predicting evolution of papaya sex chromosomes and elucidating sex determination mechanisms via identifying genes for sex determination in papaya. It is hypothesized that genes on the Y or Y^h chromosome were degenerated, while the X chromosome maintained most of active genes from the ancestral chromosome (Ming et al. 2011). Nonetheless, according to present gene prediction in the X and Y^h reference sequences, more genes were found in the Y^h chromosome than the X chromosome. On the other hand, less number of expressed genes was uniquely mapped on the Y^h chromosome (46 genes) though RNA-seq analysis than that uniquely mapped on the X chromosome (153), indicating that non-active genes (pseudogenes) might be accumulated in the Y^h chromosome as hypothesized. In previous reports by Na et al. (2012) and Wang et al. (2012), by ordering the bacterial artificial chromosome (BAC) clone sequences of the papaya X and Y^h chromosomes, physical maps of 5.4 Mb and 8.1 Mb have been constructed. Those sizes were similar to the present results of assembly (4.3 Mb in X and 7.7 Mb in Y^h chromosome). By mapping papaya EST sequences to their constructed maps, 98 and 96 annotated transcription units were identified in the X and Y^h chromosome, respectively, and of these transcription units, 70 units were shared by both sex chromosomes,

Discrepancy between number of these genes (transcription units) and that in the present study was probably owing to difference of employed transcript data (EST and RNA-seq data).

In the previous study, it was suggested that length of the Y^h chromosome is approximately 1.8 times larger than that of the X chromosome, possibly due to several large rearrangements (Na et al. 2012). In the present study, occurrences of several genomic rearrangements were presumed during the development of the Y^h chromosome, according to comparison of chromosomal position of putative allelic genes between the X and Y^h chromosome. Deduced from chromosomal location of these genes, a large DNA was suggested to be inserted between the S1 and S2 region in the Y^h chromosome (Figure 2.6). Other rearrangement events were speculated that two relatively large inversion have occurred before insertions in two regions in the Y^h chromosome as shown in the predicted scheme of Y^h chromosome (Figure 2.6). These several large genomic rearrangement events are supposed to expand recombination suppression region between X and Y^h chromosome. Total insertion length in the Y^h chromosome, compared to the X chromosome, was estimated as approximately 3.5 Mb, where 293 predicted genes were located. Only 63 expressed genes were obtained by RNA-seq were mapped to this region. These suggest that genes in the Y^h chromosome were degenerated even in its stretched genome sequence.

As one of characteristics of sex chromosome, gene dosage effect of the X chromosome was verified by estimating expression ratio from RNA-seq data between female and hermaphrodite leaves. Generally, phenotypic differences were not observed in vegetative phase of papaya plants among sex types. Therefore,

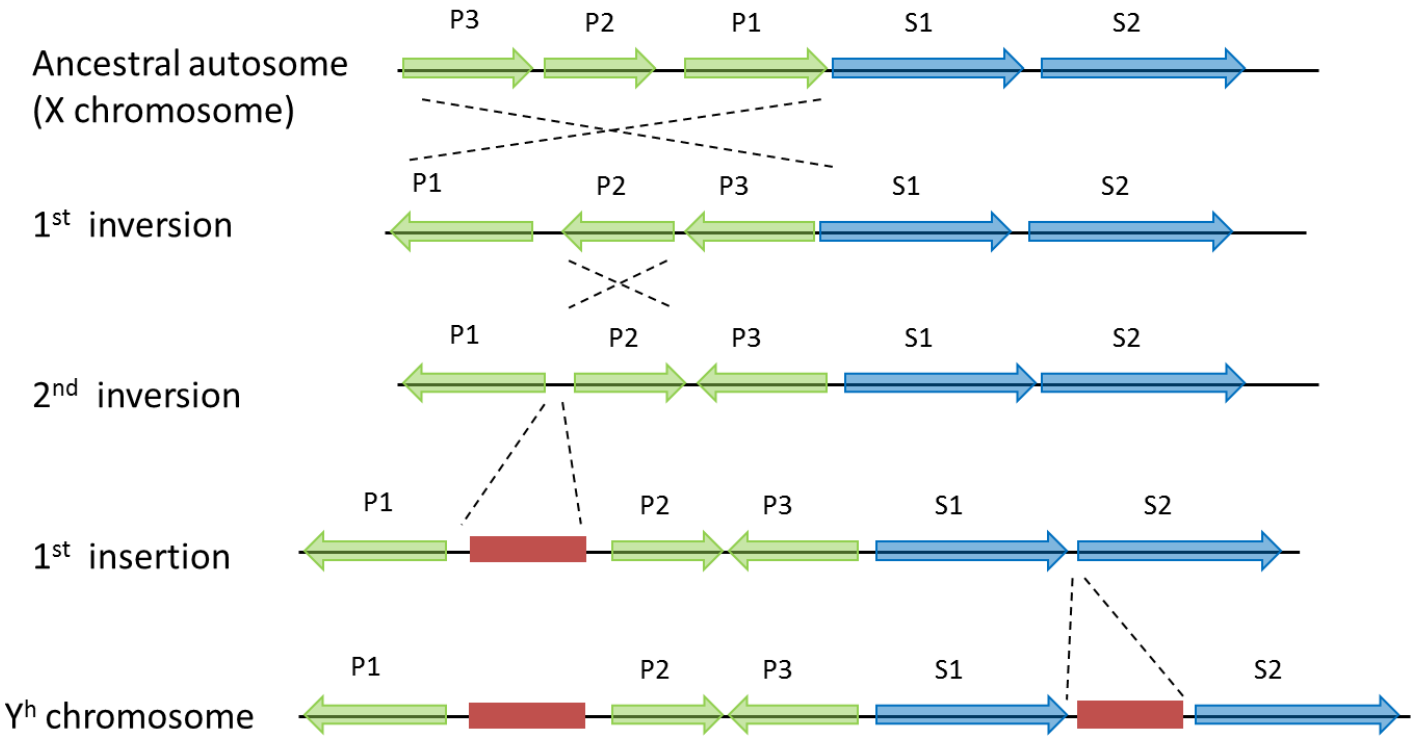

Figure 2.6 Predicted process of genomic rearrangements in the Y^h chromosome

According to comparison of order of putative alleles between the X and Y^h chromosome, genomic rearrangements during the Y^h chromosome development were predicted. Blue (S1, S2) and green (P1, P2, P3) arrows were described in Figure 2.3. Inserted regions were represented by brown colored bars.

transcript profiles in leaves were comparable between female and hermaphrodite, and appropriate for revealing gene dosage effect on the X chromosome. Actually, by comparing gene expression between female and hermaphrodite, X chromosome specific genes mostly showed similar expression level and distribution of their expression ratio is similar to that of autosomal genes. These observations suggested that any kind of gene dosage compensation was present in the genes on papaya X chromosome. In mammalian female, whole region of either X chromosome was inactivated by heterochromatinization, and transcription activity was totally repressed. In papaya X chromosome, because significantly differentially expressed genes were also found between female and hermaphrodite leaves, coordinated gene dosage compensation system, as shown in mammals or *Drosophila*, is not present. These observations implied that transcription in these genes on the X chromosome was controlled at individual gene level rather than chromosomal level. Genes for sex determination will be involved in unique genes in either X or Y^h chromosomes or allelic genes carrying any polymorphisms between two sex chromosomes. At least, hundreds of genes could be candidates for sex determination genes, according the present results. Therefore, further studies, including more gene expression analysis, are necessary for finding actual genes for female and hermaphrodite determination.

Chapter 3 Comparative study between Y and Y^h chromosome for identifying for male-hermaphrodite determination

3-1 Introduction

Of the three sex chromosomes in papaya (X, Y^h and Y), genome sequence information of the Y chromosome was limited. This is due that male plant (XY) is not useful in papaya cultivation or breeding.

In the hypothesized evolution of the papaya sex chromosome, the Y^h chromosome has recently diverged from the Y chromosome and partial genome sequence analysis also suggested that Y and Y^h chromosome were separated around 74,000 years ago (Yu et al. 2008b). Therefore, sequence divergence between the Y^h and Y chromosomes was expected to be small and not as complex as the case of the X and Y^h chromosome, as observed in previous chapter. Then, it was expected that there was a higher chance to identify the genes responsible for male–hermaphrodite differentiation by simply comparing the genome sequences of Y^h and Y chromosomes. For this purpose, structural differences were surveyed between Y and Y^h chromosome by mapping sequence reads from male genomic DNA to the reference sequence of the Y^h chromosome, which was constructed in previous chapter. If any structural differences (polymorphisms) were identified in genes on these sex chromosomes, it is highly possible that these genes are concerned to male-hermaphrodite determination. In the present study, SuperSAGE data of flower samples as transcriptome data was incorporated in finding genes for sex determination and transposon-insertion were also investigated as structural difference between the Y and Y^h chromosome.

3-2 Materials and Methods

Plant materials and genomic DNA preparation

Plant materials used in this study were the Hawaiian papaya cultivar 'Sunrise Solo' for hermaphrodite plants and TM1 for male plants derived from a cross between cv. 'Wonder frea' and 'IG4', which is a locally grown line in Okinawa. Genomic DNA was extracted from mature leaves of greenhouse-grown papaya plants using DNeasy Plant Mini Kit (Qiagen).

Prediction of putative protein coding sequences in the Y^h reference sequences

Putative gene regions in the Y^h reference sequence were predicted on the basis of sequence similarity to known cDNA. Papaya unigenes obtained by the ASGPB Papaya genome project (http://asgpb.mhpcc.hawaii.edu/papaya/) were mapped to the Y^h reference sequences using Splign program (Kapustin et al. 2008) in the NCBI genomics workbench (http://www.ncbi.nlm.nih.gov/tools/gbench/), with minimal exon identity set to 0.9. Additionally, the Y^h reference sequence were subjected to TBLASTX searches (with an e-value cutoff score of 0.1) against the cDNA sequences of *A. thaliana* and *Oryza sativa* as well as EST sequences of *Vitis vinifera* and *Populus trichocarpa*. From results of the BLAST search, gene models were created using Blast92gff3.pl. Additionally, program based gene prediction from chapter 2 was employed again with additional analysis with parameters provided for rice and wheat. These new annotations were employed together with the gene and transposon prediction in the previous chapter.

When different gene prediction procedures (program-based and sequence similarity-based) resulted in overlapping of multiple predicted genes, they were

manually integrated into a single predicted gene. ORFs in the genes were investigated, and domain structure of each ORF was surveyed by analyzing their deduced amino acid sequences with Pfam domain search in CLC Genomics Workbench software.

Identification of expressed genes in the flower in the Y^h reference sequence

For defining expressed genes in flowers, 26-bp tag sequences obtained by Ht-SuperSAGE analysis used in chapter 1 were mapped to the Y^h reference sequence using BLASTN. In this analysis, only tags with more than 10 counts were retrieved to exclude false tags caused by PCR or sequencing errors, and any mismatches in the mapping process by BLAST searching were not allowed. Finally, all of the analyzed data were converted to the gff3 format and imported into the CLC Genomics Workbench software.

Whole genome sequencing of male plant

From the genomic DNA of male plant, paired-end sequencing libraries were prepared using TruSeq DNA Sample Prep Kits, as described in chapter 2. Briefly, genomic DNA was fragmented by sonication and adapters were ligated to their ends after end-repairing and dA-tailing. Amplified DNA from the prepared library was applied to an Illumina Genome Analyzer llx for sequencing (110 bp paired-end read). Sequencing reads were used for *de novo* assembly by CLC Assembly Cell and "male scaffold" sequences were obtained.

Mapping of short sequence reads to the Y^h reference sequence

Short sequence reads from male and hermaphrodite genomic DNA were mapped to the Y^h reference sequences with CLC Genomics Workbench software. The mapping procedure was conducted in the default setting except for insertion cost and deletion cost which was set to 3, and mismatch cost set to 2. From these read mapping data, DNA sequence variation was studied for the Y^h reference sequences; Y^h in Sunrise Solo, and Y chromosome, using "probabilistic variant detection" tool in CLC genomics workbench. Only the sequence polymorphisms were specifically addressed between the Y and Y^h chromosome and those between the reference and Sunrise Solo Y^h were regarded as irrelevant to the differentiation between the Y and Y^h chromosome. Genome structure of large unmapped regions in the Y^h-reference sequences was determined by mVISTA (Frazer et al. 2004).

Expression analysis of selected genes

Expression of candidate genes for sex determination was analyzed by RT-PCR as follows. Total RNA was extracted from young flower buds (no longer than 7 mm) of each sex type of papaya plant as described in, chapter 1. RNA was reverse transcribed using Primescript reverse transcriptase (TAKARA Bio) with oligo dT primer, and the synthesized first-strand cDNA was used as the template for PCR.

For full-length cDNA isolation, GeneRacer Kit (Invitrogen) was employed following the manufacturer's instruction (Matsumura et al. 2006). Briefly, the cap sequence of mRNA was replaced by a 5′ -end adapter sequence, followed by reverse transcription using an adapter-oligo-dT primer. Each end of cDNA was amplified once using a combination of a gene specific primer and adapter primer (first PCR), and the specific cDNA fragment was amplified using nested primers for the gene

and adapter sequences (second PCR).

Alignment and phylogenic analysis of the *SVP-like* gene

Amino acid sequence of, 11 SHORT VEGETATIVE PHASE protein from 8 plant species and 9 closely related genes were employed to clarify the physiological relationship. By using the sequence of SQUA family gene from *Coffea Arabica* as an out group, A. A. sequences were aligned by clustalW (Larkin et al. 2007), and phylogenic tree was constructed by MEGA6 program (Tamura et al. 2013) based on Maximum Likelihood method.

3-3. Result

Additional gene prediction on the Y^h reference sequence

Since it is predicted that Y^h chromosome was derived from Y chromosome, it is expected that structural change of any genes has occurred in the Y^h chromosome. Therefore, genes for male-hermaphrodite determination might be inactivated by any mutation in the Y^h chromosome. Then, in addition to predicted genes by program, putative protein-encoding regions, even pseudogenes, were also explored in the Y^h chromosome by sequence similarity with known genes. In the Y^h reference sequence, homologous region with papaya unigenes or cDNA/EST of higher plants were searched using Splign program and TBLASTX respectively. Additionally geneid program was employed for gene prediction again, with parameter set for rice, and for wheat. These data was imported to the CLC genomics workbench software together with the predicted gene data in the Y^h chromosome in the chapter 2. After elimination of transposable elements, which was predicted in the chapter 2, and integration of gene models from different gene prediction data, 545 genes, except for putative transposable elements, were identified in the Y^h reference sequences (Table 3.1). Among these, 124 genes were predicted from their similarity to the papaya unigenes and 130 genes from their similarity to known cDNA or EST in higher plants. The remaining 291 genes were predicted by geneid program. Since genes for sex determination were expected to express in the flower, SuperSAGE tags in flower samples, as described in the chapter 1, were mapped to the Y^h_reference sequence by BLAST program. Since SuperSAGE tags were extracted from the 3′ -end of cDNAs, predicted genes located within 1.5 kb upstream of the tag sequence were defined as the corresponding gene to the tag in this study. These

Table 3.1 Predicted genes in the Y^h-reference sequences and sequence variation between the Y^h and Y chromosome

			sequence polymorphisms				
contig ID	predicted genes	expressed genes in flowers*	nucleotide substitution	in/del	gaps**	polymorphic genes***	uniquely expressed genes in males and/or hermaphrodites*
HSY_Ref1	177	60	5,055	349	197	51	14
HSY_Ref2	125	32	5,065	345	25	23	2
HSY_Ref3	49	11	1,918	102	4	4	0
HSY_Ref4	52	22	570	60	2	14	1
HSY_Ref5	36	10	776	30	3	5	0
HSY_Ref6	32	8	580	43	0	5	0
HSY_Ref7	22	10	21	3	15	3	0
HSY_Ref8	15	13	66	5	0	5	0
HSY_Ref9	8	1	249	8	47	1	0
HSY_Ref10	10	3	144	15	0	3	0
HSY_Ref11	10	7	0	0	1	1	0
HSY_Ref12	9	6	84	5	0	5	0
Total	545	183	14,528	965	294	120	17

* Expressed genes were identified by mapping Ht-SuperSAGE tags to the predicted genes in the Y^h reference sequences.

** No corresponding sequence reads from male genomic DNA was found.

*** Predicted genes carrying sequence polymorphisms between males and hermaphrodites.

Ht-SuperSAGE tags in papaya flowers were mapped on 183 predicted genes. Of these predicted genes, 45 were presumed to be uniquely expressed from either the Y and/or Y^h chromosomes based on tag-expression patterns among the sex types. Although most of the 45 genes had expression in both male and hermaphrodite flowers, two genes showed hermaphrodite-specific and another two genes showed male-specific expression (Table 3.2).

Comparison of genome sequences between Y^h and Y chromosome

For identifying structural differences between Y^h and Y chromosome, hermaphrodite and male genome sequences were compared by sequence read mapping against the Y^h reference sequence.

Male genomic DNA was sequenced using Illumina GAIIx sequencer in the same manner as hermaphrodite genome in chapter 2 (Table 3.3). Short sequence reads from male genomic DNA by paired-end sequencing were mapped against the Y^h_reference sequence with CLC genomic workbench software, followed by an analysis using probabilistic variant detection method, revealing 17,936 nucleotide substitutions (single or multiple nucleotide substitutions) and 1,131 small in/dels (<7 bp) between the Y and Y^h chromosome. Simultaneously, short reads from the hermaphrodite (cv. Sunrise Solo) genomic DNA was similarly mapped to the Y^h reference sequences. The results revealed 5,782 nucleotide substitutions and 304 short in/dels. Since the Y^h reference sequences mostly consisted of BAC clone sequences of cv. SunUp, the identified sequence differences represent polymorphisms between the two hermaphrodite cultivars (SunUp and Sunrise Solo). These variations are not related to Y^h/Y differentiation, so they were not included

Table 3.2 Genes in the Y^h or Y chromosomes, expressed only in hermaphrodite or male flowers.

Tag ID*	position in the Y^h reference sequences	tag count**						predicted gene function
		male		female		hermaphrodite		
		P1	P2	P3	P4	P5	P6	
Cp12204	HSY_Ref1 580,079..580,104	0	0	0	0	22	42	monodehydroascorbate reductase
Cp13833	HSY_Ref1 2,456,326..2,456,351	0	0	0	0	42	10	No significant homologies found（CPMSY10 in Table 3.5）
Cp28552	HSY_Ref3 466,759..466,784	0	15	0	0	0	0	Late embryogenesis abundant protein
	HSY_Ref3 496,756..496,781							
Cp32574	HSY_Ref 1,591,473..1,591,498	9	2	0	0	0	0	No significant homologies found

* Previously analyzed Ht-SuperSAGE tags in Urasaki et al. 2012

** Tag counts from early stage (P1,P3, P5) or late stage (P2, P4, P6) of flowers in each sex type

Table 3.3 Summary of sequencing analysis and *de novo* assembly of male and hermaphrodite genomic DNA in papaya

sample	read summary*			*de novo* assembly result		
	total length (bp)	number of reads	depth (x)	number of scaffolds	total length of scaffolds (bp)	scaffolds (contigs) N50 (bp)
male	19,426,551,160	178,225,240	43.9	133,166	285,683,349	11,189 (8,950)

* Paired-end sequencing data of 100 bp read

for consideration to identify genes specific to Y^h and Y sequences. Thus, 14,528 nucleotide substitutions and 965 short in/dels were suggested to be polymorphisms specific between the Y^h and Y (Table 3.1), showing 1.8 SNPs and 0.13 in/dels per 1 kb on average. Of these sequence polymorphisms, 934 SNPs and 78 in/dels were located in the sequences of 164 predicted genes. Focused on genes expressed in the flower buds (from chapter 1), sequence variations between the Y^h and Y were found in the 120 predicted genes (Table 3.1).

Apart from these sequence polymorphisms, 294 regions with no male genomic DNA reads mapped (gaps) were found (Table 3.1). These gaps were mainly concentrated in the three regions in the Y^h chromosome. Since these regions were supposed to be missing in the Y chromosome, their genome structures were further investigated. Since reads might not be mapped to the Y^h reference sequences because of the threshold for short read mapping, *de novo* assembled scaffolds of male genomic DNA were surveyed, which were homologous to these large unmapped regions. Several scaffolds showed sequence similarities with two large unmapped regions in the Y^h-reference sequences and were presumed to be homologous regions in the Y chromosome (Figure 3.1, Table 3.4). In the unmapped region in HSY_Ref9, scaffolds of the male genome were observed throughout the entire region, indicating that this region is highly divergent between the Y^h and Y. Two large unmapped regions were found in HSY_Ref1, where corresponding male genome scaffolds were identified between positions 2,130,001 and 2,190,000, but a gap and partial inversion were observed in scaffold15430. Between positions 2,420,000 and 2,543,606 in HSY_Ref1, no corresponding male scaffolds were found.

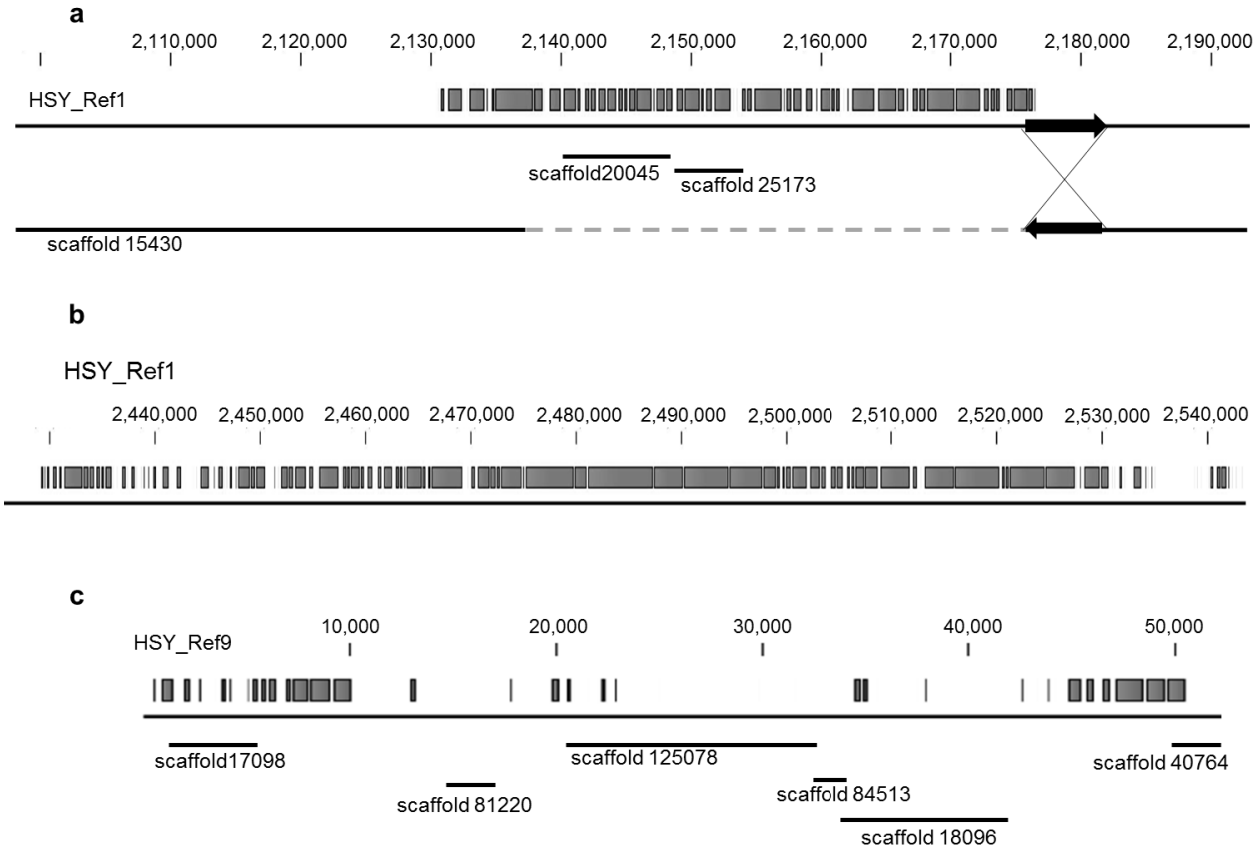

Figure 3.1 Highly-diverged or gap regions between the Yʰ and Y chromosome sequences

Three large genomic regions in the Yʰ chromosome, where short sequence reads of male genomic DNA were unmapped, are illustrated: (a) 2,130,001–2,190,000 bases in HSY_Ref1; (b) 2,420,000–2,543,606 bases in HSY_Ref1; and (c) 1–60,000 bases in HSY_Ref9. Detailed positions of unmapped regions are indicated in Table 3.4.

Lines indicate the Yʰ reference contigs or scaffolds of male genomic DNA, and numbers are positions in the Yʰ reference contigs. Gray rectangles represent unmapped regions in the Yʰ chromosome by short reads from male genomic DNA.

Arrows in (a) points to the sequence inversion between the Yʰ and Y chromosome.

Table 3.4 Sequences in the male genome that correspond to the highly diverged region in the Y^h reference sequences.

Contigs and regions of variable regions in the Y^h chromosome	scaffold ID of male genome	corresponding regions in the Y^h chromosome to male genome sequences
HSY_Ref1 2,130,001..2,190,000	15,430	2,130,001..2,130,536
		2,179,095-2,181,800 (inverted)
		2183229-2189516
	20,045	2,153,226..2,159,358
	25,173	2,158,181..2,162,743
HSY_Ref1 2,420,000..2,543,606	No significantly similar sequences are found.	
HSY_Ref9 1..60,000	17,098	1,342..5,336
	81,220	14,714..17,673
	125,078	20,795..32,393
	84,513	32,410..34,123
	18,096	34,032..42,673
	40,764	50,501..53,268

Candidate genes for male–hermaphrodite determination

As described above, 45 uniquely expressed genes in male and/or hermaphrodite flowers were selected. Among them, only 14 genes carried sequence polymorphisms between the Y^h and Y, and their expression patterns were confirmed in flowers of three sex types by RT-PCR (Figure 3.2). According to the deduced amino acid sequences of these 14 genes, all the identified polymorphisms were synonymous changes or located outside of ORFs (Table 3.5). We also surveyed the structural differences of the genes between the Y^h and Y by transposon-insertion. According to gene and transposon predictions, putative transposons were inserted in the sequences of 41 predicted genes (Table 3.6). Among them, copia-like retrotransposon insertion was specifically observed in a gene in the Y^h (CPMSYT4) encoding a MADS-box protein similar to the SHORT VEGETATIVE PHASE (SVP) protein in *A. thaliana* (Table 3.5, Figure 3.3), which was also identified as a unique gene in the Y^h chromosome and absent in the X chromosome in the study of the chapter 1. PCR amplification of the 5′ -end of the *SVP-like* gene from hermaphrodite genomic DNA demonstrated that the length of the first intron was longer than that of the male allele because of transposon insertion (Figure 3.3a). RT-PCR and RACE showed that the cDNA of the *SVP-like* gene was also different between the male and hermaphrodite (Figure 3.3b, c, d, Figure 3.4). The 5′ - end of cDNA in the Y allele was replaced by a partial sequence of the second intron of the Y^h allele, indicating an alteration of the transcription initiation site. The deduced amino acid sequences from cDNAs showed that the Y allele codes for the intact SVP-like protein with both MADS-box and K-box domains, whereas Y^h allele codes for the truncated version of the protein with only part of the K-box domain (Figure 3.3d, Figure 3.5).

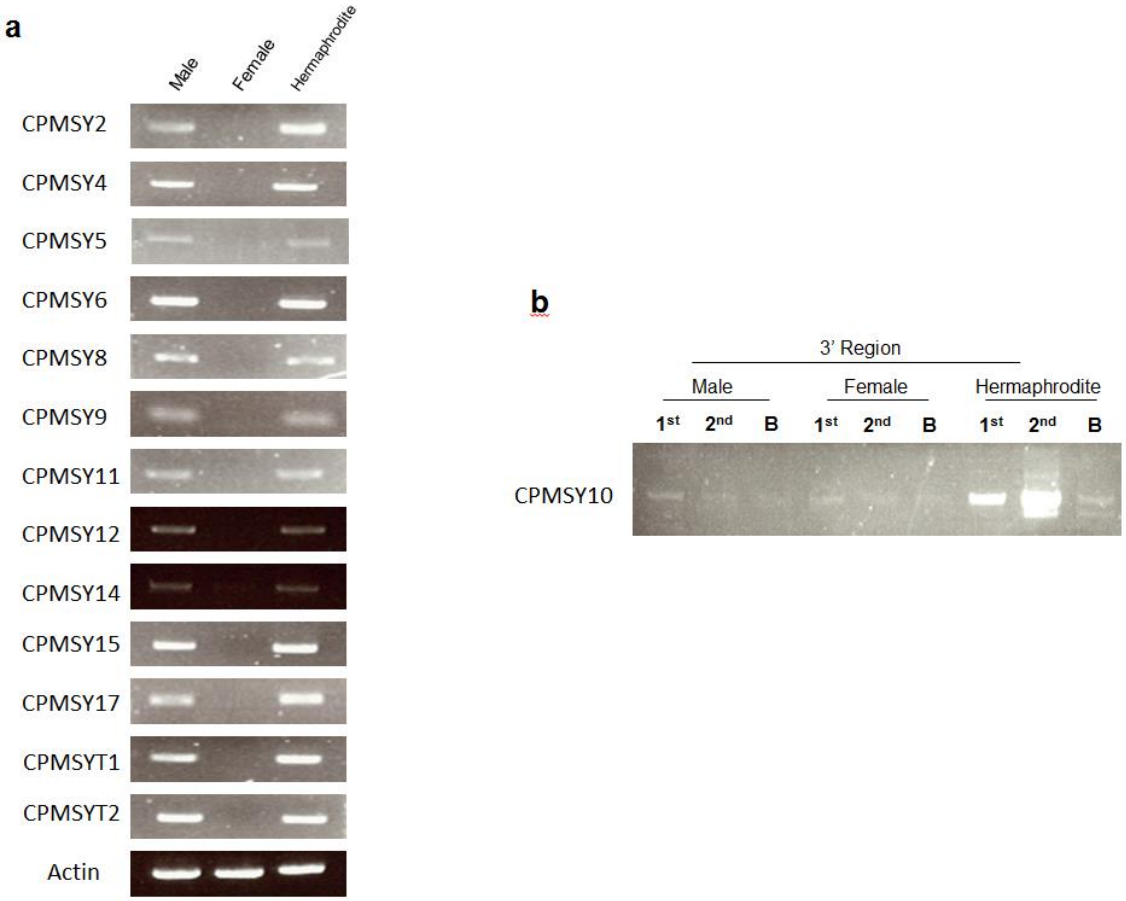

Figure 3.2 Expression analysis of selected genes in flower buds of three sex types

Single-stranded cDNA from male, female, and hermaphrodite flowers were used as templates for amplifying the genes shown in Table 3.5 (Analysis for CPMSYT4 are integrated in Figure 3.3). Fragments were amplified using gene-specific primers (a), except for CPMSY10, which were analyzed by 3′-RACE (b). In panel b, product of the first and second PCR of RACE analysis are applied to the "1st" and "2nd" lanes, respectively, and an amplified product without gene-specific primers was loaded.

Table 3.5 List of uniquely expressed genes in male and hermaphrodite flowers carrying polymorphisms

gene ID	contig ID	position in the contig	amino acid change (+/-)	deduced protein*
CPMSY2	HSY_Ref1	1,836,675..1,845,388	-	No significant homologies found
CPMSY4	HSY_Ref1	1,008,712..1,009,588	-	No significant homologies found
CPMSY5	HSY_Ref1	1,253,004..1,265,132	-	chromatin assembly factor-1
CPMSY6	HSY_Ref1	1,590,522..1,591,410	-	No significant homologies found
CPMSY8	HSY_Ref1	191,263..195,925	-	No significant homologies found
CPMSY9	HSY_Ref1	2,206,830..2,218,436	-	No significant homologies found
CPMSY10	HSY_Ref1	2,455,644..2,463,186	-	No significant homologies found
CPMSY11	HSY_Ref1	823,834..832,922	-	member of ARF GTPase family
CPMSY12	HSY_Ref1	916,301..923,130	-	storekeeper transcriptional regulator
CPMSY14	HSY_Ref2	449,560..460,165	-	protein kinase
CPMSY15	HSY_Ref2	48,093..51,564	-	rpp21 subunit-like protein
CPMSY17	HSY_Ref4	190,356..200,416	-	Vps51/Vps67 family protein isoform 1
CPMSYT1	HSY_Ref1	1,094,723..1,120,377	-	hiv tat-specific factor 1 homolog
CPMSYT2	HSY_Ref1	1,592,916..1,637,199	-	2-phosphoglycolate phosphatase 1
CPMSYT4	HSY_Ref5	133,349..168,762	+	Short vegetative phase protein

* Protein showing significant similarity with each predicted gene by BLASTX searching against non-redundant protein database in GenBank

Table 3.6 List of predicted genes with a putative transposon in its gene region.

Gene ID	Contig ID	position	notes
CPHSYT1	HSY_Ref1	164,834..497,149	
CPHSYT2	HSY_Ref1	580,186..597,812	
CPHSYT3	HSY_Ref1	623,796..633,058	
CPHSYT4	HSY_Ref1	728,591..740,996	
CPHSYT5	HSY_Ref1	948,613..972,849	
CPHSYT6	HSY_Ref1	1,094,723..1,120,377	CPMSYT1 in Table 3.5
CPHSYT7	HSY_Ref1	1,592,916..1,637,199	CPMSYT2 in Table 3.5
CPHSYT8	HSY_Ref1	1,627,125..1,636,889	
CPHSYT9	HSY_Ref1	1,864,870..1,884,893	
CPHSYT10	HSY_Ref1	1,991,629..1,992,070	
CPHSYT11	HSY_Ref1	2,202,695..2,253,965	
CPHSYT12	HSY_Ref1	2,253,180..2,271,838	
CPHSYT13	HSY_Ref1	2,337,053..2,356,305	
CPHSYT14	HSY_Ref1	2,418,151..2,426,751	
CPHSYT15	HSY_Ref10	32,268..37,747	
CPHSYT16	HSY_Ref2	9,161..25,632	
CPHSYT17	HSY_Ref2	215,336..224,825	
CPHSYT18	HSY_Ref2	858,470..931,848	
CPHSYT19	HSY_Ref2	1,140,367..1,195,842	
CPHSYT20	HSY_Ref2	1,160,521..1,344,093	
CPHSYT21	HSY_Ref2	1,575,843..1,576,385	
CPHSYT22	HSY_Ref3	483,983..486,431	
CPHSYT23	HSY_Ref3	526,450..534,657	
CPHSYT24	HSY_Ref3	606,099..624,485	
CPHSYT25	HSY_Ref4	81,219..116,496	
CPHSYT26	HSY_Ref4	403,943..410,412	
CPHSYT27	HSY_Ref4	461,427..475,251	
CPHSYT28	HSY_Ref4	500,407..509,649	
CPHSYT29	HSY_Ref4	531,209..541,402	
CPHSYT30	HSY_Ref4	571,919..586,426	
CPHSYT31	HSY_Ref4	597,415..603,229	
CPHSYT32	HSY_Ref4	603,023..609,828	
CPHSYT33	HSY_Ref5	133,349..168,762	CPMSYT4 in Table 3.5
CPHSYT34	HSY_Ref5	307,107..379,230	

CPHSYT35	HSY_Ref5	443,319..508,395
CPHSYT36	HSY_Ref6	49,259..58,336
CPHSYT37	HSY_Ref6	260,134..274,631
CPHSYT38	HSY_Ref7	176,700..186,160
CPHSYT39	HSY_Ref8	41,232..50,242
CPHSYT40	HSY_Ref8	97,941..106,835
CPHSYT41	HSY_Ref9	84,037..118,993

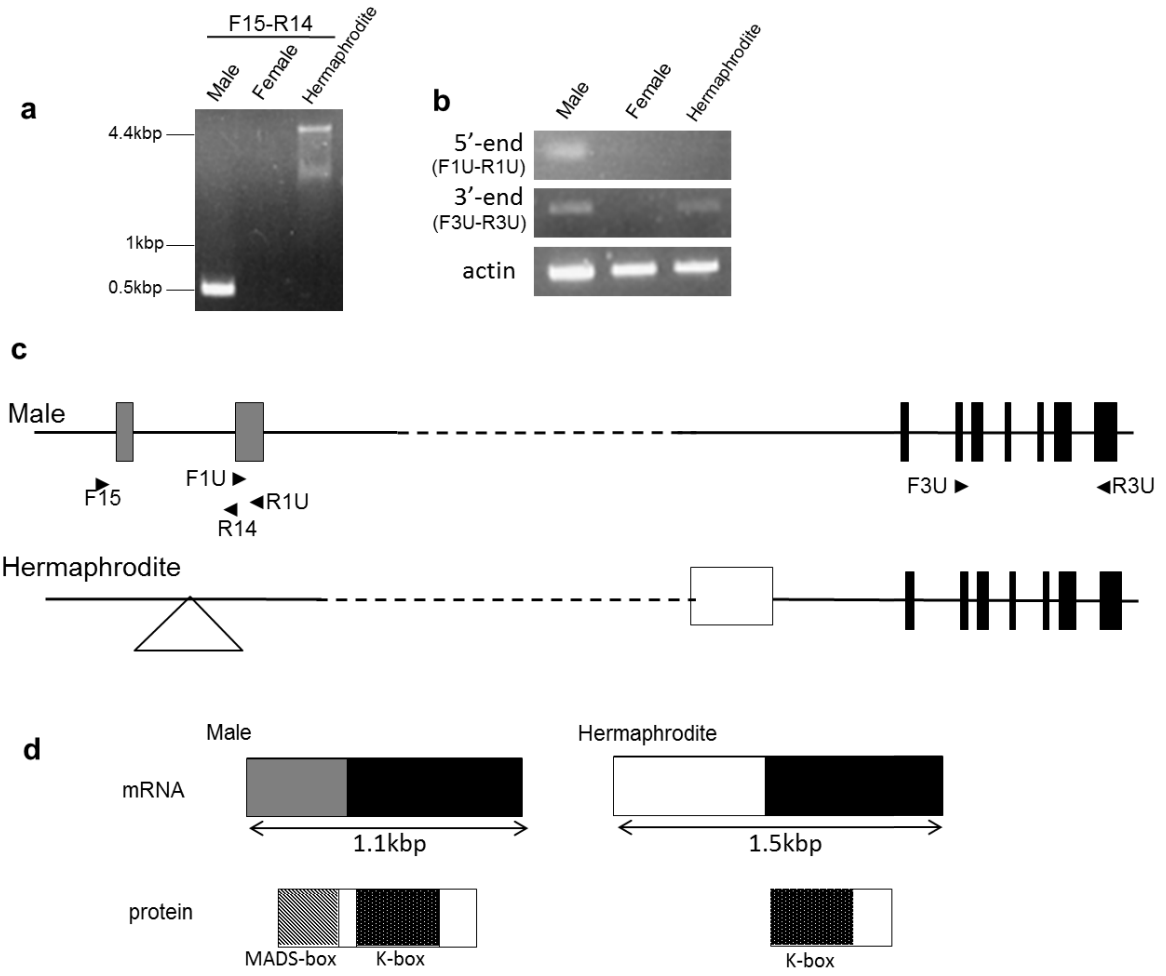

Figure 3.3 Structure of *SVP-like* genes and their products in males and hermaphrodites

Structural differences in the *SVP-like* gene (a) or cDNA ends (b) were demonstrated by PCR. The location of each PCR primer is indicated in (c). Each lane indicates the amplified product from genomic DNA or cDNA of flowers for males, females, or hermaphrodites. The structures of the *SVP-like* gene of males and hermaphrodites were predicted by 5' and 3' RACE is illustrated in (c). Black rectangles are commonly observed exons in both males and hermaphrodites, and gray or white rectangles indicate unique exons in males or hermaphrodites, respectively. A triangle in the hermaphrodite gene shows the insertion of the copia-like transposon. A schematic diagram of mRNAs and deduced proteins of the *SVP-like* gene in males and hermaphrodites is shown in (d). Black, gray, and white sections in mRNAs correspond to the exons in the panel (c). In proteins, two domains (MADS-box and K-box) are indicated.

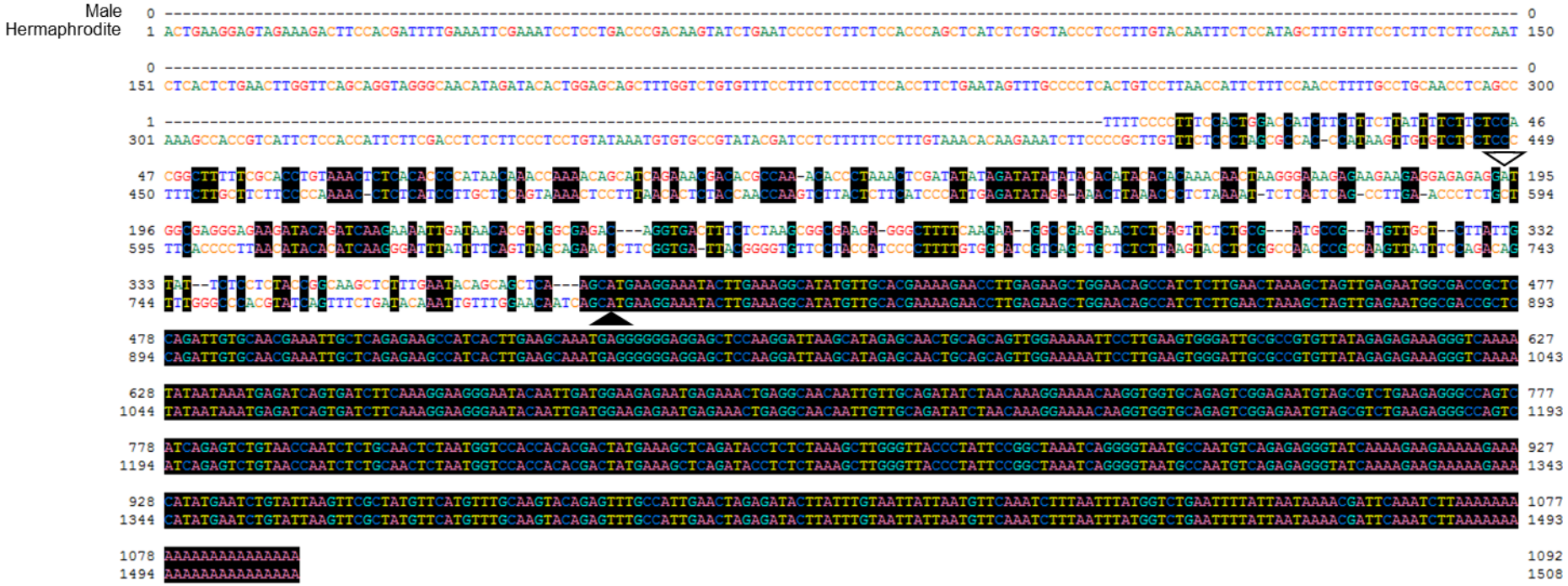

Figure 3.4 cDNA sequences of the *SVP-like* gene

Amplified cDNA sequences were aligned between male (upper) and hermaphrodite (lower) alleles. Identical sequences are shown in black boxes, and predicted start codons are indicated by an open triangle in the male allele and a closed triangle in the hermaphrodite.

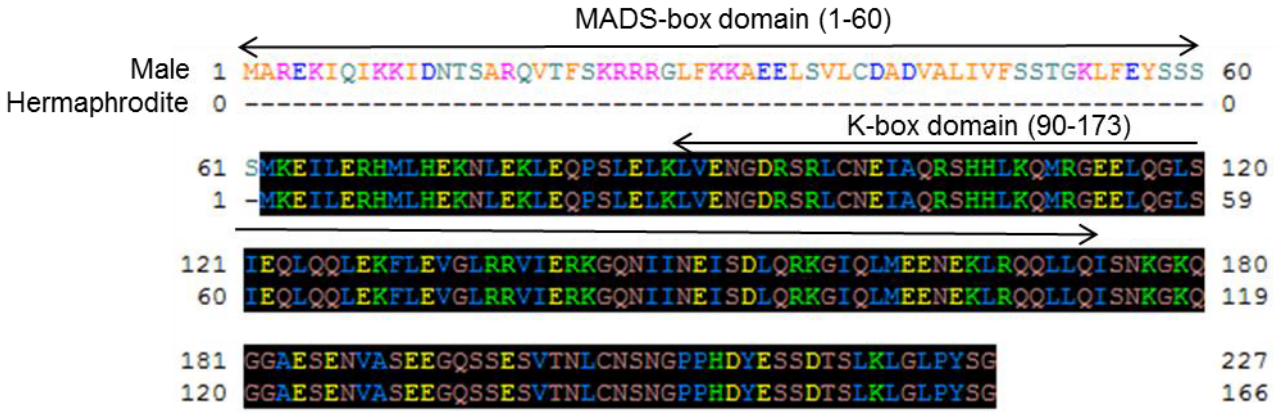

Figure 3.5 Deduced amino acid sequences of the SVP-like protein

Deduced amino acid sequences of the SVP-like protein from males and hermaphrodites are compared. Identical sequences are shown in black boxes, and the arrows indicate the MADS-box or K-box domain.

Amino acid sequence comparison of the *SVP-like* proteins

To identify any paralogous genes of the *SVP-like* gene in papaya genome, its full-length cDNA sequence was applied to BLAST searching against the papaya draft genome sequence downloaded from ASGPB Papaya genome project, and *de novo* scaffold sequence from female (chapter 2). A putative paralogue was found in the autosome, and its alignment with cDNA sequence of *SVP-like* gene in the Y chromosome showed ORF structure, encoding an intact MADS-box and K-box domains, showing 93 % identity based on predicted CDS (Figure 3.6a, b). RT-PCR analysis revealed that this gene was expressed in the flowers of all sex types (Figure 3.6c), although the expression was relatively low compared to the *SVP-like* gene in the Y chromosome. (Figure 3.6d)

SVP homologues in dicot, including two papaya *SVP-like* genes, were compared for their structural classification. Deduced amino acid sequences of 11 *SVP* genes and 9 closely related genes were aligned by clustalW and phylogenic tree was drawn based on the Maximum Likelihood method (Figure 3.7). According to this phylogenetic tree, most of the *SVP* genes were grouped to the same clade. Both paralogues of *SVP-like* genes in papaya belonged to the clade including the *A. thaliana SVP* gene, indicating their structural similarity.

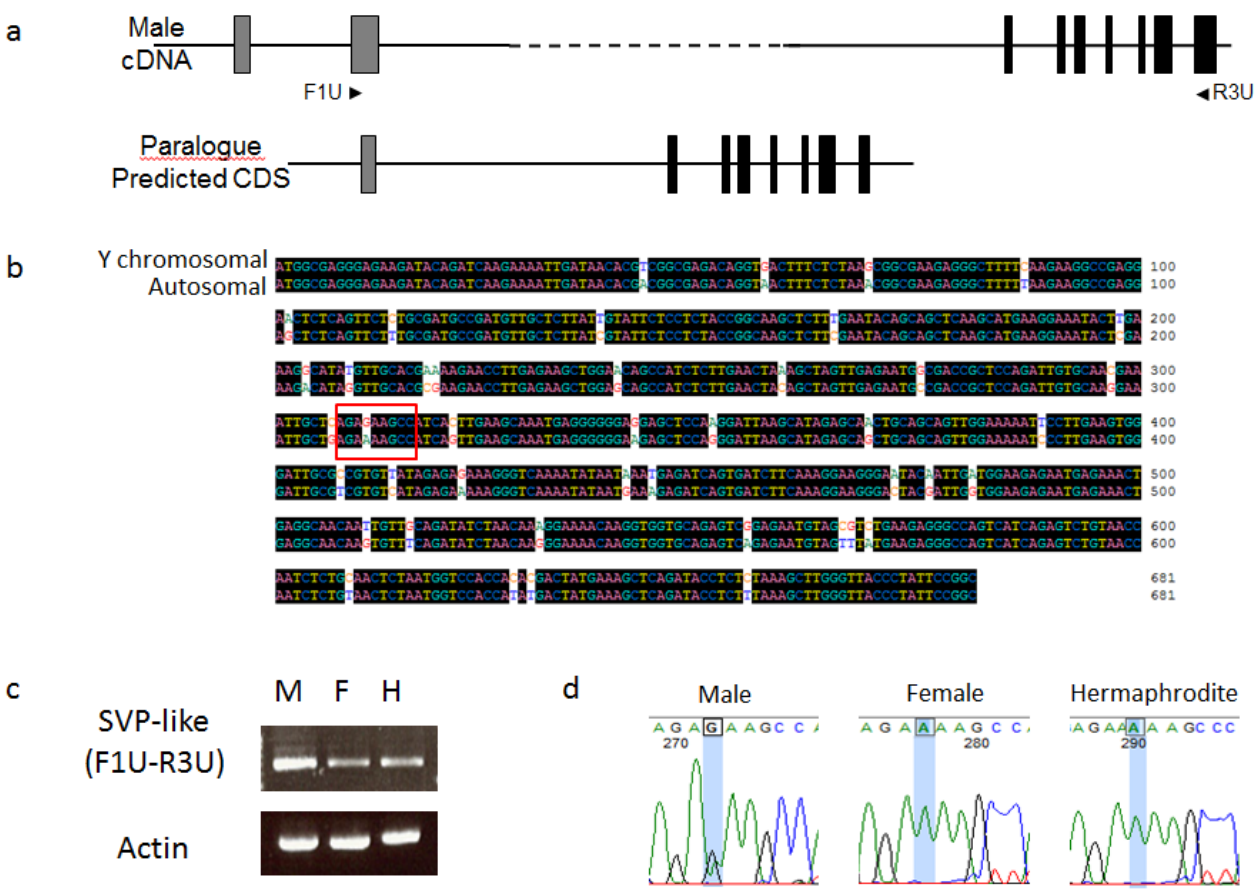

Figure 3.6 Structure and expression of autosomal paralogue of *SVP-like* gene in papaya

The predicted structures of the *SVP-like* gene in the Y chromosome and autosome were compared (a). Squares represented predicted exons. Alignment of CDS sequence of the Y chromosomal and autosomal *SVP-like* gene (b). Identical amino acids are high lightened, polymorphic sequences for discriminating these two genes were surrounded by red line. Expression of the *SVP-like* genes (c), were analyzed by RT-PCR analysis of in male (M), female (F), hermaphrodite (H). Sequencing analysis of the RT-PCR products of *SVP-like* (d), transcript (cDNA) from each sex type can be discriminated by their SNP (A/G).

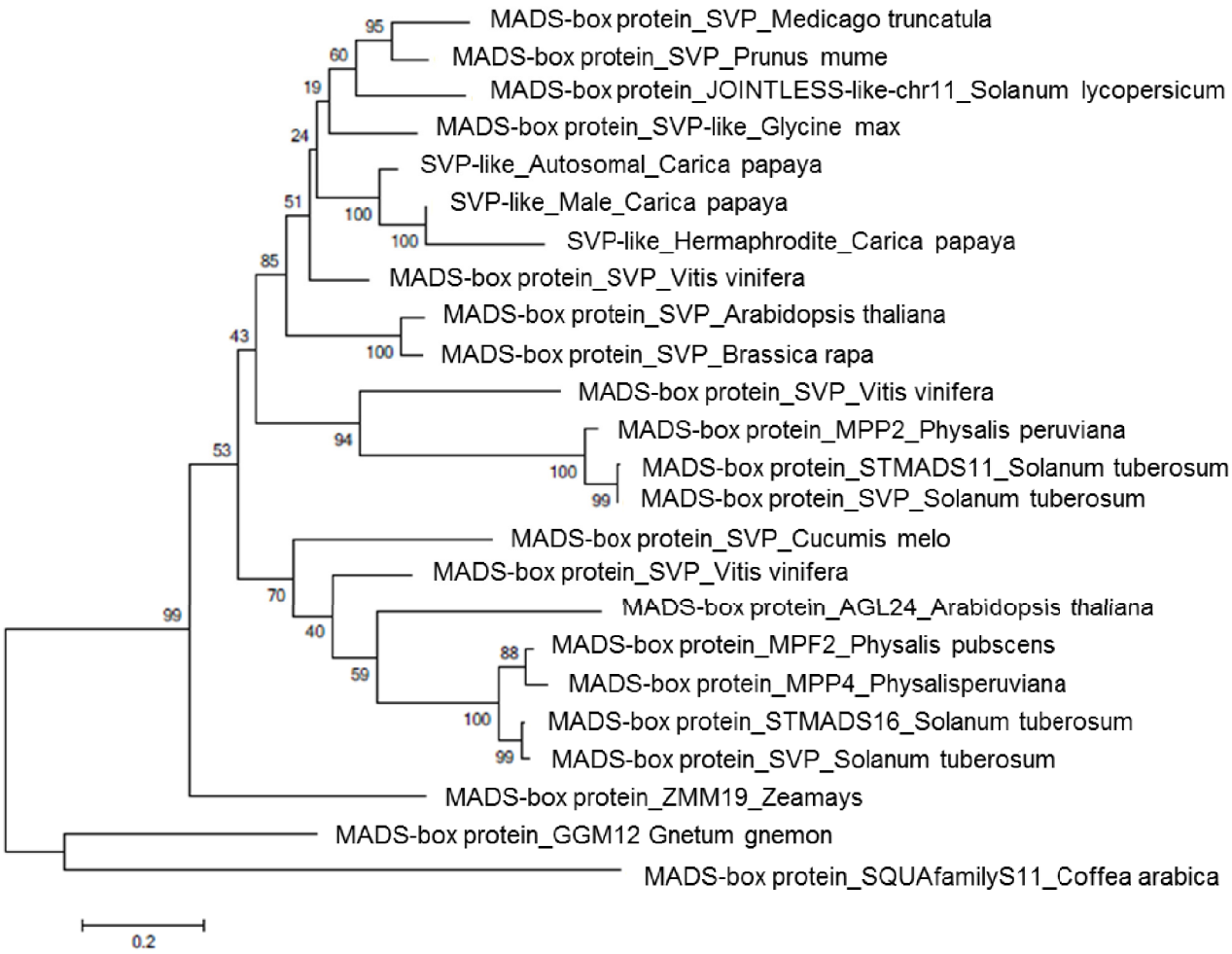

Figure 3.7 Phylogenic tree of *SVP* homologues in higher plants

According to amino acid sequences of 11 SVP gene products from 8 plant species and gene products from 10 SVP related genes (including SQUA family gene from *Coffea arabica* as an out group), SVP-like identified in this study were classified and phylogenetic was illustrated.

3-4. Discussion

As shown in the study of chapter 2, papaya genome sequences are highly diverged between X and Y^h chromosomes, including composition of genes. On the other hand, Y and Y^h chromosomes are predicted to be recently diverged (Yu et al. 2008a, b), and structural or functional differences of genes on these chromosomes are presumed to be limited. Therefore, the present study was focused on comparing sequences of the Y^h and Y chromosomes, to explore the genes responsible for male or hermaphrodite determination. Previous studies revealed high sequence similarity in an 88 kb sequence region between the BAC clone sequences in the Y^h and Y chromosome of papaya (Yu et al. 2008b). In those analyzed sequences, 236 SNPs and 36 small in/dels were observed, together with a large gap spanning 8.3 kb, indicating 2.7 SNPs and 0.4 small in/dels per 1 kb on average. The present analysis of 7.7 Mbp sequences of the Y^h and Y chromosome revealed that the frequency of sequence variation (1.8 SNP and 0.13 in/del per 1 kilobase) was not significantly different from that in the previous study. By mapping the short read of male genomic DNA to the Y^h reference sequences, large, highly diverged sequence regions were identified, which may be evidence of rearrangements.

It was expected that any of these nucleotide substitutions or in/dels between the Y^h and Y chromosome would contribute to male-hermaphrodite determination. However, within the uniquely expressed genes in male and hermaphrodite flower buds, no mutations rendering amino acid changes were found, even in the large, highly diverged regions. Together with sequence polymorphisms, it was presumed that transposon-insertion in the gene could be responsible for determining sex type. In the *SVP-like* MADS-box gene, it was found that a hermaphrodite-specific

transposon-insertion is responsible for the alteration of its cDNA and encoded protein. Interestingly, the same *SVP-like* gene was previously identified as a Y and Y^h chromosome-specific gene, because this allele is not present in the X chromosome (chapter 1). According to the hypothesis of sex chromosome evolution, genes in the Y chromosome were likely to be lost during diversification of the X and Y chromosomes. Then, this kind of Y (Y^h)-specific active gene was supposed to have any important role in sex determination as described in the chapter 1. Furthermore, the present study revealed that Y and Y^h chromosomes carried different alleles of the *SVP-like* gene, and only the allele in the Y chromosome encodes the full protein with both the MADS-box and K-box domains, which we presume to be functional. Since the early stage of male flower development shows a hermaphroditic phenotype, it is hypothesized that the *SVP-like* gene has a function as a suppressor of carpel development in papaya flowers. In *A. thaliana*, *SVP* was originally identified to be involved in controlling flowering time by transcriptional regulation of the *FT* gene. Further, it also plays roles in floral organ development via the regulation of B and C class genes in *A. thaliana* and other plant species (Gregis et al. 2006, Gregis et al. 2009, Méndez-Vigo et al. 2013, Jaudal et al. 2014, Wu et al. 2012). In dioecious poplars, *SVP-like* gene expression was higher in male flower buds than in female ones (Song et al. 2013), suggesting that it might have a role in sexual development. Therefore, the *SVP-like* gene in papaya is expected to be important in sex determination. By BLAST searching against whole papaya genome sequence, an autosomal paralogue of the *SVP-like* gene was also found, which retained an intact ORF and expressed in the flowers of all sex types. Since phylogenetic analysis showed that both *SVP-like* genes in the Y chromosome and autosome were classified

with *SVP* gene in *A. thaliana*, their molecular function might be similar. Further biochemical analyses, like evaluation of activities for DNA binding and transcriptional regulation, are necessary for elucidating physiological functions of papaya *SVP-like* genes.

Additionally, it was unclear how transposon insertion in the first intron caused an alternation of the transcription initiation site and the 5′ -end structure of mRNA in hermaphrodites (Figure 3.4). In *Cucumis melo*, gynoecy was caused by the repression of *CmWip1* expression, which was caused by retrotransposon-mediated DNA methylation (Martin et al. 2009). Thus, mutation by transposons not only disrupts ORFs, but transcriptional regulation might also be affected.

The present gene prediction in the Y^h reference sequences found that 183 genes were expressed in flowers, which are significantly more than the number of transcription units identified by Wang et al. (2012). This was possibly due to the depth of transcriptome analysis by Ht-SuperSAGE in the present study. Nonetheless, of the previously predicted transcription units reported by Wang et al. (2012), 34 units (genes) were not represented in gene list of the present study (data not shown). Among them, 12 transcription units were not found in the BAC clone sequences (SH60M19, SH71E16, and AM125I09). The remaining 22 units (genes) were individually evaluated for their expression patterns using previous transcriptome data, and for the presence of sequence polymorphisms between males and hermaphrodites. They did not show sex-dependent expression or any polymorphisms between males and hermaphrodites, so no candidate genes other than the *SVP-like* gene were identified that are probably involved in male–hermaphrodite differentiation.

Besides structural differences in genes, it is possible that sex determination is governed by differential gene expression between males and hermaphrodites. By mapping previously analyzed Ht-SuperSAGE tags to the Y^h reference sequences, tags for four genes appeared only in male or hermaphrodite flower buds. However, it has already been shown in our previous study that the allele of the *monodehydroascorbate reductase 4* gene, carrying an SNP in the tag sequence, is existent in the X chromosome and expressed in flowers of all sex types. For the other two male-specific expressed tags, putative alleles in the X chromosome or paralogs in the autosomes were found (data not shown); thus, specificity of their expression in males or hermaphrodites should be confirmed by further study. The expression of the gene corresponding to Cp13833, which is identical to CPMSY10 in Table 3.5, was observed only in the hermaphrodite flower by RT-PCR analysis (Figure 3.2). In these genes, several sequence polymorphisms were found in their upstream sequences (data not shown). To clarify whether they are responsible for sex determination, further analysis of their expression and transcriptional regulation should be carried out.

The present study focused on the uniquely expressed genes in male and/or hermaphrodite flowers and succeeded in finding a candidate gene (*SVP-like* gene) for sex determination. Although alleles on the X chromosome of male/hermaphrodite determination genes might be expressed in female flowers, it is still difficult to distinguish an allele on the X chromosome from paralogous genes on other chromosomes. More precise genome sequence information in papaya will be important for separately analyzing genes in the sex chromosomes and autosomes.

General discussion

In this study, any genes for sex determination in papaya were explored by comparing structural difference and expression analysis of genes in the sex chromosome. For identifying the gene for female/hermaphrodite determination, unique expressed genes in the Y^h chromosome was surveyed, and polymorphic genes between Y^h and Y chromosomes were investigated for finding the gene for male/hermaphrodite determination. In both approaches, the same gene, encoding MADS-box protein similar to SVP in *Arabidopsis thaliana*, was identified as a candidate of sex determination gene. This is the first discovery of a candidate gene for sex determination in papaya, and it must be greatly helpful for elucidating its sex determination mechanisms. According to the hypothesis of sex chromosome development, a dominant female sterile gene on the Y chromosome was supposed to determine male phenotype. Since an allele of the *SVP-like* gene in the Y chromosome encoded an intact protein with MADS-box and K-box, while its allele was absent in the X chromosome, it coincides with the predicted male determination gene. Also, considering that Y^h chromosome was predicted to be derived from Y chromosome, hermaphrodite was determined by alternation of genes in the Y chromosome. In the *SVP-like* gene, an insertion of a transposable element caused structural alternation of the allele in the Y^h chromosome that satisfied requirement of male/hermaphrodite determination gene. In papaya, since less difference of phenotypes, including gene expression, was observed among sex types in vegetative phase at least, differentiation of floral organs was assumed to be their major distinction among different sex types. Well-known genes for floral development

were not found in the genome sequences of papaya sex chromosomes, indicating that other regulatory genes for floral organs might be responsible for sex determination. It was not expected that such regulatory genes were frequently located in the sex chromosomes. Therefore, although male/female or male/hermaphrodite differentiation was independently occurred event, it is reasonable that alleles of the same gene in the sex chromosomes determined male, female and hermaphrodite.

As described in the chapter 1 and 3, *SVP* in *Arabidopsis* is known to encode a MADS-box transcription factor. In *A. thaliana*, SVP directly represses expression of *FT* gene, and its loss of function caused early flowering time. Considering the results of sequence similarity and phylogenetic analysis, the product of *SVP-like* gene in papaya seems to have a function as transcriptional factor like SVP protein in *A.thaliana*. It was presumed that there were two possible target genes controlled by *SVP-like* gene in papaya. One is the *FT* gene, which is known as the determinable gene for flowering. In papaya genome, since an autosomal paralogue of the *SVP-like* gene was also found and expressed in all the sex types (Figure 3.6c, d), this gene could be responsible for authentic role of SVP as a regulator of flowering time. Therefore, the *SVP-like* gene in the sex chromosome might have additional function, even when it regulated expression of the *FT* gene. Recent studies showed that *FT* had multiple functions other than flowering control in higher plants, like regulation of stomatal opening (Kinoshita et al. 2011) or storage organ formation in potato (Navarro et al. 2011). Present study did not show that *FT* gene was transcriptionally regulated by the *SVP-like* gene in papaya, and no evidences were demonstrated about relationship between *FT* and sex determination.

However, interestingly, *FT* gene is located in the sex chromosomes in papaya, and sequence polymorphisms were found between its alleles in the X and Y^h chromosomes (data not shown), Another possibility was that product of the *SVP-like* gene in the Y chromosome controlled genes for development of floral organs in papaya. Recent studies have reported involvement of *SVP* in the development of flowers, particularly regulation of B and C class genes in the ABC model of flower development (Gregis et al. 2006, Gregis et al. 2009, Méndez-Vigo et al. 2013, Jaudal et al. 2014, Wu et al. 2012). Since B and C class genes are involved in determination of stamens and pistil, *SVP* had potential to control development of sexual floral organs.

The *SVP-like* gene encoded MADS-box domain for DNA binding, and K-box domain for interaction with other protein. Generally, MADS-box proteins are known to form homo- or hetro- dimers when binding to target DNA, and by interaction with different proteins, the targets and function are known to change. In the Y^h chromosome, allele of the *SVP-like* gene was transcribed, regardless of transposon-insertion and it encoded truncated version of protein carrying only the K-box. It was still unknown whether this truncated protein was actually expressed and functional, but it might be associated with ambiguity of hermaphroditic phenotype in papaya plants. In papaya, growth of hermaphrodite plant at high temperature sometimes causes male flowers. It is called as 'sex reversal' and is one of obstacles in agricultural production of papaya fruits. As described in chapter 3, both male and hermaphrodite flowers carried both male and female floral organs at the early stage of flowers, and development of pistils was arrested in the male plant. This sex reversal might be due to instability of truncated SVP-like protein carrying

the K-box only in hermaphrodite flowers.

Thus, *SVP-like* gene in the sex chromosome was the most likely candidate of the gene for sex determination. For its certification, functional analysis of this gene, including transgenic analysis, was necessary in papaya plant. Also, further biochemical studies will allow revealing target sequences of binding or proteins interacting with product of the SVP-like gene, and understanding mechanisms of sex determination in papaya.

Other than the *SVP-like* gene, unique genes in the Y^h chromosome, but not in the X chromosome, were seen in the predicted genes. These genes could be also candidate genes of female/hermaphrodite or female/male determination, although further studies, including detailed structural and expression analysis, were necessary. On the other hand, between the Y and Y^h chromosome, no other possible candidate genes were found in the present study. By incorporating information of expressed genes in other tissues or stages, it might be possible to discover any genes carrying polymorphisms with amino acid change between the Y and Y^h chromosome as candidates of male/hermaphrodite determination genes.

In the evolutional model of sex chromosomes in higher plants including papaya, the ancestral autosome has diverged to X and Y chromosomes, due to male and female sterile mutations, followed by recombination suppression and accelerated genomic rearrangements. Consequently, genes in the Y chromosome were likely to be degenerated, resulting lethality in pairing of two Y chromosomes. Particularly, in papaya, Y^h chromosome was predicted to be derived from Y chromosome recently. Genome sequencing analysis and gene prediction in the present study proved some of these hypothesis of sex chromosomal evolution in papaya. Between the X and Y^h

chromosome, several genomic rearrangements with possible insertion and inversion events were found. It was detected twice as much putative transposons in Y^h chromosome (2,283 loci) as those in the X chromosome (1,014 loci). These results represented that genomic rearrangements and transposon accumulation were occurred in Y (Y^h) chromosome rather than X chromosome, which caused difference of sequence length between the X and Y^h chromosome, 4.5Mb and 7.7 Mb, respectively. By comparison of genome structure between the Y^h and Y chromosome, three large highly diverged regions were found, other than dispersed sequence polymorphisms. This also showed that genomic rearrangements were likely to be occurred in sex chromosomes in papaya. In contrast to difference of located genes between the X and Y^h chromosome, gene composition was almost identical and non-synonymous sequences differences were limited between the Y^h and Y chromosome, demonstrating that the Y and Y^h chromosomes were recently diverged. It is hypothesized that the accumulation of mutations led to gene degeneration in Y chromosome, and unique genes in X chromosome was actually more abundant than those in Y chromosome in human. In papaya, 553 and 929 genes on X and Y^h chromosomes, respectively, were predicted. According to this results, frequency of genes per kilobase was 0.13 and 0.12 in X and Y^h chromosomes, respectively, which is not significantly different. However, by RNA-seq analysis, 173 unique X chromosomal genes were expressed, while Y^h chromosome-specific transcribed genes were only 46 genes. Therefore, even though predicted genes in the Y^h chromosome was more abundant than those in the X chromosome, actually transcribed (active) genes were probably infrequent in the Y^h chromosome, which agreed with the prediction that Y-chromosomal genes are likely to be degenerated.

It was also supposed that unique active gene in the Y chromosome was limited, since the X chromosome should have alleles of Y chromosomal genes mostly. However, present analyses of genome sequence and transcriptome of mature leave demonstrated that 46 Y^h chromosome-specific expressed genes were found. As discussed in the chapter 1, it was expected that unique genes in the Y (Y^h) chromosome can be responsible for sex determination, but exploration of genes for sex determination might not be effective based on only this criterion (Y (Y^h) chromosome-specific genes).

Regardless of different gene dosage of 173 genes in the papaya X chromosome (X chromosome-specific expressed genes) between female and hermaphrodite, gene dosage effect in their expression level was not observed in majority of those genes. Currently, known sex chromosome gene dosage compensation mechanisms are, inactivation of either X chromosome on XX (mammals), activation of genes in X chromosome on XY (*Drosophila*) and partial expression activation of limited genes on the sex chromosome in heterogametic sex type (birds, platypus), probably based on gene-to-gene expression regulation (Mank et al. 2011). According to hermaphrodite/female expression ratio in the X chromosome unique genes and autosomal genes, their distribution of expression ratio did not show significant difference, indicating that expression bias was not probably present in expression level between autosomal and X chromosomal genes. Therefore, it was suggested that transcription level of the X chromosome could be coordinately regulated. In *S. latifolia,* expression of genes on the X and Y chromosome linked genes were analyzed by RNA-seq (Muyle et al. 2012). They reported that expression rate of most X chromosomal genes was elevated in males than that in females, when

expression of their alleles in Y chromosome was reduced in male. Still, further analysis was required, but any gene dosage compensation system was probably present in sex chromosomes in higher plants, according to previous and present studies. In contrast to mammalian sex chromosomes, only a portion of entire chromosome is responsible for sex determination and recombination suppression in higher plants. Therefore, it was interesting to elucidate how gene dosage was regulated in these sex chromosomes. As described above, genome sequencing of sex chromosomes and transcriptome analysis facilitated to reveal structure and function of the sex chromosomes or located genes in papaya. These results allowed to demonstrate several hypothesis of sex chromosome evolution in higher plant, and to identify a notable candidate gene for papaya sex determination, *SVP-like* gene. Although further studies were necessary for demonstrating its function in papaya sex determination, studies of *SVP-like* or its related genes could be greatly helpful for understanding whether sex determination was conserved or not in other dioecious species. Also, in this study, it was demonstrated that next generation sequencing technology was quite useful and accelerated to elucidate sex determination in papaya. Similar genomic approaches will be applicable to sex chromosome in other plant species, because recent NGS technologies including genome assembly analysis allow to constructing contigs or scaffolds for even in non-model organisms. Additionally, high-throughput long read sequencing (> 3000base/read), like a sequencer in Pacific Biosciences, is also available, which will facilitate obtaining longer contiguous sequences of complex genome like sex chromosomes.

As applications of the present results to papaya breeding, development of

genetically fixed hermaphrodite cultivar could be developed, which avoids female and hermaphrodite segregation in the seeds of self-fertilized hermaphrodite flowers, because only hermaphrodite plants are desired for papaya fruits production. If the *SVP-like* gene was responsible for hermaphrodite determination, its introduction in female could convert it to hermaphrodite. Also, its modification in hermaphrodite can improve stability of sexual phenotype, since high temperature often induces male flowers even in hermaphrodite plants (Tamaki et al. 2011). However, it could be difficult to change sexual phenotype only by manipulating a single gene. Then, introduction of lethality to female (XX) might be applicable to fix hermaphrodite plant genetically. For this purpose, any house-keeping gene on the X chromosome should be identified. Additionally, its gene should have an active allele in the Y^h chromosome. By its loss of function in allele of the X chromosome, pairing of XX chromosome must be lethal, but XY^h (hermaphrodite) is survived. Genome sequences of the sex chromosomes, their predicted genes and transcriptome data in the present study are inevitably applicable to find genes for introducing female lethality. Once candidate target genes were found, 'gene editing' technology, like CAS/CRISPR system, is recently available, which allows to introduce a mutation in the desired gene. Once genetically fixed hermaphrodite cultivar was developed, it must greatly contribute to production of papaya all over the world.

References

Ackerman CM, Yu Q, Kim S, Paull RE, Moore PH, Ming R (2008) B-class MADS-box genes in trioecious papaya: two paleoAP3 paralogs, CpTM6-1 and CpTM6-2, and a PI ortholog CpPI. Planta **227**: 741-753

Alvarez-Buylla ER, Azpeitia E, Barrio R, Benitez M, Padilla-Longoria P (2010) From ABC genes to regulatory networks, epigenetic landscapes and flower morphogenesis: Making biological sense of theoretical approaches. Seminars in Cell & Developmental Biology **21**: 108-117

Bai S-N, Xu Z-H (2013) Unisexual cucumber flowers, sex and sex differentiation. International Review of Cell and Molecular Biology, Vol 304 **304**: 1-55

Bhowmick BK, Satta Y, Takahata N (2007) The origin and evolution of human ampliconic gene families and ampliconic structure. Genome Research **17**: 441-450

Carvalho FA, Renner SS (2012) A dated phylogeny of the papaya family (Caricaceae) reveals the crop's closest relatives and the family's biogeographic history. Molecular Phylogenetics and Evolution **65**: 46-53

Charlesworth B, Charlesworth D (2000) The degeneration of Y chromosomes. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences **355**: 1563-1572

Cheng MK, Disteche CM (2006) A balancing act between the X chromosome and the autosomes. Journal of Biology **5**: 2-2

Delichere C, Veuskens J, Hernould M, Barbacar N, Mouras A, Negrutiu I, Moneger F (1999) SIY1, the first active gene cloned from a plant Y chromosome,

encodes a WD-repeat protein. EMBO Journal **18**: 4169-4179

Disteche CM (2012) Dosage compensation of the sex chromosomes. Annual Review of Genetics, **46**: 537-560

Filatov DA (2005) Substitution rates in a new Silene latifolia sex-linked gene, SlssX/Y. Molecular Biology and Evolution **22**: 402-408

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, Chen ZH, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology **29**: 644-652

Gregis V, Sessa A, Colombo L, Kater MM (2006) AGL24, SHORT VEGETATIVE PHASE, and APETALA1 redundantly control AGAMOUS during early stages of flower development in Arabidopsis. Plant Cell **18**: 1373-1382

Gregis V, Sessa A, Dorca-Fornell C, Kater MM (2009) The Arabidopsis floral meristem identity genes AP1, AGL24 and SVP directly repress class B and C floral homeotic genes. Plant Journal **60**: 626-637

Gschwend AR, Wai CM, Zee F, Arumuganathan AK, Ming R (2013) Genome size variation among sex types in dioecious and trioecious Caricaceae species. Euphytica **189**: 461-469

Hartmann U, Hohmann S, Nettesheim K, Wisman E, Saedler H, Huijser P (2000) Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis. Plant Journal **21**: 351-360

Jang S, Torti S, Coupland G (2009) Genetic and spatial interactions between FT, TSF and SVP during the early stages of floral induction in Arabidopsis.

Plant Journal **60**: 614-625

Jaudal M, Monash J, Zhang L, Wen J, Mysore KS, Macknight R, Putterill J (2014) Overexpression of Medicago SVP genes causes floral defects and delayed flowering in Arabidopsis but only affects floral development in Medicago. Journal of Experimental Botany **65**: 429-442

Johnson MH (1982) X-chromosome inactivation and the control of gene expression. Nature **296**: 493-494

Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schuetz F, Daish T, Gruetzner F, Kaessmann H (2012) Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. PLoS Biology **10**: e1001328

Kapustin Y, Souvorov A, Tatusova T, Lipman D (2008) Splign: algorithms for computing spliced alignments with identification of paralogs. Biology Direct **3**: 20

Kazama Y, Fujiwara MT, Koizumi A, Nishihara K, Nishiyama R, Kifune E, Abe T, Kawano S (2009) A SUPERMAN-like gene is exclusively expressed in female flowers of the dioecious plant Silene latifolia. Plant and Cell Physiology **50**: 1127-1141

Kinoshita T, Ono N, Hayashi Y, Morimoto S, Nakamura S, Soda M, Kato Y, Ohnishi M, Nakano T, Inoue S-i, Shimazaki K-i (2011) FLOWERING LOCUS T regulates stomatal opening. Current Biology **21**: 1232-1238

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and clustal X version 2.0. Bioinformatics **23**: 2947-2948

Mank JE (2009) The W, X, Y and Z of sex-chromosome dosage compensation. Trends

in Genetics **25**: 226-233

Mank JE, Hosken DJ, Wedell N (2011) Some inconvenient truths about sex chromosome dosage compensation and the potential role of sexual conflict. Evolution **65**: 2133-2144

Martin A, Troadec C, Boualem A, Rajab M, Fernandez R, Morin H, Pitrat M, Dogimont C, Bendahmane A (2009) A transposon-induced epigenetic change leads to sex determination in melon. Nature **461**: 1135-1138

Matsumura H, Bin Nasir KH, Yoshida K, Ito A, Kahl G, Kruger DH, Terauchi R (2006) SuperSAGE array: the direct use of 26-basepair transcript tags in oligonucleotide arrays. Nat Methods **3**: 469-474

Matsumura H, Ito A, Saitoh H, Winter P, Kahl G, Reuter M, Kruger DH, Terauchi R (2005) SuperSAGE. Cellular Microbiology **7**: 11-18

Matsumura H, Reich S, Ito A, Saitoh H, Kamoun S, Winter P, Kahl G, Reuter M, Kruger DH, Terauchi R (2003) Gene expression analysis of plant host-pathogen interactions by SuperSAGE. Proceedings of the National Academy of Sciences of the United States of America **100**: 15718-15723

Matsumura H, Yoshida K, Luo S, Kimura E, Fujibe T, Albertyn Z, Barrero RA, Krueger DH, Kahl G, Schroth GP, Terauchi R (2010) High-Throughput SuperSAGE for digital gene expression analysis of multiple samples using next generation sequencing. PLoS One **5**: e12010

Matsumura H, Yoshida K, Luo S, Kruger DH, Kahl G, Schroth GP, Terauchi R (2011) High-throughput SuperSAGE. Methods in molecular biology (Clifton, N.J.) **687**: 135-146

Mendez-Vigo B, Martinez-Zapater JM, Alonso-Blanco C (2013) The flowering

repressor SVP underlies a novel arabidopsis thaliana qtl interacting with the genetic background. PLoS Genetics **9**: e1003289

Ming R, Bendahmane A, Renner SS (2011) Sex chromosomes in land plants. Annual Review of Plant Biology 62: 485-514

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 DW, Albert H, Wang M-L, Zhu YJ, Schatz M, Nagarajan N, Acob RA, 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-Perez R, Torres MJ, Feltus FA, 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, dePamphilis 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

Ming R, Yu Q, Moore PH (2007) Sex determination in papaya. Seminars in Cell & Developmental Biology **18**: 401-408

Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GAB (2012) Rapid *de novo* evolution of X chromosome dosage compensation in Silene latifolia, a plant with young sex chromosomes. PLoS Biology **10**: e1001308

Na J-K, Wang J, Murray JE, Gschwend AR, Zhang W, Yu Q, Navajas-Perez R,

Feltus FA, 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

Navarro C, Abelenda JA, Cruz-Oro E, Cuellar CA, Tamaki S, Silva J, Shimamoto K, Prat S (2011) Control of flowering and storage organ formation in potato by FLOWERING LOCUS T. Nature **478**: 119-122

Nicolas M, Marais G, Hykelova V, Janousek B, Laporte V, Vyskot B, Mouchiroud D, Negrutiu I, Charlesworth D, Moneger F (2005) A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. PLoS Biology **3**: e4

Song Y, Ma K, Ci D, Chen Q, Tian J, Zhang D (2013) Sexual dimorphic floral development in dioecious plants revealed by transcriptome, phytohormone, and DNA methylation analysis in Populus tomentosa. Plant Molecular Biology **83**: 559-576

Tamaki M, Urasaki N, Sunakawa Y, Motomura K, Adaniya S (2011) Seasonal variations in pollen germination ability, reproductive function of pistils, and seeds and fruit yield in papaya (Carica papaya L.) in Okinawa. Journal of the Japanese Society for Horticultural Science **80**: 156-163

Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Molecular Biology and Evolution **30**: 2725-2729

Urasaki N, Tarora K, Shudo A, Ueno H, Tamaki M, Miyagi N, Adaniya S, Matsumura H (2012) Digital transcriptome analysis of putative sex-determination genes in papaya (Carica papaya). PLoS One **7**: e40904

van Bakel H, Stout JM, Cote AG, Tallon CM, Sharpe AG, Hughes TR, Page JE (2011) The draft genome and transcriptome of Cannabis sativa. Genome Biology **12**: R102

VanBuren R, Ming R (2013) Organelle DNA accumulation in the recently evolved papaya sex chromosomes. Molecular Genetics and Genomics **288**: 277-284

Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene-expression. Science **270**: 484-487

Veyrunes F, Waters PD, Miethke P, Rens W, McMillan D, Alsop AE, Gruetzner F, Deakin JE, Whittington CM, Schatzkamer K, Kremitzki CL, Graves T, Ferguson-Smith MA, Warren W, Graves JAM (2008) Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. Genome Research **18**: 965-973

Wang X, Wang H, Wang J, Sun R, Wu J, Liu S, Bai Y, Mun J-H, Bancroft I, Cheng F, Huang S, Li X, Hua W, Wang J, Wang X, Freeling M, Pires JC, Paterson AH, Chalhoub B, Wang B, Hayward A, Sharpe AG, Park B-S, Weisshaar B, Liu B, Li B, Liu B, Tong C, Song C, Duran C, Peng C, Geng C, Koh C, Lin C, Edwards D, Mu D, Shen D, Soumpourou E, Li F, Fraser F, Conant G, Lassalle G, King GJ, Bonnema G, Tang H, Wang H, Belcram H, Zhou H, Hirakawa H, Abe H, Guo H, Wang H, Jin H, Parkin IAP, Batley J, Kim J-S, Just J, Li J, Xu J, Deng J, Kim JA, Li J, Yu J, Meng J, Wang J, Min J, Poulain J, Hatakeyama K, Wu K, Wang L, Fang L, Trick M, Links MG, Zhao M, Jin M, Ramchiary N, Drou N, Berkman PJ, Cai Q, Huang Q, Li R, Tabata S, Cheng S, Zhang S, Zhang S, Huang S, Sato S, Sun S, Kwon S-J, Choi S-R, Lee T-H, Fan W, Zhao X, Tan X, Xu X, Wang Y, Qiu Y, Yin Y, Li Y, Du Y,

Liao Y, Lim Y, Narusaka Y, Wang Y, Wang Z, Li Z, Wang Z, Xiong Z, Zhang Z (2011) The genome of the mesopolyploid crop species Brassica rapa. Nature Genetics **43**: 1035-1039

Wu R-M, Walton EF, Richardson AC, Wood M, Hellens RP, Varkonyi-Gasic E (2012) Conservation and divergence of four kiwifruit SVP-like MADS-box genes suggest distinct roles in kiwifruit bud dormancy and flowering. Journal of Experimental Botany **63**: 797-807

Xiong Y, Chen X, Chen Z, Wang X, Shi S, Wang X, Zhang J, He X (2010) RNA sequencing shows no dosage compensation of the active X-chromosome. Nature Genetics **42**: 1043-1047

Yu Q, Hou S, Feltus FA, Jones MR, Murray JE, 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-linked genes. Plant Journal **53**: 124-132

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. Molecular Genetics and Genomics **278**: 177-185

Yu Q, Navajas-Perez R, Tong E, Robertson J, Moore P, Paterson A, Ming R (2008) Recent origin of dioecious and gynodioecious y chromosomes in papaya. Tropical Plant Biology **1**: 49-57

Yu Q, Tong E, Skelton RL, Bowers JE, Jones MR, Murray JE, Hou S, Guan P, Acob RA, Luo M-C, 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

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 Research **18**: 1938-1943

Zluvova J, Lengerova M, Markova M, Hobza R, Nicolas M, Vyskot B, Charlesworth D, Negrutiu I, Janousek B (2005) The inter-specific hybrid Silene latifolia x S-viscosa reveals early events of sex chromosome evolution. Evolution & Development **7**: 327-336

Acknowledgments

I would firstly like to deeply thank my supervisor, Dr. Hideo Matsumura for his detailed and constructive comments, and his important supports during this work. I also wish to express my gratitude to Dr. Naoya Urasaki and Mr. Kazuhiko Tarora in Okinawa prefectural agriculture research center, and Dr. Ryohei Terauchi, Dr. Kentaro Yoshida and Mr. Satoshi Natsume in Iwate Biotechnology Research Center for their helpful advice and technical supports.

The reviewers of this thesis, Dr. Hiroyasu Ebinuma, Dr Nobuaki Hayashida, Dr, Tomoaki Horie and Dr. Akira Kanno provided advice that helped make this thesis more substantial.

This work was partly supported by Grant-in-Aid for Global COE Program, Grants for Excellent Graduate Schools by the Ministry of Education, Culture, Sports, Science and Technology. This work was also supported by JSPS (Japan Society for the Promotion of Science) grant no. 25450004, and the Program for Promotion of Basic and Applied Researches for Innovations in Bio-oriented Industry (PROBRAIN).

Finally I would like to express my deepest appreciation to my parents for their numerous supports, every members of the Division of Gene Research, Department of Life Science, research Center for Human and Environmental Science, Shinshu University, and also to Dr. Takuo Hoya and Dr. Teruhiko Hamanaka for their medical attendance done for the author's glaucoma.

List of Publications

Naoya Urasaki, Kazuhiko Tarora, Ayano Shudo, Hiroki Ueno, Moritoshi Tamaki, Norimichi Miyagi, Shinichi Adaniya, Hideo Matsumura, Digital Transcriptome Analysis of Putative Sex-Determination Genes in Papaya (*Carica papaya*), PLoS one **7**: e40904

Hiroki Ueno, Naoya Urasaki, Satoshi Natsume, Kentaro Yoshida, Kazuhiko Tarora, Ayano Shudo, Ryohei Terauchi and Hideo Matsumura, Genome sequence comparison reveals a candidate gene involved in male-hermaphrodite differentiation in papaya (*Carica papaya*) trees. Molecular Genetics and Genomics: in press