

## ABSTRACT

### THE EVOLUTIONARY TEMPO OF SEX CHROMOSOME DEGRADATION IN *CARICA PAPAYA*

by Meng Wu

The young Y chromosome in papaya provides a good model to study the evolutionary tempo of Y chromosome degradation. In this study, comparative genetic analyses of 70 X-Y gene pairs showed that Y-linked genes have significantly more deleterious mutations than X-linked genes, suggesting the degeneration on the Y. However, this asymmetric evolutionary pattern is only confined to the oldest stratum, indicating a relatively slow degeneration tempo of the papaya Y chromosome compared to neo-sex chromosomes in *Drosophila*. Population genetic analyses based on protein-coding sequence variation of six X-Y gene pairs on the oldest stratum also identified elevated nonsynonymous polymorphisms and reduced codon bias on the target X region relative to the neutrally evolving autosomal regions. This may be explained by weaker purifying selection and a widespread genetic hitchhiking. Here we discuss how the unique genomic features of pericentromeric location of the papaya sex-determining region may have influenced its evolution.

THE EVOLUTIONARY TEMPO OF SEX CHROMOSOME DEGRADATION IN  
*CARICA PAPAYA*

A Thesis

Submitted to the

Faculty of Miami University

in partial fulfillment of

the requirements for the degree of

Master of Science

Department of Biology

by

Meng Wu

Miami University

Oxford, Ohio

2014

Advisor \_\_\_\_\_

Dr. Richard C. Moore

Reader \_\_\_\_\_

Dr. James J. Hickey

Reader \_\_\_\_\_

Dr. Chun Liang

©

Meng Wu

2014

## TABLE OF CONTENTS

Section	Page
INTRODUCTION .....	1
THESIS OVERVIEW.....	9
MATERIALS AND METHODS .....	9
RESULTS .....	13
DISCUSSION .....	17
CONCLUSION .....	21
LITERATURE CITE .....	23
TABLES .....	30
FIGURES .....	34
APPENDIX .....	40

## LIST OF TABLES

Table 1. PAML-based tests of evolutionary rate differences between X-linked and Y-linked genes in different strata (with pseudogenes), using homologous genes of <i>V. monoica</i> as outgroup sequences.....	30
Table 2. Summary of nucleotide diversity for focal loci .....	31
Table 3. Tests of selection based on the site frequency spectrum ( <i>D</i> and <i>H</i> ), haplotype frequency spectrum ( <i>EW</i> ), or combinations thereof ( <i>DH</i> and <i>DHEW</i> ) .....	32
Table 4. Coding sequence polymorphism and divergence in X/Y-linked loci .....	33
 Appendix:	
Table S1. <i>C. papaya</i> and <i>J. dolichaula</i> accessions used for analysis .....	40
Table S2. Comparison of phylogenetic distances between <i>V. monoica</i> and <i>J. dolichaula</i> relative to <i>C. papaya</i> .....	43
Table S3. Summary of focal loci source and probable function .....	45
Table S4 Diversity ( $\theta_w$ ) and site frequency spectrum tests of selection for 25 autosomal loci used as neutral comparisons .....	47
Table S5. Primer list for Y-linked genes' amplifications .....	48
Table S6. Primer list for X-linked genes' amplifications .....	50
Table S7. List of functional X-Y gene pairs .....	52
Table S8. List of X-Y gene pairs of which Y-linked gene is pseudogene .....	54
Table S9. List of X-Y gene pairs of which X-linked gene is pseudogene .....	55
Table S10. List of X-Y gene pairs of which X-linked and Y-linked gene are both pseudogenes .....	56
Table S11. PAML-based tests of rate differences between each X-Y gene pair, using homologous genes of <i>V. monoica</i> as outgroup sequences .....	57
Table S12. Sequencing coverage .....	62
Table S13. Maximum-likelihood HKA analysis of silent polymorphism of papaya sex-linked loci against 25 autosomal loci. ....	63
Table S14. Distribution of functional genes and pseudogenes in the X/Y-specific regions .....	64

## LIST OF FIGURES

Figure 1. Diagram of chromosomal regions used in this study .....	35
Figure 2. Matched-pair analysis comparing the difference in synonymous or nonsynonymous mutations between X and Y alleles versus mean number of differences between X and Y alleles .....	36
Figure 3. Linkage disequilibrium plots of parsimoniously informative sites for X-linked and Y-linked alleles in the target regions .....	37
Figure 4. Plot of $K_a/K_s$ versus $K_s$ of X/Y-linked genes .....	38
Figure 5. Comparison $\pi_{(a)}/\pi_{(s)}$ ratio among X/Y-linked and autosomal genes .....	39

## ACKNOWLEDGEMENTS

First, I would thank to my advisor Dr. Richard Moore for his kind help in this two years. He was very nice to every student and provided very helpful advice to me along my research project. His enthusiasm and interest in molecular evolution inspires me to become a motivated student in his lab. I would also like to thank all the lab members in the Moore lab, in particular Ms. Sandra Mardonovich and Ms. Fiona Lappin, for their guidance and support to me. I would also like to thank my friend Mr. Zheng Li for frequently sharing interesting scientific papers and discussing research project with me. I also appreciate the help from my thesis committee members, Dr. James Hickey and Dr. Chun Liang, to promote my research proposal and thesis writing. This project is also supported by the help from our collaborator Dr. Ray Ming in University of Illinois Champaign-Urbana for providing sequencing data, and Dr. Oscar Rocha in University of Kent State University for preparing papaya materials. This project is funded by the National Science Foundation and the Academic Challenge grants through the Department of Botany at Miami University.

# INTRODUCTION

## Overview of sex determining mechanisms

Most multicellular sexual species are dioecious, consisting of two sexes, male and female. The separation of male and female sex function in separate individuals is believed to promote the species fitness even though only one sex (female) is capable of bearing offspring. First, separation of sex allows for the male or female developments in two different individuals, thus enhancing the efficiency of resource allocation (Smith 1978). Second, hybridization between male and female individuals with different genetic backgrounds can also prevent reduced fitness associated with close inbreeding (Charlesworth 1991; Charlesworth and Charlesworth 1987).

The evolution of separate sexes has occurred independently in a diversity of organisms, and diverse developmental mechanisms can affect or interact with each other to determine the expression of separate sexes (Bull 1979; Werren and Beukeboom 1998). In some organisms, sex expression is determined by environmental factors. For example, temperatures during the incubation time of eggs influence the sex ratio of offspring, such as in alligators (Gilbert 2010). Alternatively, the number of chromosomes in the genome can influence the sex determination. In the insect order *Hymenoptera*, sexes depend on ploidy level. Fertilized eggs develop into females and unfertilized eggs develop into males (Grimaldi and Engel 2005). Also, genomic imprinting can regulate the expression of DNA segments in the genome and thus epigenetically determine the sex, as in some insects. Individuals with all euchromatic chromosomes become females while those with heterochromatinised paternal genome become males (Nur 1990). It has also been reported that sex determination can be affected by quantitative trait loci. For example, polygenic sex determination was found in zebrafish such that the development of maleness or femaleness is affected the number of sex determining genes with minor effects (Liew et al. 2012). Here is just a partial list of ways gender can be expressed, and there are far more sex determination systems not reported right here.

The most familiar form of genetic sex determination involves sex chromosomes. Sex chromosomes have independently evolved in various eukaryotic species including animals, plants and fungi. There are three principle types of sex chromosome systems in nature: male heterogamety (XY system), female heterogamety (ZW system) and haploid phase determination (UV system; Bachtrog et al. 2011). Male heterogamety (XY) and female heterogamety (ZW) are found only in diploid species (Bull 1983). The XY male heterogametic system is the most



familiar, as it is prevalent in mammalian species, including humans; it is also found in dipterans, such as the fruit fly. The ZW female heterogametic system is prevalent in birds, snakes, and lepidopteran species. In haploid dominant species (e.g. some algae and bryophytes), the male and female sexes are determined during the haploid phase of life cycle (Bull 1983). The haploid spores with the V chromosome develop into male while the haploid spores with the U chromosome develop into female.

### **The two-step mutational model of sex chromosome formation**

Studies of recently evolved sex chromosomes support the hypothesis that sex chromosomes evolve from autosomes through the acquisition of a pair of feminizing and masculinizing mutations (Moore et al. 2003; Liu et al. 2004; Spigler et al. 2008; Yin et al. 2008). In this model of sex chromosome evolution, the initiation of sex chromosomes follows two-step mutations on a pair of autosomes having sex determining loci that lead to the emergence of separate sexes from a hermaphrodite ancestor (Bachtrog 2013). In this case, the pair of autosomes bearing the sex determining loci is called proto-sex chromosomes. In the evolution of XY sex chromosome system, one male-determining dominant mutation arises that suppresses female function and promotes male reproductive traits (mutation:  $f \rightarrow Su^F$ ; Charlesworth 2013), and female-determining recessive mutation on the homologous chromosome causes male-sterility but is necessary for female development (mutation:  $M \rightarrow m$ ; Charlesworth 2013). The chromosome bearing  $Su^F$  and  $M$  alleles becomes the proto-Y chromosome and its homologue bearing  $f$  and  $m$  alleles becomes the proto-X chromosome (Charlesworth 2013). Still, this pair of chromosomes can recombine with each other, resulting in a shuffling of these sex determining alleles (Charlesworth 2013). The recombination of these alleles will lead to either hermaphrodites or neuters. In hermaphrodites, the available resources devoted to female sex expression would be reduced under the effect of the male-promoting allele. Alternatively, a neutral could not produce offspring (Charlesworth 2013). Thus, selection for greater fitness drives the population to inhibit recombination between the two sex determination loci (Rice 1987; Charlesworth 2013). Through chromosomal rearrangements such as chromosomal inversion, suppression of recombination between the proto-sex chromosomes can be achieved (Bachtrog 2013). This region of suppressed recombination is called the sex determination region (SDR). In this way, a nascent pair of sex chromosomes forms.

### **The expansion of the SDR through the acquisition of sexually antagonistic loci**

To further promote fitness in dioecious population, the Y tends to be masculinized by accumulation of genes beneficial to males (male-biased genes) while the X tends to be feminized through accumulating more female-biased genes (Rice 1987). For examples, a potential male-biased gene can promote pollen production in male flower and a potential female-biased gene can increase female flower size. Under natural selection, the frequency of sexually-antagonistic (SA) genes (i.e. male-/female-biased genes) around the SDR regions may be influenced. For example, if a gene closely linked to the SDR region causes male-beneficial or a female-detrimental effect, it might be selectively favored to be transmitted with male individuals and thus fixed on the Y chromosome (Rice 1987). Selection for the reduced recombination on the SDR adjacent SA genes may be responsible for expansion of the SDR (Rice 1987; Bachtrog 2006). A step-wise expansion of the SDR can be supported by the presence of distinct evolutionary strata in the SDR of some sex chromosome systems. Evolutionary strata of the SDR showing different levels of X-Y sequence divergence are caused by a series of chromosomal inversions to initiate recombination suppression at different evolutionary periods of time (Bergero and Charlesworth 2009). The presence of evolutionary strata in sex chromosomes was first discovered in mammals (Lahn and Page 1999; Sandstedt and Tucker 2004). Evolutionary strata have since been discovered in avian and snake ZW sex chromosomes (Lawson-Hadnley et al. 2004; Nam and Ellegren 2008; Vicoso et al. 2013). In smut fungi, there is a record of evolutionary strata along a small, mating-type-specific region (Votintseva and Filatov 2009). Increasing evidence of evolutionary strata also have been found in various plant species, including flowering plants and mosses (Bergero et al. 2007; Wang et al. 2012; McDaniel et al. 2013). Thus, this step-wise expansion of the SDR by chromosomal inversions seems to be common, although recent genomic analyses have questioned the reality of strata in humans and in *Silene* (Chibalina and Filatov 2011; Sayres and Makova 2013). Following recombination suppression between the X and Y, X/Y-linked alleles located in the SDR would evolve independently in separate lineages (XY system). Analyzing the molecular evolutionary signatures of X/Y-linked genes sampled from different aged strata could provide important clues as to how sex-linked genes evolve after different time periods experiencing recombination suppression.

## **Mechanisms influencing Y chromosome evolution**

The SDR of the Y chromosome, termed the male-specific region of the Y (MSY), tends to accumulate deleterious mutations in an irreversible manner similar to what happens in an asexual genome. This evolutionary process within the MSY can be explained by the Hill-Robertson interference effect on regions of reduced or suppressed recombination. The Hill-Robertson interference effect describes the overall reduced selective efficiency in finite populations due to the linkage between sites under selection (Hill and Robertson 1966). Its application to understanding the evolution of the Y chromosome can be seen in several ways. First, a neutral site on the Y which is linked to another site under directional selection would experience reduced effective population size,  $N_e$ , decreasing the efficacy of selection by enhancing drift effects (Gordo and Charlesworth 2006). Considering there is same numbers of male and female individuals in a population, there is only one Y chromosome relative to four autosomes in the population; thus, any site linked to the MSY will have a  $N_e$  equal to  $\frac{1}{4}$  that of autosomes. Fluctuations in allele frequency are largely affected by  $N_e$  in finite populations. In small population, any allele is easily lost or fixed randomly due to genetic drift no matter it is beneficial or not. Thus, the relatively small  $N_e$  of the MSY may lead to the stochastic fixation of deleterious mutations or loss of beneficial mutations. In an asexual or non-recombining genome, if mutation-free individuals or haplotypes are stochastically lost in a finite population, they cannot be regained due to the lack of recombination (Manning and Thompson 1984). This stochastic process is called Muller's ratchet and is one way to drive the Y chromosome to accumulate deleterious mutations leading to the Y chromosomal degeneration (Charlesworth 1996). Moreover, the low  $N_e$  of loci on the MSY restricts the ability of purifying selection to purge deleterious mutation out off population, as the efficacy of purifying selection is proportional to the  $N_e$ . In addition, the tight linkage between sites on the MSY may also elevate the influence of genetic hitchhiking. When there is positive selection on a strong beneficial mutation on the MSY, linked deleterious mutations are likely to be drags along to fixation in the population (Bachtrog 2013). Under the influence of the Hill-Robertson interference effect on the recombination-suppressed MSY region, Y-linked genes are theorized to accumulate amino acid substitutions at an accelerated rate due to increasing deleterious mutations (Bachtrog and Charlesworth 2002). Furthermore, the Hill-Robertson interference effect also limits the adaptive

evolution of Y chromosomes by a process called the “Ruby in the Rubbish” (Bachtrog 2013). This process happens when a slightly beneficial mutation is linked to deleterious mutations. Because there is no recombination, the beneficial allele is less likely to become independent of surrounding deleterious mutations. And when purifying selection works on deleterious mutations, this beneficial allele would be purged simultaneously, thus making it difficult for the MSY region to gain beneficial mutations.

### **Mechanisms influencing X chromosome evolution**

Unlike the MSY region, the X-specific region (corresponding X region of the MSY) can still recombine with another X chromosome in females. Thus, the X chromosome is thought to be non-degenerative and to evolve as other regions of the genome (Bachtrog et al. 2009; Wilson and Makova 2009). However, selective forces are expected to be stronger and more visible on the X relative to the autosome (Schaffner 2004). One explanation for this is that the increased exposure of recessive alleles to selection on the X-specific region, as there is only one X chromosome in the male genome. For example, X-recessive deleterious alleles would be more efficiently weeded out (Schaffner 2004). Also, because the X can only recombine when present as part of a female genome, theoretically only two thirds of the X population is undergoing recombination compared to 100% of the autosomes. Genomics study in the human genome has supported that the recombination rate of the X chromosome is almost two thirds of the autosomes (Kong et al. 2002). Thus, linkage disequilibrium (LD) would be greater on the X chromosome than the autosomes (Schaffner 2004). The Hill-Robertson interference effect might be elevated to some extent on the X due to the greater LD. However, evidence from a large dataset of single nucleotide polymorphism sites (Sachidanandam et al. 2001) does not support increased rates of selection on the X chromosome in humans; this needs further study.

### **Neo-sex chromosomes in animals**

Studying young sex chromosome systems is helpful to understand the mechanisms driving sex chromosome evolution in early stages. Young sex chromosomes have been found in animals such as *Drosophila* and sticklebacks. However, the young sex chromosomes in these systems are not evolved from a new pair of autosomes by the two-step mutations. Instead, they were formed by a chromosomal infusion between an autosome and an ancestral Y chromosome and are called

“neo-sex chromosomes” to distinguish them from those sex chromosomes that form *de novo* from autosomes. The neo-Y is derived from the Y-autosomal fusion and takes on the characteristics of the Y chromosome, including suppressed recombination between it and its previous autosomal homolog (the neo-X chromosome; Bachtrog and Charlesworth 2002; Peichel et al. 2004; Kitano et al. 2009). The neo-sex chromosomes in animals are good models to study the effect of recombination suppression on sex chromosome evolution, especially the degeneration of the Y chromosome (Kitano and Peichel 2012). However, the whole neo-Y chromosome is completely sheltered from recombination immediately after the Y-autosomal fusion (Bachtrog et al. 2009). Thus, the neo-sex chromosomes are not ideal to study the step-wise expansion of the SDR. Also, lack of records on sex chromosomes of a variety of ages in diverse animal species, from proto-sex chromosomes to heteromorphic sex chromosomes, limits the study on evolutionary history of sex chromosomes.

### **Young *de novo* sex chromosomes in plants**

Unlike neo-sex chromosomes in animals with limited age variance, young sex chromosomes at different evolutionary stages have been discovered in flowering plants (Ming et al. 2007b). The least divergent sex chromosomes were reported in *Fragaria virginiana*; in this taxon, the male and female sexes are controlled by two sex-determining loci on a pair of homologous autosomes (Spigler et al. 2008). The production of hermaphrodites and neuters in populations indicates that recombination still occurs between the two loci. This pair of chromosomes with their loosely linked sex determining loci represents a good example of proto-sex chromosomes (Spigler et al. 2008). A second type of young sex chromosomes can be found in *Asparagus officinalis*; here a non-recombining region has developed surrounding the sex determining loci (Telgmann-Rauber et al. 2007). However, the YY genotype is still viable, indicating that important genes are still functional on the Y; thus, degeneration process of Y chromosome has not been initiated. A third type of young sex chromosome exists wherein the non-recombining MSY region has expanded and the degeneration of the Y chromosome has been initiated (Ming et al. 2007b). The *Carica papaya* sex chromosome is at this stage: the YY genotype is lethal, but the MSY region is still relatively small and the sex chromosomes appear homomorphic at the cytological level (Ming et al. 2007a). In a later stage, the expansion of the MSY region is expedited by the accumulation of transposable elements and sequence duplications, and the Y

chromosome is further degraded (Ming et al. 2007b). The sex chromosomes in *Silene latifolia* have these properties and the Y chromosome is physically larger than the X chromosome. The various young sex chromosomes at different evolutionary stages found in flowering plants provide good materials to study the evolutionary process of sex chromosomes. Since the MSY region in papaya is at the third evolutionary stage of Y chromosome evolution, when degeneration is initiated, papaya is a good model to study the early stage of degeneration of the Y chromosome

### **Sex chromosomes in papaya**

*Carica papaya* is an herbaceous tropical plant in the family *Caricaceae*. In this family, there are 32 dioecious species, two polygamous species (*Vasconcellea cundinamarcensis* and *C. papaya*), and one monoecious species (*Vasconcellea monoica*; Ming et al. 2007a). The prevalence of dioecy in the family *Caricaceae* suggests that the sex determination system may have evolved early in this family and is shared in most of these species (Wu et al. 2010). Papaya is a polygamous species with three basic sex forms: male, female and hermaphrodite. Morphologically, male individuals have long and pendulous inflorescences bearing many slender staminate flowers without pistils; female individuals are characterized by short inflorescences with few large pistillate flowers lacking stamens. Hermaphrodite individuals, have short inflorescences like female individuals, but bear bisexual flowers (Ming et al. 2007a).

Papaya has a relatively small genome of 372 Mb with 9 pairs of chromosomes (Arumuganathan and Earle 1991; Ming et al. 2008). The X/Y sex chromosome system is on the largest linkage group (LG1). Two slightly different Y chromosomes exist: male-specific Y and hermaphroditic Y<sup>h</sup>. The Y<sup>h</sup> shares nearly 99% DNA similarity with Y (Yu et al. 2008) and I refer to both as the papaya Y in this work when the distinction is not important. Papaya has an active-Y sex determination system; males are XY, hermaphrodites are XY<sup>h</sup>, and females are XX. (Ming et al. 2007a). The MSY region is ~8.1 Mbps, comprising ~13% of the Y chromosome (Yu et al. 2009) and is a little larger than its X corresponding region (~5.3 Mbps; Wang et al. 2012). The relatively small SDR in papaya and low divergence between the SDR of the X and Y chromosomes support that X/Y chromosomes diverged from each other very recently, perhaps after the divergence of papaya (Wu et al. 2010). Even though dioecy predominates within the *Caricaceae*, molecular evidence indicated sex chromosomes evolved independently among

different species in this family (Wu et al. 2010).

Sequence comparisons between the X and the Y<sup>h</sup> have revealed two evolutionary strata in the SDR of papaya which correspond to two chromosomal inversions. The history of the older stratum can be dated back to ~7.0 MYA, while the most recent stratum was formed ~1.7 MYA (Wang et al. 2012). In addition, there is a collinear region on the SDR that was not formed by chromosomal inversion but was somehow incorporated into the SDR recently, potentially through the recruitment of sexually antagonistic mutations (Wang et al. 2012). The two evolutionary strata and the collinear region on the SDR of papaya contain 70 X/Y-paired genes at different divergence times, and thus provide a good model to study the temporal dynamics of the evolution of young sex chromosomes.

Another unique feature of the papaya MSY region and its corresponding X region is that they span the centromere and are therefore found in the pericentromeric region of the LG1 (Fig 1). Recombination rates in pericentromeric regions tend to be lower than that of most genomic regions (Choo 1998). In papaya, the reduced recombination rates were found in all nine major linkage groups, including the X chromosome (Chen et al. 2007; Yu et al. 2009). Thus, though the X-specific region may recombine in females, it exhibits a much lower rate than the most parts of the genome (Yu et al. 2009), while recombination is suppressed in the MSY region, just as in Y chromosomes of other systems. As such, both the MSY and the X-specific region of papaya sex chromosomes are likely to experience some of the side effects of suppressed/reduced recombination associated with the Hill-Robertson interference, including relaxed purifying selection, elevated genetic hitchhiking, greater Muller's ratchet, and slower adaptive evolution rates. If that is true, the Hill-Robertson interference effect should be stronger on the MSY than on the X-specific region, because recombination on the MSY is completely suppressed.

Previous studies have shown there are accumulations of repetitive sequences and pseudogenes on both the X/Y-specific region compared to the homologous autosomal regions in *V. monoica* (Gschwend et al. 2012; Wang et al. 2012), suggesting the Hill-Roberson inference effect may be enhanced on both the X/Y-specific region. Greatly reduced genetic diversity was also observed at previous target X alleles, indicating ongoing selection along the X-specific region (Weingartner and Moore 2012). As selective sweep is the reason leading to genetic hitchhiking, genetic hitchhiking may contribute more deleterious mutations to be fixed on the low-recombining X-specific region than other recombining region, similar to the Hill-Robertson

interference effect on the Y.

## **THESIS OVERVIEW**

In this study, I focused on the evolution of coding regions on the sex chromosomes of papaya. I expected to find the signature of gene decay on the MSY as has been found in other organisms (Bachtrog 2013). My comparative genetics analyses of X/Y-linked genes showed there are more deleterious mutations on the MSY relative to its X counterpart. However, this signature is confined to the oldest Stratum 1, suggesting that purifying selection has been significantly relaxed in the oldest stratum of the MSY to lead to the degeneration of Y-linked genes in that region. I also conducted population genetic analyses on one 0.8-Mb region of the MSY in stratum 1 surrounding a region suspected to contain the female suppressing sex determination gene (R. Ming, University of Illinois, personal communication) and its corresponding 1.4-Mb pericentromeric X region. Interestingly, I observed a similar degeneration pattern in the target 1.4-Mb X region as the MSY region, in which all X focal loci are highly linked and experience elevated nonsynonymous mutations and reduced codon bias. Then I proposed that the degeneration occurring on the target X region is caused by weaker purifying selection and a widespread hitchhiking effect due to the enhanced Hill-Robertson inference effect on this low-recombining pericentromeric region.

## **MATERIALS AND METHODS**

### **Plant materials**

Forty male and two hermaphrodite papayas were used for amplifying focal Y loci, and 24 female papaya individuals were used for focal X loci amplification in the population analyses. They were all wild individuals collected from five geographically dispersed regions of Costa Rica (Table S1; Brown et al. 2012). Plant tissues were stored at -80°C before DNA extraction. *V. monoica* was used as the outgroup species in sequence analyses of all sex-linked loci and some autosomal loci. The reason to use *V. monoica* is that it is only monoecious species in the *Caricaceae* and definitely has no sex chromosomes. The homologous sequences of sex-linked loci in *V. monoica* were pulled out through BLAST against the whole-genome short gun sequences (<http://www.life.illinois.edu/ming/NSFCaricaProjectSite/Introduction.html>). One more species *Jacaratia dolichaula* was used as outgroup for some autosomal loci from which



homologous sequences have been sequenced in our previous study (Brown 2013), when there were no homologous sequences found in the *V. monoica* database. Tajima's relative rate test (Tajima 1993) was conducted to compare dates of divergence of *V. monoica* and *J. dolichaula* relative to *C. papaya*. No significant difference of Tajima's relative rate (Avg.  $P = 0.47$ ) showed there is a similar phylogenetic distance from *V. monoica* and *J. dolichaula* to *C. papaya*, thus either of the two species could be used as outgroup species for sequence analysis (Table S2).

### Genes studied

70 X-Y gene pairs in the SDR were used for the comparative analyses, including 50 functional X-Y gene pairs, ten Y-pseudogene pairs (functional X-allele + Y-pseudogene), six X-pseudogene pairs (functional Y-allele + X-pseudogene) and four X-Y-pseudogene pairs (Table S7-9). Genes were sorted into three classes based on their positions on stratum 1, stratum 2 or the collinear region of the SDR (Wang et al. 2012). Gene sequences were pulled out based on the gene annotations (Wang et al. 2012). Coding sequences were ascertained through BLAST against the papaya expressed sequence tag dataset (EST; <http://www.ncbi.nlm.nih.gov/nucest>). X-Y paired coding sequences were aligned to their corresponding EST in BioEdit software (Hall 1999). Only that portion of the predicted X-Y coding sequence that corresponded to the EST was used in the analysis or ~82.78% (77057 bps) of the predicted X-Y paired coding sequences (Wang et al. 2012). Orthologous outgroup sequence from the *V. monoica* shotgun sequenced database totaled ~51.94% (48356 bps) of the predicted X-Y paired coding sequence. Complete outgroup sequence was difficult to obtain due to the many gaps and ambiguous sequences in the whole-genome short gun sequences of *V. monoica*.

Six X-Y gene pairs in the oldest stratum of the SDR were used for population genetic analyses of coding sequence variation (Fig. 1 and Table S3). The focal Y-linked genes are located in one 0.8-Mb Y region surrounding the putative female suppressing sex determination gene (R. Ming, University of Illinois, unpublished). Their X homologs, among which three are predicted pseudogenes, are distributed across a 1.4-Mb region (Fig. 1b). Twenty-five autosomal loci dispersed across the genome of papaya were used to serve as genomic comparisons for the target sex-linked genes (Table S4). Neutral tests based on the site frequency spectrum suggested there are no signatures of selection on any of these autosomal genes (Table S4; Brown 2013), thus they were believed to be under neutral evolution and an ideal control group in this study.

### Comparative genetic analyses of all X-Y gene pairs

The codon substitution model (CODEML) in PAML (Yang 1997; Yang 2007) was used to discern the asymmetric protein evolutionary rates in X-Y gene pairs along the SDR. An unrooted phylogenetic tree with three branches consisting of X-Y gene pairs and their homologous sequences from *V. monoica* was used. A control model (model = 0) that assumes all branches evolve in a same ratio of nonsynonymous to synonymous substitutions ( $\omega_x = \omega_y = \omega_z$ ) was compared to a free-ratio model (model = 1) in which every branch has a unique ratio ( $\omega_x \neq \omega_y \neq \omega_z$ ). Significant differences between the two models were assessed using a likelihood ratio test.

DnaSp v5 (Librado and Rozas 2009) was used to estimate the number of synonymous or nonsynonymous substitutions between X/Y-linked locus (X+Y) as well as the substitution numbers of X/Y-linked locus relative to *V. monoica* (X+V, Y+V). The number of X-specific substitutions was then calculated by subtracting the total number of substitutions separating the Y-linked locus from *V. monoica* (Y+V) from the sum of the total substitutions separating the X-linked locus and *V. monoica* (X+V) and the number of substitutions separating the X/Y-linked locus (X+Y) and dividing the difference by 2 (i.e.,  $X = [(X+Y) + (X+V) - (Y+V)]/2$ ); Similarly, the number of Y-specific substitution was calculated by:  $Y = [(X+Y) + (Y+V) - (X+V)]/2$ . All the X-Y gene pairs were classified into three datasets based on their positions in different evolutionary strata. X/Y-specific substitutions were calculated for both synonymous sites and nonsynonymous sites from these data. Numbers of synonymous or nonsynonymous substitutions between X/Y-linked genes for each evolutionary stratum were compared using a paired t-test in JMP Pro 10 (SAS Institute Inc.).

Synonymous nucleotide divergence ( $K_s$ ) and nonsynonymous nucleotide divergence ( $K_a$ ) per site on each gene was calculated by DnaSp v5 using *V. monoica* as outgroup (Librado and Rozas 2009). To look for an association between the evolutionary rate of X/Y-linked genes and the time since recombination suppression, the  $K_a/K_s$  value of each X/Y-linked genes was plotted against the  $K_s$  values between X and Y. In addition, the  $K_a/K_s$  values based on the concatenated datasets of all X/Y-linked genes on stratum 1, stratum 2 and the collinear region were calculated and compared, which reflects the evolutionary rates of X/Y-linked genes in different evolutionary strata

### **Amplification and sequencing of six focal X/Y-linked loci**

Genomic DNA was isolated from stored leaf tissues using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA). PCR primers were designed by Primer3 (Rozen and Skaletsky 2000). Primers were placed in noncoding regions but adjacent to the boundary of target exon regions to amplify more coding sequences of the six focal X-Y paired genes from the papaya populations. Due to the limited range of cycle sequencing reaction (< 1200 bps), large focal loci were divided into smaller blocks for amplification (Table S5 and S6). In order to allow differential amplification of X/Y-linked genes, the primers were designed to be X-specific or Y-specific. Designed primers were synthesized by Integrated DNA Technologies. PCR was performed with GoTaq® Colorless Master Mix (Promega, Madison, WI). PCR reaction consisted of initial denaturation for 2 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s annealing at 56°C, 30 s at 72°C, and final extension for 5 min at 72°C. Products were confirmed by agarose gel electrophoresis. Single-band products were cleaned by using a combination of 5 U exonuclease I and 0.5 U shrimp alkaline phosphatase at 37°C for 40 min and inactivating the enzymes at 80°C for 15 min. Purified products were labeled through cycle sequencing using BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA), followed by a post-sequencing clean-up step using ethanol precipitation. Precipitated sequencing reactions were re-suspended in Hi-Di formamide before loading them into an ABI 3130xl or 3730 Genetic Analyzer (Applied Biosystems). Sequencing results were assembled and aligned using BioEdit and BioLign software (Hall 1999). Polymorphic sites were examined through visualizing the original ABI chromatogram file. Homologous sequences from outgroup species and protein coding sequences were added into alignments.

### **Population genetic analyses of six focal X/Y-linked loci**

Nucleotide diversity of the six focal loci were estimated using DnaSP v5 (Librado and Rozas 2009). Indel polymorphisms in the sequence alignment were included into data analysis while other gaps and ambiguous base pairs due to lower sequencing coverage or quality were excluded. After assigning coding regions of sequences and defining sequence sets into populations (natural papaya and outgroup), the level of polymorphism was calculated and reflected by two estimates of the population mutation parameter ( $\theta = 4N_e\mu$ ): theta ( $\theta_w$ ), which is based on the number of segregating sites (Watterson 1975), and pi ( $\pi$ ), pairwise nucleotide diversity (Tajima 1989).

Neutral tests based on the site frequency spectrum including Tajima's  $D$  (Tajima 1989) and Fay and Wu's  $H$  (Fay and Wu 2000), were performed in DnaSP v5 to examine whether inspected populations evolve under neutral expectations. The compound tests  $DH$  and  $DHEW$  (Zeng et al. 2007) combining individual tests based on site spectra ( $D$  and  $H$ ) and haplotype frequency spectrum ( $EW$ ) were implemented as well, to sensitively discern the pattern of positive selection. Linkage disequilibrium (LD) considering only informative sites on the target X/Y regions was analyzed in DnaSP v5 using concatenated sequence files. Two estimates of codon usage bias, the effective number of codons (ENC) and the codon bias index (CBI), were also calculated in DnaSP v5 for the X, Y and autosomal concatenated protein coding dataset, respectively.

Polymorphism patterns of silent sites on the six focal X-Y gene pairs were compared to 25 neutrally evolving autosomal loci using the Maximum Likelihood Hudson, Kreitman, and Aguade (MLHKA) test (Wright and Charlesworth 2004). The selection parameter  $k$  in the test, measures the excess ( $k > 1$ ) or deficiency ( $k < 1$ ) of polymorphism relative to neutral expectation. A selection model was used to estimate the  $k$  values of focal X/Y-linked genes by comparing polymorphism in the focal X/Y-linked genes with 25 autosomal loci ( $k = 1$ ). The differences between the selection model with a neutral model in which all loci including sex-linked and autosomal genes evolve neutrally ( $k = 1$ ), were assessed by the likelihood ratio test.

The number of synonymous and non-synonymous polymorphic and divergent substitutions in the exon regions of the focal loci was counted by DnaSP v5 (Librado and Rozas 2009). The ratio of nonsynonymous ( $P_A$ ) to synonymous ( $P_S$ ) polymorphisms as well as the ratio of amino-acid ( $D_A$ ) to synonymous ( $D_S$ ) divergences at the X/Y-linked loci were calculated based on the concatenated protein coding sequences of the six focal X/Y-linked genes. Similar data were drawn from concatenated coding sequences from 25 autosomal genes. The  $P_A/P_S$ ,  $D_A/D_S$  and  $(P_A/P_S)/(D_A/D_S)$  ratios were compared between X/Y-linked loci and autosomal loci.

## RESULTS

### Faster protein sequence evolution on the Y versus the X

In order to detect asymmetric protein evolution of X/Y-linked genes after X-Y divergence, the CODEML program in PAML was run using all the X-Y gene pairs and their homologs from the outgroup species *V. monoica* (Table S11). A free ratio model that allows  $\omega$  ( $d_N/d_S$ ) to vary in all branches of an unrooted tree of X, Y and *V. monoica* sequences was compared to a

constrained model in which each branch has the same  $\omega$ . In this test, few individual X-Y gene pairs showed a significant asymmetric evolutionary pattern. Lack of asymmetric protein evolution by tests on individual X-Y gene pairs may be due to 1) a paucity of segregating sites due to the recent divergence between X and Y chromosomes and the limited amounts of codons in individual genes used for analyses (Table S7), and 2) a lack of recurrent strong selective forces acting on individual X/Y genes. To address the first point and increase the segregating sites for analysis, all coding sequences were concatenated according to their positions on stratum 1, stratum 2 or the collinear region of the SDR and analyzed for patterns of asymmetric evolution by stratum (Table 1). Stratum 1 significantly fits the free-ratio model over the constrained model ( $P = 0.009$ ), showing an elevated  $\omega$  on the Y ( $\omega = 0.4078$ ) relative to its X counterpart ( $\omega = 0.264$ ). This indicates the faster protein evolution on the Y relative to the X in stratum 1. However, there is no strong signature of asymmetric pattern of protein evolution found in stratum 2 ( $P = 0.069$ ) or the collinear region ( $P = 0.139$ ).

To distinguish whether an excess of nonsynonymous mutations drives the higher  $d_N/d_S$  ratio on the Y relative to the X, a matched pair analysis was implemented to compare the number of synonymous and nonsynonymous substitutions at Y-linked loci versus X-linked loci. Across the entire SDR region, synonymous mutations are similar between X and Y-linked loci ( $P = 0.1244$ ; Fig. 2a). However, there is a significant excess of amino acid replacement mutations at Y-linked loci compared to X-linked loci ( $P = 0.0016$ ; Fig. 2b). The excess of nonsynonymous substitutions is localized in stratum 1 of the SDR region ( $P = 0.0024$ ; Fig. 2d). No significant differences in the distribution of nonsynonymous substitutions were found in stratum 2 ( $P = 0.0898$ ; Fig. 2f) or the collinear region ( $P = 0.9619$ ; Fig. 2h). These results parallel the PAML branch-site results, indicating that the young Y chromosome of papaya has begun to degenerate, primarily in the oldest evolutionary stratum.

Comparing the  $K_a/K_s$  values of X/Y-linked genes in different strata relative to their orthologs in *V. monoica* enables us to assess the evolutionary tempo of degeneration after the events of recombination suppression. The strata-specific concatenated datasets of coding sequences was used to summarize the  $K_a/K_s$  values of loci found in each stratum.  $K_a/K_s$  values for both the X/Y-specific regions increase with the age of strata (Fig. 3a).  $K_a/K_s$  values are almost the same between X and Y in the collinear region and stratum 2, but higher on the Y in stratum 1. The result indicates the rates of protein evolution of both the X/Y-linked genes are positively

correlated with the age of strata and only Y-linked loci in the oldest stratum have higher rates than their X-homologs. A plot of  $K_a/K_s$  versus  $K_s$  for each individual sex-linked gene was also conducted to assess the association between the degree of functional constraint and time since recombination suppression between each X/Y-linked locus (Fig. 3b). A significant positive correlation between  $K_a/K_s$  and  $K_s$  is found on the Y ( $P = 0.031$ ), but not on the X ( $P = 0.450$ ). This pattern again shows the Y evolves faster than the X and the degree of functional constraint at the Y-linked loci is positively correlated to the divergence time from their X homologs.

### **A strong hitchhiking effect on the target X region but not on the target Y region**

In order to further explore the evolutionary mechanisms on the oldest stratum of the SDR where the strongest degeneration pattern of the MSY is exhibited, population analyses on six X-Y genes pairs around a region in stratum 1 were conducted. As with a previous study, contrasting patterns of nucleotide diversity were found on X- and Y-linked loci (Table 2; Weingartner and Moore 2012). Two estimates of the population mutation parameter  $\pi$  and  $\theta_w$  reflect the genetic diversity on the X and Y. The average  $\pi_{\text{sil}}$  is  $0.00091 \pm 0.00033$  (SE) for the Y-linked loci, and  $0.00034 \pm 0.00015$  (SE) for the X-linked loci (Table 2). The average  $\theta_w$  values at silent sites are  $0.00084 \pm 0.00027$  (SE) for the Y-linked loci, and  $0.00093 \pm 0.00042$  (SE) for the X-linked loci (Table 2). Both  $\pi$  and  $\theta_w$  show an elevated polymorphism pattern on the target Y or reduced nucleotide variation on the corresponding X relative to the neutral diversity (X/Y=3) based on the effective population size.

Based on Tajima's  $D$  test, all the X-linked genes have negative  $D$  values, consistent with excesses of rare alleles. Significant negative  $H$  values of all the X-linked genes further support a signature of selective sweep on the X. According to the compound  $DH$  and  $DHEW$  test, the X-linked loci vary significantly from neutral expectations, except for  $CpXYh14X$  and  $PXCpXY^h6X$  which are limited by the number of polymorphic sites. However, for all the Y-linked loci, both  $D$  and  $H$  values tend to be near zero, indicating no strong signature of positive selection. The MLHKA test was performed on the silent sites of the focal X/Y genes to assess whether there are skewed polymorphism levels relative to divergence (Table 13). For all the Y-linked genes, the selection coefficient  $k$  is larger than 1, indicating a relative high polymorphism relative to divergence (Table S13). Among them, this pattern is significant in  $CpXY^h13Y^h$  ( $P = 0.0014$ ) and marginally significant in  $PXCpXY^h6Y^h$  ( $P = 0.063$ ). In contrast, the

$k$  values for most of X loci are near one, except for one predicted pseudogene *PXCpXY<sup>h</sup>10X*, which shows a significant elevated polymorphism ( $P = 0.0022$ ; Table S13). The concatenated dataset of Y has a significant  $k$  value of 2.79 ( $P = 0.0046$ ) while the X-concatenated data set has a non-significant  $k$  of 1.70 ( $P = 0.25$ ; Table S13). The significant large  $k$  value on the Y supports an increased nucleotide variation on the Y than expected.

The signature of a selective sweep drives allele structure in nearly all the segregating sites of the six focal X loci across the target X region. Strong LD exists across the entire 1.4-Mb region of the X chromosome (Fig. 4a). This strong LD may be explained by a selective sweep on the unusual genomic background of this target X region. This region is very close to the corresponding X region of the Y-specific heterochromatic Knob 4 (Fig. 1a). As Knob 4 of the MSY is assumed to be either associated with or adjacent to the centromere (Zhang et al. 2008) where recombination is extremely reduced (Yu et al. 2009), the target X region should be under a very low recombination rate. Strong LD also exists across the 0.8-Mb target region of Y, consistent with no recombination in the MSY (Fig. 4b).

### **Elevated nonsynonymous polymorphism and divergence on both the target X and Y regions**

To assess the polymorphic pattern of relaxed purifying selection on the focal X/Y loci, the  $\pi_{(a)}/\pi_{(s)}$  ratios based on the concatenated protein coding dataset of focal X/Y loci were calculated and compared with the value in autosomal coding sequences (Fig. 5). Since nonsynonymous sites are more selectively constrained,  $\pi_{(a)}/\pi_{(s)}$  value is expected to be much lower than 1. However, for the concatenated protein coding dataset of Y-linked loci, the  $\pi_{(a)}/\pi_{(s)}$  value is 0.84, close to 1 and much higher than the  $\pi_{(a)}/\pi_{(s)}$  value (0.17) in the autosomal genes. The  $\pi_{(a)}/\pi_{(s)}$  value for the concatenated X protein coding sequence is 0.31, and though less than the Y, is elevated relative to the autosomal loci.

To compare polymorphism and divergence pattern in coding sequence of the six focal X-Y gene pairs with autosomal loci, the number of amino-acid and synonymous mutations were tabulated within papaya populations and between papaya and two similarly related species, *V. monoica* or *J. dolichaula* (Table 4). Due to limited polymorphic sites in individual loci, this test was performed using the concatenated dataset of the coding sequences from the six focal X/Y loci and compared with the dataset of autosomal coding sequences. An excess of

nonsynonymous mutations at both divergence and polymorphism level was found on the focal X/Y-linked loci relative to the autosomal loci (Table 4). The  $D_A/D_S$  on the Y (1.32) and the X (1.29) are significantly greater than that on the autosomes (0.60). The three X predicted pseudogenes (1.31) just show slightly higher  $D_A/D_S$  ratio than the three functional X loci (1.27). The  $P_A/P_S$  is similarly elevated on the Y (1.75) and the X (2.00), relative to the autosome (0.45). The three predicted pseudo X-linked genes (2.50) exhibit a higher  $P_A/P_S$  ratio than the functional X loci (1.67). While the pattern of excess of nonsynonymous substitutions relative to polymorphisms ( $P_A/P_S < D_A/D_S$ ) is observed in autosomal loci, the X/Y-linked loci show more nonsynonymous polymorphisms relative to divergence ( $P_A/P_S > D_A/D_S$ ). This is likely to be caused by relaxed purifying selection on the X/Y-specific regions to purge new deleterious nonsynonymous mutations efficiently and thus  $P_A/P_S$  is elevated to be higher than  $D_A/D_S$ .

As relaxed purifying selection can also lead to reduced codon bias, two measures of codon bias were estimated from the concatenated datasets (2801 Y-linked codons, 2618 X-linked codons, and 1298 autosomal codons). The effective number of codons (ENC) is greatest for focal Y-linked loci (51.6), intermediate for X-linked loci (51.0) and lowest for autosomal loci (50.6). The codon bias index (CBI), which ranges from 0 (weak bias) to 1 (strong bias), is lowest for Y-linked loci (0.273), intermediate for X-linked loci (0.290) and greatest for autosomal loci (0.321).

A previous analysis of pseudogene composition found there are 24 (25% of 96) predicted pseudogenes on the MSY and 14 (14% of 98) predicted pseudogenes in the corresponding X region (Wang et al. 2012). There are slightly significant higher proportions of pseudogenes on the Y than on the X (Fisher's exact test;  $P=0.044$  Table S14). Furthermore, there are a significantly greater proportion of pseudogenes (regardless of sex-linkage) in the oldest stratum 1 than in either stratum 2 ( $P=0.048$ ) or the collinear region ( $P=0.003$ ). In the target 1.4-Mb X region, there are 16 X-specific and five X-Y paired genes and seven out of these 21 X-linked genes (33%) are predicted pseudogenes. By comparison to the entire X/Y-specific regions (14% and 25%, respectively), the pseudogenes percentage is higher on the target X region (33%).

## **DISCUSSION**

### **Evolutionary tempo of Y chromosomal degradation in papaya**

The Y chromosome tends to accumulate more deleterious mutations and has a relatively



slower adaptive evolutionary rate than most of recombining region of the genome due to the lack of recombination (Bachtrog 2013). In animal species, like *Drosophila miranda*, the degeneration of the neo-Y chromosome happened very rapidly. The 1 MY old neo-Y has lost almost half of its previously functional genes; either these genes have become pseudogenes or they are completely missing (Bachtrog et al. 2008). Degeneration of the young Y chromosome also happens in plant species, as described in *Silene latifolia* (Marais et al. 2008). However, the degeneration progress seems to be much slower. At most 20% of the genes on this 10 MY old Y chromosome have lost their functions (Bergero and Charlesworth 2011; Chibalina and Filatov 2011). Considering the shorter generation time in *Drosophila miranda*, a more accurate evaluation of age of Y chromosome should be based on the rates of synonymous protein divergence between X- and Y-linked genes (Bachtrog 2011). According to the synonymous protein X-Y divergence in *S. latifolia* (2.5%-10%) and *D. miranda* (1-7%), the Y chromosome in *S. latifolia* is still slightly older than that in *D. miranda* (Bachtrog 2011). It was proposed that the retarded degeneration of Y-linked genes in *S. latifolia* is due to the expression of Y-linked alleles during the haploid gametophyte phase found in plants that results in stronger selective pressure on detrimental mutations (Chibalina and Filatov 2011).

Papaya is another good model to study the degeneration of early Y chromosome in plants. Its non-recombining SDR consists of two evolutionary strata formed at about 7.0 and 1.9 MYA (Avg.  $K_s = 9.4\% \pm 1.1\%$  and  $3.1\% \pm 0.7\%$ , respectively) and one collinear region (Avg.  $K_s = \sim 0.7\% \pm 0.2\%$ ) probably due to recent involvements of adjacent sexually antagonistic mutations (Wang et al. 2012). According to the synonymous protein X-Y divergence, the age of sex chromosomes in papaya is comparable to that of *S. latifolia*. The Y chromosome potentially has begun to degenerate, as evidenced by nonviable YY plants (Ming et al. 2007), increased abundance of transposable elements and pseudogenes on the Y chromosome (Wang et al. 2012). Our test found evidence of faster amino acid replacements and reduced codon bias on the Y than the X, consistent with an accumulation of unpreferred codons for translation. Although this difference can be also explained by higher per-generation mutation rate in males than females (Filatov and Charlesworth 2002), the comparison of synonymous mutations between X and Y (Fig. 2a) shows there is no significant difference in substitution rates. But this asymmetric nonsynonymous mutations pattern is confined to the oldest evolutionary stratum, suggesting the decay of Y-linked genes may have happened between the formations of the two strata (1.9~7.0 MYA;  $K_s$ : 3.1-9.4%).

Based on that, the tempo of Y chromosomal degeneration in papaya is relatively slow, consistent with the retarded Y degeneration of *S. latifolia* experiencing haploid purifying selection.

The relatively slow degeneration tempo of the Y chromosome in papaya may be also explained by other factors. Beneficial mutations, such as masculinization of Y genes, simultaneously fix the linked deleterious mutations on the neo-Y chromosome of *D. miranda* (Zhou and Bachtrog 2012). But a strong hitchhiking effect in the non-recombining region only happens in the case that the selective advantage of a beneficial mutation was strong enough to outweigh the selective disadvantage of its linked deleterious mutations (Bachtrog 2013). Until now, we have no evidence showing a strong selective sweep on the MSY region (Weingartner and Moore 2012). Also, the tempo of Y-chromosome degeneration is negatively correlated with the number of functional genes (Bachtrog 2008; Bachtrog 2013). Compared to the large gene-rich neo-Y chromosome in *D. miranda*, the MSY of papaya is a small pericentromeric region which starts with the characters of reduced recombination rate and low gene density. Thus, the young Y chromosome in papaya may act as other old Y chromosomes containing a small number of active genes and degenerate at a slow pace. Furthermore, despite of few genes decayed found in this study based on the number of pseudogenes and nonsynonymous substitution, the possibility that degeneration has already been initiated through reduced gene expression cannot be excluded. In the neo-Y chromosome of *Drosophila albomicans* that originated ~0.1 MYA, ~30% of Y-linked genes have been down-regulated though less than 2% of the genes have lost their functions due to accumulation of deleterious amino acid replacement mutations (Zhou et al. 2012). Formation of heterochromatin and accumulation of mutations in intronic regulatory region or promoter sequences can both lead to gene silencing (Bachtrog 2013). As the MSY region of papaya has harbored much more repetitive sequences than its X counterpart (Wang et al. 2012), change in genomic structure may influence expression of Y-linked genes. Thus, a clearer degeneration pattern may be detected at the level of gene expression.

### **Different selective forces drive the X and Y evolution**

The centromere is indicated as a recombinational cold spot associated with wide-ranging eukaryotic genomes (Choo 1998). Recombination is notably reduced at the centromeric region of all nine major linkage groups in papaya (Chen et al. 2007; Yu et al. 2009). The broad LD interval

on the X- and Y-specific regions would enhance the efficacy of various selective forces. For example, hitchhiking effect of a strong positive selection is enhanced on the pericentromeric X-specific region with low recombination, which can explain the significant linkage disequilibrium across the target 1.4-Mb X region. However, it also makes it difficult to distinguish whether there is any gene under causal positive selection. This common haplotype may expand far more than this 1.4 Mbps, containing more than 20 genes. What we know at present is at least one gene is under intense positive selection presumably for compensatory function of X alleles for degenerating Y-linked loci, causing widespread hitchhiking effect on its linked flanking region. On the other hand, all focal loci in the target Y region maintained elevated polymorphism relative to neutral expectation based on the results from MLHKA test. The elevated polymorphism could be probably explained by frequency-dependent selection on one or more loci in the MSY (e.g. polymorphisms for Y-linked suppressors of X meiotic drive; Weingartner and Moore 2012). Due to the recombination suppression of the MSY, the magnitude of frequency-dependent selection may also have been extended to the entire MSY region.

### **Relaxed selective constraint on the focal X loci**

Although there are no evident cues of selection on any focal X/Y genes based on various neutral tests explaining the observed polymorphism pattern, elevated  $\pi_{(a)}/\pi_{(s)}$  and A/S ratios, excess of nonsynonymous polymorphism relative to divergence, and reduced codon bias in X/Y coding sequences relative to autosomal coding regions, suggest there are more deleterious mutations accumulating on both the target X and Y regions relative to neutrally evolving autosomal loci. We also found a positive correlation between the degree of functional constraint and the age of strata of the X-specific region based on the concatenated data set (Fig. 5a). It is interesting to find the signature of degeneration on the X-specific region, which still recombines. If it is true, degeneration of the MSY relative to the X would be masked to some extent. However, the  $K_a/K_s$  versus  $K_s$  based on each individual sex-linked gene does not show similar patterns (Fig. 5b). This may be explained by different selective constraints on individual genes with various functions. Also, more pseudogenes in older stratum, especially the target X region adjacent to centromere, suggest that gene decay in the X region is expedited after the recombination suppression between the X and Y.

The X chromosome of papaya has not conserved the structure and gene content of its

ancestral autosome. Compared with the orthologous region in *V. monoica*, the X-specific region has expanded in size by transposable element accumulation, and genes on it have been lost and added (Gschwend et al. 2012). The remarkable changes in the structure and gene content of the X-specific region of papaya may be explained by its relatively low recombination frequency (Wang et al. 2012). Recombination suppression between X and Y forces the crossover in the X-specific region to be restricted to XX females. Thus, the recombination frequency of the X-specific region is theoretically two thirds that of autosomes. In addition, the X-specific region in papaya spans the centromere and most of the X-specific region lay in the pericentromeric region due to its small size. Recombination rate in this unusual pericentromeric region between two X chromosomes in female is also low, as evidenced by comparative genetic and physical maps (Yu et al. 2009). Thus, the X-specific region in papaya may be evolving like other genomic regions with low recombination. Reduced natural selection is associated with regions with low recombination (Betancourt and Presgraves 2002). Low codon bias present in the region near the centromeres, telomere or in the dot chromosome, was interpreted as relaxed selection for optimized translation (Kliman and Hey 1993). The genes located in the heterochromatin of chromosomes, especially the ones most proximal to the centromeres, exhibit an accelerated rate of protein sequence evolution and lower codon usage bias, implications of relaxed purifying selection (Campos et al. 2012). Because of its small size, the dot chromosome in *Drosophila*, lies near the recombination-suppressing centromere, and suffers enhanced Hill-Robertson interference effect, contributing to the accumulation of deleterious mutations (Haddrill et al. 2007; Charlesworth et al. 2009; Arguello et al. 2010).

Furthermore, the degeneration signature of the focal X loci may be explained by the strong selective sweep. When the effect of a beneficial gene outweighs the impact of linked detrimental alleles, those slightly deleterious alleles are likely to be dragged along (Birky and Walsh 1988; Hartfield and Otto 2011). Accumulation of deleterious mutations due to a hitchhiking effect is an important force contributing to the degeneration of some Y chromosomes (Charlesworth 1996; Bachtrog 2004; Zhou and Bachtrog 2012). It also explains the increased number of deleterious mutations in the genomes of domesticated species under strong artificial selection (Lu et al. 2006; Cruz et al. 2008) and the distribution of detrimental disease-associated alleles in human genome (Chun and Fay 2011). In this study, we found a strong selective sweep on the target X region. Due to the low recombination rate at this pericentromeric region, the LD caused by the

hitchhiking effect is expanded to at least 1.4-Mb region. Thus, many detrimental alleles are probably fixed under the widespread hitchhiking effect, which explains the degeneration pattern at this target X region.

## **CONCLUSION**

The SDR of papaya with two evolutionary strata and one collinear region is a good model to study the evolutionary dynamics of young sex chromosomes. Here we report a degenerating Y chromosome in papaya. The degeneration pattern of the Y relative to the X is confined to the oldest stratum of the SDR, suggesting a relatively slow tempo of degeneration. However, the papaya X-specific region similarly suffers accumulation of deleterious mutations after recombination suppression, evident in a 1.4-Mb region of stratum 1 close to the centromere. This may be explained by weaker purifying selection due to the Hill-Roberson interference effect and a widespread hitchhiking effect on the pericentromeric region with low recombination. If that is true, the degeneration tempo of the Y chromosome may be masked to some extent. Further investigation on the expression level of X/Y-linked genes, relating to their biological functions and rates of accumulation of nonsynonymous substitutions, may show a clearer pattern of degeneration on the X/Y chromosomes in papaya. This knowledge will ultimately enhance the understanding of evolutionary process of young sex chromosomes, especially for plant systems which are likely exposed to haploid-phase selection.

## LITERATURE CITED

- Arguello JR, Zhang Y, Kado T, Fan C, Zhao R, Innan H, Wang W, Long M. 2010. Recombination yet inefficient selection along the *Drosophila melanogaster* subgroup's fourth chromosome. *Mol Biol Evol.* 27:848–861.
- Arumuganathan K, Earle ED. 1991. Nuclear DNA content of some important plant species. *Plant Mol Biol Rep.* 9:208–18.
- Bachtrog D, Charlesworth B. 2002. Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* 416:323–326.
- Bachtrog D. 2004. Evidence that positive selection drives Y-chromosome degeneration in *Drosophila miranda*. *Nature* 36:518–522.
- Bachtrog D. 2006. A dynamic view of sex chromosome evolution. *Curr Opin Genet Dev.* 16: 578-585.
- Bachtrog D, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice W, Valenzuela N. 2011. Are all sex chromosomes created equal? *Trends Genet.* 27: 350-357.
- Bachtrog D. 2008 The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics* 179:1513–1525.
- Bachtrog D, Hom E, Wong KM, Maside X, de Jong P. 2008. Genomic degradation of a young Y chromosome in *Drosophila miranda*. *Genome Biol.* 9:R30.
- Bachtrog D, Jensen JD, Zhang Z. 2009. Accelerated adaptive evolution on a newly formed X chromosome. *PLoS Biol.* 7:712–719.
- Bachtrog D. 2011. Plant sex chromosomes: a non-degenerated Y? *Curr Biol.* 21:685–688.
- Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet.* 14:113–124.
- Bergero R, Forrest A, Kamau E, Charlesworth D. 2007. Evolutionary strata on the X chromosomes of the dioecious plant *Silene latifolia*: evidence from new sex-linked genes. *Genetics* 175:1945–1954.
- Bergero R, Charlesworth D. 2009. The evolution of restricted recombination in sex chromosomes. *Trends Ecol Evol.* 24:94–102.
- Bergero R, Charlesworth D. 2011. Preservation of the Y transcriptome in a 10-million-year-old plant sex chromosome system. *Curr Biol.* 21:1470–1474.
- Betancourt AJ, Welch JJ, Charlesworth B. 2009. Reduced effectiveness of selection caused by a

- lack of recombination. *Curr Biol.* 19:655–660.
- Birky CWJ, Walsh JB. 1988. Effects of linkage on rates of molecular evolution. *Proc Natl Acad Sci U S A.* 85:6414–6418.
- Brown JE, Bauman JM, Lawrie JF, Rocha OJ, Moore RC. 2012. The structure of morphological and genetic diversity in natural populations of *Carica papaya* (*Caricaceae*) in Costa Rica. *Biotropica* 44:179–188.
- Brown JE. 2013. The evolutionary mechanisms promoting sex chromosome divergence within *Carica papaya*. Doctor's Thesis. Miami University, Oxford, Ohio.
- Bull JJ. 1983. *Evolution of sex determining mechanisms*. Benjamin Cummings, Menlo Park, CA.
- Charlesworth B. 1991. The evolution of sex chromosomes. *Science* 261:1030-1033.
- Charlesworth B. 1996. The evolution of chromosomal sex determination and dosage compensation. *Curr Biol.* 6: 149-162.
- Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. *Nat Rev Genet.* 10:195–205.
- Charlesworth B, Betancourt AJ, Kaiser VB, Gordo I. 2009. Genetic recombination and molecular evolution. *Cold Spring Harb Symp Quant Biol.*74:177–186.
- Charlesworth D. 2002. Plant sex determination and sex chromosomes. *Heredity* 88: 94–101.
- Charlesworth D and Charlesworth B. 1987. Inbreeding depression and its evolutionary consequences. *Ann Rev Ecol Syst.* 18:237-268.
- Charlesworth D. 2013. Plant sex chromosome evolution. *J Exp Bot.* 64:405–420.
- Chen C, Yu Q, Hou S, et al. 2007. Construction of a sequence-tagged high-density genetic map of papaya for comparative structural and evolutionary genomics in *Brassicales*. *Genetics* 177:2481–2491.
- Chibalina MB, Filatov DA. 2011. Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr Biol.* 21:1475–1479.
- Chun S, Fay JC. 2011. Evidence for hitchhiking of deleterious mutations within the human genome. *PLoS Genet.* 7:e1002240.
- Choo KH. 1998. Why is the centromere so cold? *Genome Res.* 8:81–82.
- Comeron JM, Williford A, Kliman RM. 2008. The Hill-Robertson effect: Evolutionary consequences of weak selection and linkage in finite populations. *Heredity* 100:19–31.

- Compos JL, Charlesworth B, Haddrill PR. 2012. Molecular evolution in nonrecombining regions of the *Drosophila melanogaster* genome. *Genome Biol Evol.* 4:278–288.
- Cruz F, Vilà C, Webster MT. 2008. The legacy of domestication: accumulation of deleterious mutations in the dog genome. *Mol Biol Evol.* 25:2331–2336.
- Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413.
- Filatov DA, Charlesworth D. 2002 Substitution rates in the X- and Y-linked genes of the plants, *Silene latifolia* and *S. dioica*. *Mol Biol Evol.* 19:898–907.
- Gilbert SF. 2000. *Environmental sex determination*. Sinauer Associates, In. Publishers, Sunderland, Massachusetts, USA.
- Gordo I, Charlesworth B. 2001. Genetic linkage and molecular evolution. *Curr Biol.* 11, R684-R686.
- Grimaldi D, Engel MS. 2005. *The evolution of the insects*. Cambridge University Press, Cambridge.
- Gschwend AR, Yu Q, Tong EJ, et al (11 co-authors). 2012. Rapid divergence and expansion of the X chromosome in papaya. *Proc Natl Acad Sci U S A.* 109:13716–13721.
- Haddrill PR, Halligan DL, Tomaras D, Charlesworth B. 2007. Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. *Genome Biol.* 8:R18.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program from Windows 95/98/NT. *Nucleic Acids Symp Ser.* 41:95–98.
- Hartfield M, Otto SP. 2011. Recombination and hitchhiking of deleterious alleles. *Evolution* 65:2421–2434.
- Hill WG, Robertson A. 1966. The effect of linkage on limits to artificial selection. *Genet Res.* 8:269-294.
- Kitano J, Peichel CL. 2012. Turnover of sex chromosomes and speciation in fishes. *Environ Biol Fish.* 94: 549-558.
- Kliman RM, Hey J. 1993. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol Biol Evol.* 10:1239–1258.
- Lahn BT, Page DC. 1999. Four evolutionary strata on the human X chromosome. *Science* 286:964–967.
- Lawson-Handley LJ, Ceplitis H, Ellegren H. 2004. Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. *Genetics* 167:367–376.



- Liew WC, Bartfai R, Lim Z, Sreenivasan R, Siegfried KR, Orban L. 2012. Polygenic sex determination system in zebrafish. *PLoS One* 7:e34397.
- Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
- Liu Z, Moore PH, Ma H, et al. 2004. A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* 427:348–352.
- Lu J, Tang T, Tang H, et al. 2006. The accumulation of deleterious mutations in rice genomes: a hypothesis on the cost of domestication. *Trends Genet.* 22:126–131.
- Manning JT, Thompson DJ. 1984. Muller's ratchet and the accumulation of favorable mutations. *Acta Biotheoretica* 33: 219-225.
- Marais GA, Nicolas M, Bergero R, Chambrier P, Kejnovsky E, Moneger F, Hobza R, Widmer A, Charlesworth D. 2008. Evidence for degeneration of the Y chromosome in the dioecious plant *Silene latifolia*. *Curr Biol.* 18:545–549.
- McDaniel SF, Neubig KM., Payton AC, Quatrano RS, Cove DJ. 2013. Recent gene-capture on the UV sex chromosomes of the moss *Ceratodon purpureus*. *Evolution* 67:2811–2822.
- Ming R, Yu Q, Moore PH. 2007a. Sex determination in papaya. *Semin Cell Dev Biol.* 18:401–408.
- Ming R, Wang J, Moore PH, Paterson AH. 2007b. Sex chromosomes in flowering plants. *Am J Bot.* 94: 141-150.
- Ming R, Hou S, Feng Y, et al. 2008. The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya Linnaeus*). *Nature* 452:991–996.
- Ming R, Bendahmane A, Renner SS. 2011. Sex chromosomes in land plants. *Annu Rev Plant Biol.* 62:485–514.
- Moore RC, Kozyreva O, Lebel-Hardenack S, Siroky J, Hobza R, Vyskot B, Grant SR. 2003. Genetic and functional analysis of DD44, a sex-linked gene from the dioecious plant *Silene latifolia*, provides clues to early events in sex chromosome evolution. *Genetics* 163:321–334.
- Muyle A, Zemp N, Deschamps C, Mousset S, Widmer A, Marais GA. 2012. Rapid de novo evolution of X chromosome dosage compensation in *Silene latifolia*, a plant with young sex chromosomes. *PLoS Biol.* 10, e1001308.
- Nam K, Ellegren H. 2008. The chicken (*Gallus gallus*) Z chromosome contains at least three

- nonlinear evolutionary strata. *Genetics* 180:1131–1136.
- Nur U. 1990. Heterochromatization and euchromatization of whole genomes in scale insects (*Coccoidea: Homoptera*). *Development* 108: 29–34.
- Peichel CL, Ross JA, Matson CK, Dickson M, Grimwood J, Schmutz J, Myers RM, Mori S, Schluter D, Kingsley DM. 2004. The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Curr Biol.* 14:1416–1424.
- Rice WR. 1996. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* 41:911–914.
- Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol.* 132:365–386.
- Sachidanandam R, Weissman D, Schmidt SC, et al. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* 409: 928-933.
- Sandstedt SA, Tucker PK. 2004. Evolutionary strata on the mouse X chromosome correspond to strata on the human X chromosome. *Genome Res.* 14:267–272.
- Schaffner SF. 2004. The X chromosome in population genetics. *Nat Rev Genet.* 5:43-51.
- Skaletsky H, Kuroda-Kawaguchi T, Minx P, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423:825–837.
- Smith JM. 1978. *The evolution of sex*. Cambridge University Press, Cambridge.
- Spigler RB, Lewers KS, Main DS, Ashman TL. 2008. Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome. *Heredity* 101:507–517.
- Steinemann M, Steinemann S. 1998. Enigma of Y chromosome degeneration: neo-Y and neo-X chromosomes of *Drosophila miranda* a model for sex chromosome evolution. *Genetica* 102–103:409–420.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Vicoso B, Emerson JJ, Zektser Y, Mahajan S, Bachtrog D. 2013. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol.* 11:e1001643.
- Votintseva AA, Filatov DA. 2009. Evolutionary strata in a small mating-type-specific region of

- the smut fungus *Microbotryum violaceum*. *Genetics* 182:1391–1396.
- Wang J, Na J, Yu Q, et al. 2012. Sequencing papaya X and Y<sup>h</sup> chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc Natl Acad Sci U S A*. 109:13710–13715.
- Watterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol*. 7:256-276.
- Weingartner LA, Moore RC. 2012. Contrasting patterns of X/Y polymorphism distinguish *Carica papaya* from other sex chromosome systems. *Mol Biol Evol*. 29:3909–3920.
- Werren JH, Beukeboom LW. 1998. Sex determination, sex ratios, and genetic conflict. *Annu Rev Ecol Syst*. 29:223-261.
- Wilson MA, Makova KD. 2009. Genomic analyses of sex chromosome evolution. *Annu Rev Genom Hum Genet*. 10: 333-354.
- Wright SI, Charlesworth B. 2004. The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. *Genetics* 168:1071–1076.
- Wu X, Wang J, Na J, et al. 2010. The origin of the non-recombining region of sex chromosomes in *Carica* and *Vasconcellea*. *Plant J*. 63:801–810.
- Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci*. 13:555–556.
- Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 24:1586–1591.
- Yin T, DiFazio SP, Gunter LE, et al. 2008. Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. *Genome Res*. 18:422–430.
- Yu Q, Hou S, Feltus FA, et al. 2008. Low X/Y divergence in four pairs of papaya sex-linked genes. *Plant J*. 53:124–132.
- Yu Q, Tong E, Skelton RL, et al. 2009. A physical map of the papaya genome with integrated genetic map and genome sequence. *BMC Genomics* 10:371.
- Zeng K, Shi S, Wut CI. 2007. Compound tests for the detection of hitchhiking under positive selection. *Mol Biol Evol*. 24:1898–1908.
- Zhang W, Wang X, Yu Q, Ming R, Jiang J. 2008. DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya. *Genome Res*. 18:1938–1943.

Zhou Q, Bachtrog D. 2012. Sex-specific adaptation drives early sex chromosome evolution in *Drosophila*. *Science* 337:341–345.

Zhou Q, Zhu H, Huang Q, et al. 2012. Deciphering neo-sex and B chromosome evolution by the draft genome of *Drosophila albomicans*. *BMC Genomics* 13:109

## TABLES

**Table 1.** PAML-based tests of evolutionary rate differences between X-linked and Y-linked genes in different strata (with pseudogenes), using homologous genes of *V. monoica* as outgroup sequences

Stratum	Number of codons	Model	LnL <sup>a</sup>	$\chi^2$	df	P	free d <sub>N</sub> /d <sub>S</sub> <sup>b</sup>
SDR	15178	Model = 0	-77094.936	34.84	2	<< <b>0.001</b>	X: 0.2958
		Model = 1	-77077.514				Y: 0.4222
1	6833	Model = 0	-35902.063	9.327	2	<b>0.009</b>	X: 0.2640
		Model = 1	-35897.399				Y: 0.4078
2	5092	Model = 0	-25245.509	5.333	2	0.069	X: 0.3101
		Model = 1	-25242.842				Y: 0.3489
Collinear	4115	Model = 0	-19783.215	3.945	2	0.139	X: 0.6176
		Model = 1	-19781.243				Y: 0.2414

<sup>a</sup>LnL, log likelihood; <sup>b</sup>free d<sub>N</sub>/d<sub>S</sub>, the ratio of nonsynonymous/synonymous substitution rates in each branch under model 2

**Table 2.** Summary of nucleotide diversity for focal loci

Gene	No. of sites			$\pi^a$			$\theta^b$		
	Sil	Syn	Nonsyn	Sil	Syn	Nonsyn	Sil	Syn	Nonsyn
<i>Y-polymorphism</i>									
<i>CpXY<sup>h</sup>8Y<sup>h</sup></i>	2231	688	2330	0.00051	0.00126	0.00048	0.00042	0.00072	0.00030
<i>CpXY<sup>h</sup>14Y<sup>h</sup></i>	1466	287	871	0.00048	0.00099	0.00038	0.00063	0.00081	0.00053
<i>PXCpXY<sup>h</sup>14Y<sup>h</sup></i>	2407	401	1426	0.00035	0	0.00034	0.00048	0	0.00049
<i>PXCpXY<sup>h</sup>10Y<sup>h</sup></i>	2562	254	907	0.00059	0.00037	0.00101	0.00063	0.00091	0.00077
<i>CpXY<sup>h</sup>13Y<sup>h</sup></i>	643	65	244	0.00251	0	0	0.00217	0	0
<i>PXCpXY<sup>h</sup>6Y<sup>h</sup></i>	1939	220	749	0.00097	0	0	0.00072	0	0
<b>Average</b>	<b>1876</b>	<b>319</b>	<b>1071</b>	<b>0.00091</b>	<b>0.00044</b>	<b>0.00037</b>	<b>0.00084</b>	<b>0.00041</b>	<b>0.00035</b>
<i>X-polymorphism</i>									
<i>CpXY<sup>h</sup>8X</i>	2245	643	2225	0.00044	0.00038	0.00017	0.00120	0.00105	0.00051
<i>CpXY<sup>h</sup>14X</i>	1456	292	887	0.00006	0	0	0.00015	0	0
<i>PXCpXY<sup>h</sup>14X</i>	2125	430	1526	0.00035	0.00057	0.00032	0.00095	0.00157	0.00089
<i>PXCpXY<sup>h</sup>10X</i>	2777	245	877	0.00084	0.00067	0	0.00235	0.00184	0
<i>CpXY<sup>h</sup>13X</i>	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
<i>PXCpXY<sup>h</sup>6X</i>	1226	132	477	0	0	0	0	0	0
<b>Average</b>	<b>1966</b>	<b>348</b>	<b>1199</b>	<b>0.00034</b>	<b>0.00032</b>	<b>0.00010</b>	<b>0.00093</b>	<b>0.00089</b>	<b>0.00028</b>

n.d., no determined data (*CpXY<sup>h</sup>13X* fails to be amplified); Syn, synonymous sites; Nonsyn, nonsynonymous sites; Sil, silent sites; <sup>a</sup> $\pi$  pairwise nucleotide diversity; <sup>b</sup> $\theta$  Nucleotide site variability based on the number of segregating sites

**Table 3.** Tests of selection based on the site frequency spectrum ( $D$  and  $H$ ), haplotype frequency spectrum ( $EW$ ), or combinations thereof ( $DH$  and  $DHEW$ ).

Locus	n <sup>a</sup>	S <sup>b</sup>	$D$	$D$ (NS <sup>c</sup> )	$H$	$EW$ (P)	$DH$ (P)	$DHEW$ (P)
<i>Y-linked loci</i>								
<i>CpXY<sup>h</sup>8Y<sup>h</sup></i>	42	7	1.07	1.28	0.57	0.36	0.77	0.73
<i>CpXY<sup>h</sup>14Y<sup>h</sup></i>	42	6	-0.67	-0.53	-0.56	0.16	0.13	0.11
<i>PXCpXY<sup>h</sup>14Y<sup>h</sup></i>	42	8	-0.80	-0.65	0.43	0.14	0.42	0.37
<i>PXCpXY<sup>h</sup>10Y<sup>h</sup></i>	42	10	0.12	0.67	0.63	0.06	0.39	0.36
<i>CpXY<sup>h</sup>13Y<sup>h</sup></i>	42	6	0.41	n.a.	-0.71	0.42	0.30	0.27
<i>PXCpXY<sup>h</sup>6Y<sup>h</sup></i>	42	6	0.93	n.a.	0.47	0.82	0.93	0.91
<i>X-linked loci</i>								
<i>CpXY<sup>h</sup>8X</i>	48	17	-2.04*	-1.65	-7.05**	0.12	<<0.001	<b>0.02</b>
<i>CpXY<sup>h</sup>14X</i>	48	1	-0.87	n.a.	n.a.	n.a.	n.a.	n.a.
<i>PXCpXY<sup>h</sup>14X</i>	48	15	-1.97*	-1.63	-8.74**	0.35	<<0.001	0.09
<i>PXCpXY<sup>h</sup>10X</i>	48	29	-2.15*	n.a.	-6.59**	0.04	<<0.001	<<0.001
<i>CpXY<sup>h</sup>13X</i>	48	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
<i>PXCpXY<sup>h</sup>6X</i>	48	0	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.

n.d., no determined data (*CpXY<sup>h</sup>13X* fails to be amplified); n.a., not applicable data (no polymorphisms);

<sup>a</sup>n, number of alleles; <sup>b</sup>S, number of segregating sites. <sup>c</sup> $D$  (NS), Tajima's  $D$  at nonsynonymous sites; <sup>d</sup>P-value estimated on coalescent simulations. \*P<0.05, \*\*P<0.01.

**Table 4.** Coding sequence polymorphism and divergence in X/Y-linked loci

Genes	Class	Amino-acid substitution, A	Synonymous substitution, S	A/S
All 5 X-linked	Polymorphism	10	5	2
	Divergence <sup>a</sup>	226	175	1.29
2 functional X-linked	Polymorphism	5	3	1.67
	Divergence <sup>a</sup>	117	92	1.27
3 pseudo X-linked	Polymorphism	5	2	2.5
	Divergence <sup>a</sup>	109	83	1.31
All 6 Y-linked	Polymorphism	7	4	1.75
	Divergence <sup>a</sup>	288	211	1.36
Autosomal	Polymorphism	5	11	0.45
	Divergence <sup>b</sup>	65	109	0.60

<sup>a</sup>Outgroup = *V. monoica*; <sup>b</sup>Outgroup = *V. monoica* or *J. dolichaula*



## FIGURES

**Figure 1.** Diagram of chromosomal region used in this study. **(a)** Gene pairs used for comparative study are dispersed in the three evolutionary strata of the SDR on the X and Y chromosomes of papaya. The fourth heterochromatic knob in the Y region is indicated by a solid circle and its estimated corresponding position in the X region is indicated by an empty circle. The centromere is in or near Knob 4. The red brackets indicate the target X/Y regions for population analysis. **(b)** Relative positions of six focal loci on the target regions used for population study. The vertical grey bars indicate the X and Y sequences. Location of each target X/Y-linked gene is indicated by a black area. The width of each black area corresponds to gene size. The solid black lines link X-MSY paired genes. The red line on the right side serves as a scale in a unit of 0.2 Mbps.

**Figure 2.** Matched-pair analysis comparing the difference in synonymous or nonsynonymous mutations between X and Y alleles versus mean number of differences between X and Y alleles for **(a-b)** the entire SDR region, **(c-d)** stratum 1, **(e-f)** stratum 2, or **(g-h)** the collinear region. Solid horizontal lines represent the mean number of substitutions. Confidence intervals (95%) of the Y axis are indicated by the range between dotted horizontal lines.

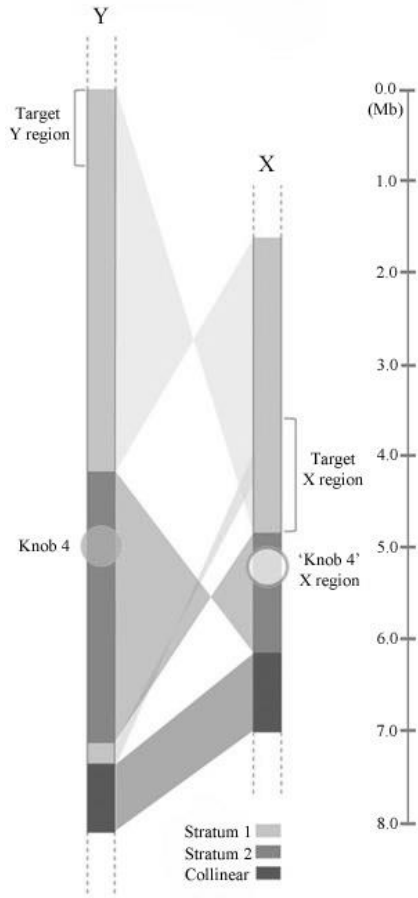
**Figure 3.** Linkage disequilibrium plots of parsimoniously informative sites for **(a)** X-linked and **(b)** Y-linked alleles in the target region. *CpXY<sup>h</sup>13X* and *PXCpXY<sup>h</sup>6X* do not have any parsimoniously informative sites. Significant LD among sites is indicated by shading ( $P < 0.001$ , dark gray;  $P < 0.01$ , medium gray;  $P < 0.05$ , light gray).

**Figure 4.** Plot of  $K_a/K_s$  versus  $K_s$  of X/Y-linked genes by **(a)** evolutionary strata or **(b)** individual genes. Divergence values are relative to orthologous sequences in *V. monoica*.

**Figure 5.** Comparison  $\pi_{(a)}/\pi_{(s)}$  ratio among X/Y and autosomal genes. Values estimated from concatenated dataset of focal X/Y loci and autosomal loci.

Figure 1

A



B

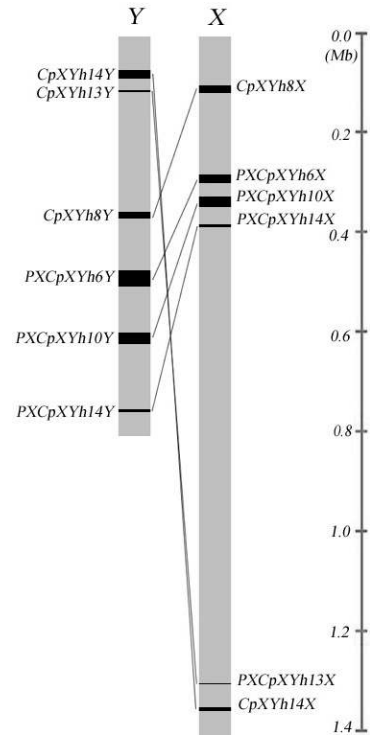


Figure 2

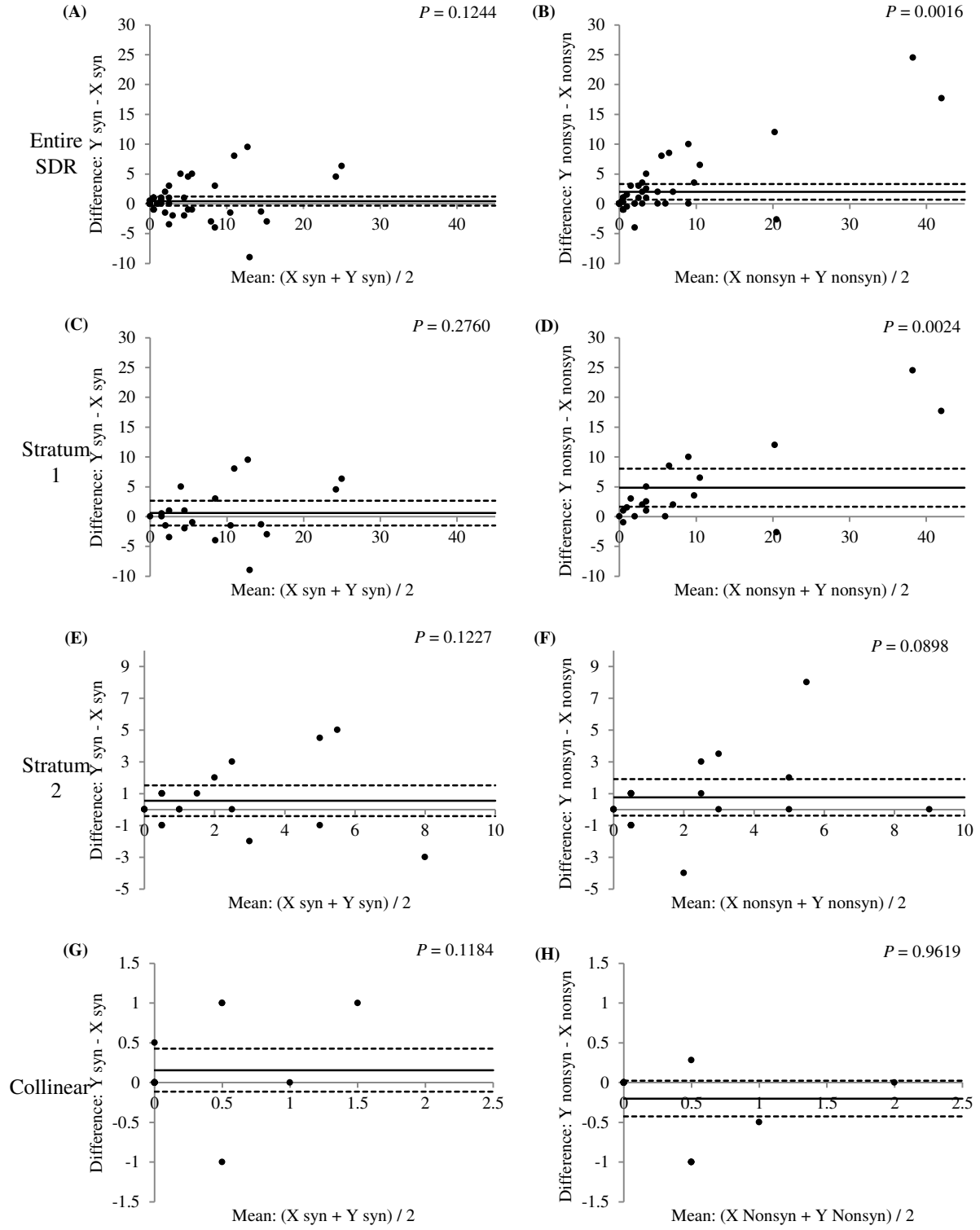


Figure 3

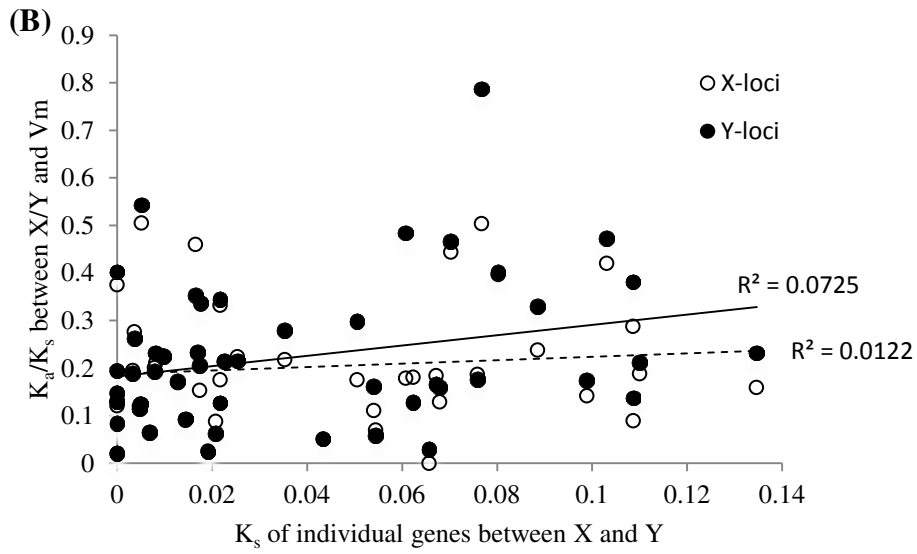
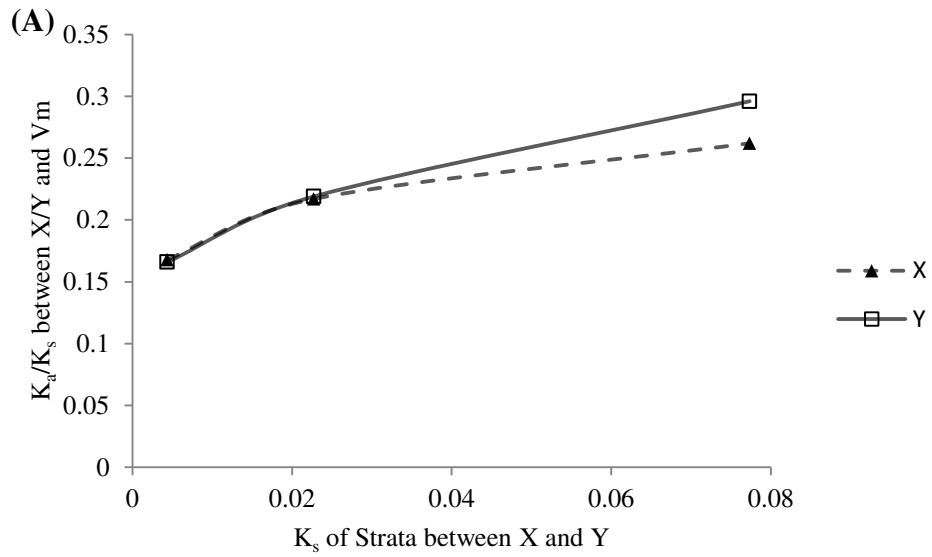
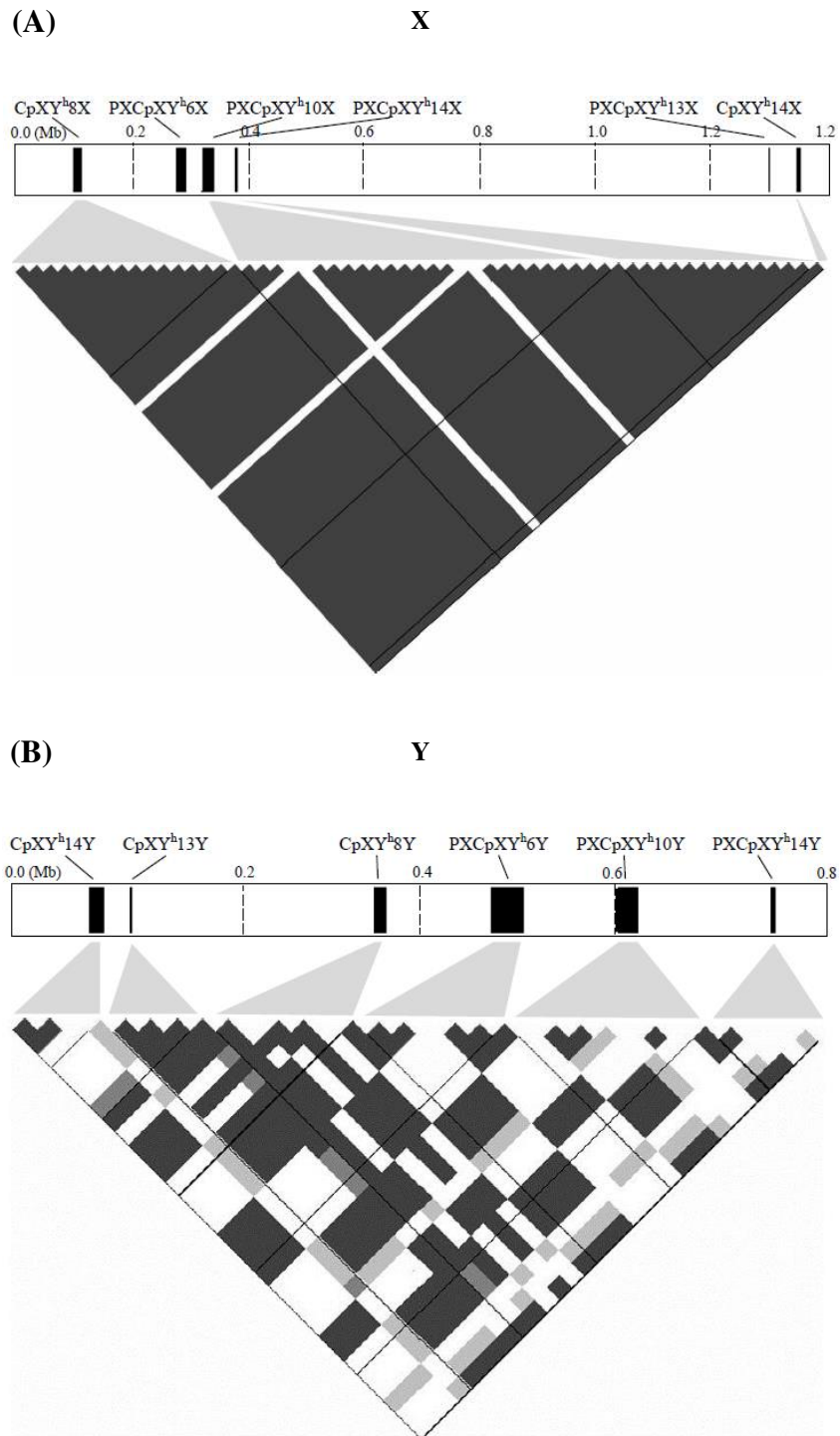
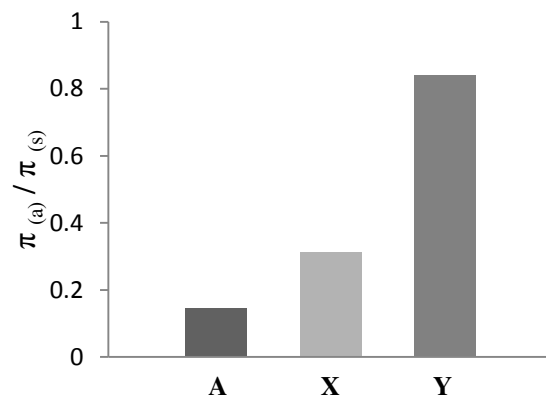


Figure 4



**Figure 5**



## APPENDIX

**Table S1.** *C. papaya* and *J. dolichaula* accessions used for analysis.

Accession	Location	Population	Region
<i>Male papayas for amplifying Y-linked loci</i>			
Cp4	9° 45.025' N, 82° 52.361' W	1	Caribbean
Cp9	9° 46.540' N, 82° 54.546' W	1	Caribbean
Cp15	9° 48.715' N, 82° 54.509' W	1	Caribbean
Cp16	9° 48.630' N, 82° 54.462' W	1	Caribbean
Cp23	9° 48.773' N, 82° 54.563' W	1	Caribbean
Cp27	9° 49.733' N, 82° 55.352' W	1	Caribbean
Cp49	10° 01.638' N, 83° 15.135' W	2	Caribbean
Cp51	10° 02.641' N, 83° 16.713' W	2	Caribbean
Cp56	10° 03.660' N, 83° 22.077' W	2	Caribbean
Cp57	10° 03.886' N, 83° 22.046' W	2	Caribbean
Cp60	10° 04.642' N, 83° 24.032' W	2	Caribbean
Cp63	10° 05.977' N, 83° 26.239' W	2	Caribbean
Cp68	10° 05.502' N, 84° 47.808' W	3	N Pacific
Cp70	10° 14.767' N, 85° 00.268' W	4	N Pacific
Cp88	10° 27.860' N, 85° 09.229' W	5	N Pacific
Cp99	10° 28.259' N, 85° 08.218' W	5	N Pacific
Cp114	10° 14.088' N, 85° 15.632' W	4	N Pacific
Cp117	10° 14.117' N, 85° 15.516' W	4	N Pacific
Cp119	10° 14.176' N, 85° 15.423' W	4	N Pacific
Cp126	9° 55.910' N, 84° 57.733' W	6	Nicoya
Cp127	9° 55.387' N, 84° 57.468' W	6	Nicoya
Cp129	9° 55.316' N, 84° 57.567' W	6	Nicoya
Cp130	9° 55.313' N, 84° 57.562' W	6	Nicoya
Cp131	9° 55.320' N, 84° 57.562' W	6	Nicoya
Cp139	9° 55.321' N, 84° 57.553' W	6	Nicoya
Cp144	9° 55.234' N, 84° 57.558' W	6	Nicoya
Cp158	9° 54.658' N, 84° 56.899' W	6	Nicoya
Cp164	9° 54.137' N, 84° 56.187' W	6	Nicoya
Cp184	9° 24.139' N, 84° 04.739' W	8	Central Pacific
Cp190	9° 23.385' N, 84° 03.693' W	8	Central Pacific

**Table S1.** Continued

<b>Accession</b>	<b>Location</b>	<b>Population</b>	<b>Region</b>
Cp193	9° 23.156' N, 84° 03.346' W	8	Central Pacific
Cp205	9° 12.837' N, 83° 49.201' W	9	S Pacific
Cp209	9° 11.443' N, 83° 46.427' W	9	S Pacific
Cp214	9° 11.142' N, 83° 45.951' W	9	S Pacific
Cp216	9° 03.650' N, 83° 38.661' W	10	S Pacific
Cp217	9° 03.645' N, 83° 38.660' W	10	S Pacific
Cp224	8° 59.882' N, 83° 33.703' W	10	S Pacific
Cp225	8° 59.963' N, 83° 33.809' W	10	S Pacific
Cp230	9° 31.391' N, 84° 25.566' W	7	Central Pacific
Cp233	9° 31.402' N, 84° 25.609' W	7	Central Pacific
Cp241	9° 35.269' N, 84° 37.295' W	7	Central Pacific
Cp244	9° 40.551' N, 84° 38.606' W	7	Central Pacific
<i>Female papayas for amplifying X-linked loci</i>			
Cp7	9° 46.521' N, 82° 54.538' W	1	Caribbean
Cp10	9° 46.540' N, 82° 54.551' W	1	Caribbean
Cp35	9° 52.794' N, 82° 57.846' W	1	Caribbean
Cp41	9° 53.837' N, 82° 58.550' W	1	Caribbean
Cp53	10° 02.642' N, 83° 16.737' W	2	Caribbean
Cp62	10° 04.664' N, 83° 24.403' W	2	Caribbean
Cp102	10° 13.556' N, 85° 07.711' W	4	N Pacific
Cp121	10° 14.160' N, 85° 15.422' W	4	N Pacific
Cp93	10° 27.690' N, 85° 08.691' W	5	N Pacific
Cp95	10° 27.698' N, 85° 08.722' W	5	N Pacific
Cp136	9° 55.324' N, 84° 57.556' W	6	Nicoya
Cp156	9° 54.607' N, 84° 57.053' W	6	Nicoya
Cp166	9° 54.136' N, 84° 56.193' W	6	Nicoya
Cp169	9° 53.900' N, 84° 56.132' W	6	Nicoya
Cp170	9° 53.852' N, 84° 56.170' W	6	Nicoya
Cp231	9° 31.415' N, 84° 25.662' W	7	Central Pacific
Cp246	9° 41.291' N, 84° 38.995' W	7	Central Pacific
Cp249	9° 41.304' N, 84° 38.753' W	7	Central Pacific
Cp176	9° 27.155' N, 84° 08.371' W	8	Central Pacific
Cp189	9° 23.384' N, 84° 03.691' W	8	Central Pacific



**Table S1.** Continued

<b>Accession</b>	<b>Location</b>	<b>Population</b>	<b>Region</b>
Cp211	9° 11.446' N, 83° 46.438' W	9	S Pacific
Cp212	9° 11.443' N, 83° 46.432' W	9	S Pacific
Cp219	9° 03.079' N, 83° 37.583' W	10	S Pacific
Cp226	8° 59.958' N, 83° 33.806' W	10	S Pacific
<i>J. dolichaula as outgroup species</i>			
Jd 36	10° 25.761' N, 84° 00.331' W	n.a.	n.a.

**Table S2.** Comparison of phylogenetic distances between *V. monoica* and *J. dolichaula* relative to *C. papaya*

Papaya Locus	$\chi^2$	<i>P</i>	Branch Length	Unique Mutations in Jd	Unique Mutations in Vm
<i>Autosomal Loci</i>					
4.44	2.17	0.14	Vm > Jd	18	28
8.72	3.93	<b>0.047*</b>	Vm > Jd	15	28
183.1	0.02	0.893	Vm > Jd	27	28
23.17	2	0.157	Jd > Vm	20	12
53.13	2.45	0.117	Vm > Jd	12	21
71.11	0.06	0.8	Jd > Vm	32	30
92.11	2.13	0.144	Vm > Jd	11	19
106.5	0.72	0.396	Vm > Jd	22	28
234.12	1.4	0.237	Vm > Jd	14	21
85.87	0.13	0.724	Vm > Jd	15	17
<i>X-linked Loci</i>					
<i>est7X</i>	0.16	0.689	Jd > Vm	52	48
<i>est40X</i>	0.17	0.683	Vm > Jd	11	13
<i>fsh176X</i>	0.03	0.857	Vm > Jd	15	16
<i>gm43X</i>	3.43	0.064	Vm > Jd	15	27
<i>gm59X</i>	1.45	0.228	Vm > Jd	18	26
<i>gm128X</i>	5	<b>0.025*</b>	Vm > Jd	5	15
<i>msy33X</i>	0.02	0.886	Vm > Jd	24	25
<i>gm88X</i>	1.26	0.262	Vm > Jd	16	23
<i>fgene39X</i>	1	0.317	Vm > Jd	15	21
<i>msy26X</i>	1.4	0.237	Vm > Jd	14	21
<i>Y-linked Loci</i>					
<i>est7Y</i>	0.65	0.419	Jd > Vm	53	45
<i>est40Y</i>	0	1	Jd = Vm	13	13
<i>fsh176Y</i>	0.13	0.715	Vm > Jd	14	16
<i>gm43Y</i>	5.49	<b>0.019*</b>	Vm > Jd	13	28
<i>gm59Y</i>	1.45	0.228	Vm > Jd	18	26
<i>gm128Y</i>	6.37	<b>0.012*</b>	Vm > Jd	4	15
<i>msy33Y</i>	0.18	0.668	Vm > Jd	23	26
<i>gm88Y</i>	0.64	0.423	Vm > Jd	17	22
<i>fgene39Y</i>	0.03	0.866	Vm > Jd	17	18

**Table S2.** Continued

<b>Papaya Locus</b>	<b><math>\chi^2</math></b>	<b><i>P</i></b>	<b>Branch Length</b>	<b>Unique Mutations in Jd</b>	<b>Unique Mutations in Vm</b>
<i>msy26Y</i>	1.88	0.17	Vm > Jd	17	26

Vm, *V. monoica*; Jd, *J. dolichaula*

**Table S3.** Summary of focal loci source and probable function

<b>Papaya locus<sup>a</sup></b>	<b>CDS sequenced<sup>b</sup></b>	<b>X/Y-linked BAC clone<sup>c</sup></b>	<b>Putative Arabidopsis homolog<sup>d</sup></b>	<b>Functional Description (from corresponding <i>A. thaliana</i> locus)</b>
<i>CpXY<sup>h</sup>8X</i>	2868 (79.3%)	XBAC SH55B21 (FJ429367)	At5g16270	Sister chromatid cohesion 1 protein 4 (SYN4) localizes to the nucleolus and is essential for gametogenesis. The Arabidopsis cohesion proteins play essential roles in mitosis and meiosis.
<i>CpXY<sup>h</sup>8Y</i>	3018 (83.6%)	YBAC SH81O12 (AC239165)		
<i>PXCpXY<sup>h</sup>6X</i>	1179 (82.0%)		AT5g36890	Beta glucosidase 42 (BGLU42) function in cation binding, beta-glucosidase activity, hydrolase activity, hydrolyzing O-glycosyl compounds, catalytic activity; involved in cellulose catabolic process, carbohydrate metabolic process
<i>PXCpXY<sup>h</sup>6Y</i>	969 (66.1%)	YBAC SH81O12 & SH32G10 (AC238596)		
<i>PXCpXY<sup>h</sup>10X</i>	1122 (75.1%)		AT5g16260	Early flowering 9 (ELF9) is involved in RNA interference, methylation-dependent chromatin silencing, negative regulation of flower development, nuclear-transcribed mRNA catabolic process, nonsense-mediated decay; functions in RNA binding, mRNA binding
<i>PXCpXY<sup>h</sup>10Y</i>	1161 (63.3%)	YBAC SH65D15 (AC238767)		
<i>PXCpXY<sup>h</sup>14X</i>	1956 (80.8%)	XBAC SH65C15 (AC239160)	At1g65470	Chromatin/Nucleosome assembly factor-1 (CAF-1) is associated with DNA methylation, DNA recombination, DNA replication initiation
<i>PXCpXY<sup>h</sup>14Y</i>	1827 (66.8%)	YBAC SH50M09 (AC238602)		
<i>CpXY<sup>h</sup>13X</i>	0 (0%)	XBAC SH46O19 (AC238599)	At5g40350	R2R3 Myb24 transcription factor (MYB24) is associated with transcription factor activity, DNA binding; involved in stamen development
<i>CpXY<sup>h</sup>13Y</i>	309 (53.4%)	YBAC SH90D06 (AC238772)		

**Table S3.** Continued

<b>Papaya locus<sup>a</sup></b>	<b>CDS sequenced<sup>b</sup></b>	<b>X/Y-linked BAC clone<sup>c</sup></b>	<b>Putative Arabidopsis homolog<sup>d</sup></b>	<b>Functional Description (from corresponding <i>A. thaliana</i> locus)</b>
<i>CpXY<sup>h</sup>14X</i>	609 (51.3%)	XBAC SH46O19 (AC238599)	At1g63940	Monodehydroascorbate reductase 4 (MDAR4) functions in ATP binding; is
<i>CpXY<sup>h</sup>14Y</i>	1158 (80.6%)	YBAC SH90D06 (AC238772)		involved in response to cadmium ion, response to cold

<sup>a</sup>Pseudogenes are denoted by a "P"; <sup>b</sup>Values in parentheses represent the ratio of sequenced CDS (coding sequences) to the predicted total CDS sizes of loci; <sup>c</sup>genbank accession indicated in parentheses; <sup>d</sup>BLAST e values  $\approx 0$

**Table S4.** Diversity ( $\theta_w$ ) and site frequency spectrum tests of selection for 25 autosomal loci used as neutral comparisons.

Locus <sup>a</sup>	N <sup>b</sup>	S <sup>c</sup>	$\theta_w$ <sup>d</sup>	<i>D</i>	<i>H</i>	<i>DH (P)</i> <sup>e</sup>
3.55	90	5	0.00077	0.093	-1.473	0.385
3.326	90	3	0.00054	-0.790	-2.105	0.067
4.44	84	1	0.00019	1.65	-0.822	0.834
6.349	84	4	0.00083	-1.167	0.295	0.260
8.72	84	2	0.00047	-0.152	0.430	0.538
9.252	84	6	0.00144	-1.289	0.336	0.26
10.98	74	3	0.00079	0.419	-3.125	0.524
14.173	84	8	0.00218	1.458	0.424	0.863
21.34	82	1	0.00032	1.637	0.513	0.840
30.5	84	3	0.00084	-0.088	0.643	0.720
41.23	84	4	0.00119	-0.803	0.462	0.464
56.38	82	4	0.00097	-0.515	-4.622	0.326
59.97	84	11	0.00241	0.672	-5.633	0.628
62.2	84	5	0.00155	0.569	-1.365	0.585
71.31	84	2	0.00054	-0.909	0.126	0.217
72.43	84	1	0.00051	-0.909	0.126	0.227
79.41	84	8	0.00228	-0.316	-5.375	0.198
85.87	84	6	0.00143	0.352	-1.398	0.4814
92.11	84	6	0.00125	-0.68	-4.753	0.064
103.38	84	1	0.00024	1.327	0.672	0.830
106.5	82	6	0.00146	2.100	0.269	0.958
107.25	90	4	0.00113	1.289	0.025	0.796
125.41	83	2	0.00105	0.986	-0.109	0.7372
132.2	90	2	0.00051	0.388	-4.631	0.496
198.38	82	4	0.00115	0.876	-1.104	0.690
<b>Average</b>			<b>0.001</b>	<b>0.296</b>	<b>-1.149</b>	

<sup>a</sup>Locus=ID based on location on papaya genome supercontigs; <sup>b</sup>N=number of sampled alleles; <sup>c</sup>S=number of segregating sites; <sup>d</sup> $\theta_w$ =Waterson's theta at silent sites; <sup>e</sup>P-value estimated on coalescent simulations

**Table S5.** Primer list for Y-linked genes' amplifications

<b>Locus name</b>	<b>Primer Name</b>	<b>Primer Direction</b>	<b>Primer (5'-3')</b>
<i>CpXY<sup>h</sup>8Y<sup>h</sup></i>			
Block 1	FSH19Y_R1_F	Forward	CGTATGGTTTTATAGGATAGTGTTTT
	FSH19Y_R1_R	Reverse	TTCATGTTCAAGTTCCTAATCC
Block 2	FSH19Y_R2_F	Forward:	CCTCTGTTTGATGATTACAGTTGTT
	FSH19Y_R2_R	Reverse:	TAACCTTACAAGCCACATGC
Block 3	FSH19Y_R3_F	Forward:	TCTTTTCCACATGGTGAATACT
	FSH19Y_R3_R	Reverse:	ATAGAAACACAACAAAAACAAGC
Block 4	FSH19Y_R4_F	Forward:	GCAATACTAGGTTAGTTGAATTGTGC
	FSH19Y_R4_R	Reverse:	AAGTGGGAGGAGAAGGTGTG
<i>CpXY<sup>h</sup>14Y<sup>h</sup></i>			
Block 1	FSH11Y_R1_F	Forward:	CATGGTGTTCCCTGAAGC
	FSH11Y_R1_R	Reverse:	GCCTAGAATTGAAAAATGTAACAG
Block 2	FSH11Y_R2_F	Forward:	GTTGGTTAACTGTTGGTTAACTGT
	FSH11Y_R2_R	Reverse:	GCAATTTTTGGTGTAATAATCG
<i>PXCpXY<sup>h</sup>14Y<sup>h</sup></i>			
Block 1	GENE11Y_R1_F	Forward:	GCAAACCTTCCCTGTTTTGG
	GENE11Y_R1_R	Reverse:	ACTAACTGAATATTTTTCTATTTCGC
Block 2	GENE11Y_R2_F	Forward:	TTATTTCTGGGGTTTTCCACC
	GENE11Y_R2_R	Reverse:	GCAATTTGAAAGAGAGAGAGGA
Block 3	GENE11Y_R3_F	Forward:	TCTAAGAAACCACTTGGACTGC
	GENE11Y_R3_R	Reverse:	CCAAATGACATGTTTCACACC
Block 4	GENE11Y_R4_F	Forward:	CAAGCAGTAGTGTCTTACCTGTATTG
	GENE11Y_R4_R	Reverse:	GCTTTGGTCCTTCAGTAGCC
<i>CpXY<sup>h</sup>13Y<sup>h</sup></i>			
Block 1	GENE183Y_R2_F	Forward:	CCCATTACCCTCATCAGC
	GENE183Y_R2_R	Reverse:	ATCTGTTGGGAGAACA AAAAATTAC
<i>PXCpXY<sup>h</sup>6Y<sup>h</sup></i>			
Block 1	EST5Y_R1_F	Forward:	ACAATAAAATTGATGAGGGTATAAG
	EST5Y_R1_R	Reverse:	GCCAAATCATAACAACATAATGG
Block 2	EST5Y_R2_F	Forward:	ACACATTCTGTTGATTATTTTCTCC
	EST5Y_R2_R	Reverse:	CAAGGTTTATGCTGCCATT
Block 3	EST5Y_R3_F	Forward:	TCATGGCTGTAATGATTGTGG
	EST5Y_R3_R	Reverse:	GCAACGAAAAGAAAAGAGAAGTT

**Table S5.** Continued

<b>Locus name</b>	<b>Primer Name</b>	<b>Primer Direction</b>	<b>Primer (5'-3')</b>
<i>PXCpXY<sup>h</sup>10Y<sup>h</sup></i>			
Block 1	GENE4Y_R1_F	Forward:	TGGCCTAAGCAAATCAGAATA
	GENE4Y_R1_R	Reverse:	TAGTGGATAGCCCTCTTTTCC
Block 2	GENE4Y_R2_F	Forward:	AGAAATTGGCATGTTAGTGG
	GENE4Y_R2_R	Reverse:	CTTCCGTTATCCTCTCTGC
Block 3	GENE4Y_R3_F	Forward:	TTCTCGTCTCTCCTCCTACC
	GENE4Y_R3_R	Reverse:	ACAAGTTGTCCAGAAAACAATATG
Block 4	GENE4Y_R4_F	Forward:	GCGCATTGTTTATGATGGCA
	GENE4Y_R4_R	Reverse:	GGAGGAATCCCTCTTTCAGC



**Table S6.** Primer list for X-linked genes' amplifications

<b>Locus name</b>	<b>Primer Name</b>	<b>Primer Direction</b>	<b>Primer (5'-3')</b>
<i>CpXY<sup>h</sup>8X</i>			
Block 1	FSH19X_R1_F	Forward	TGGTTTTACAGGATAGTGTGG
	FSH19X_R1_R	Reverse	CCTTCATGTTCAAGTTTCTAATCC
Block 2	FSH19X_R2_F	Forward:	GTTACCTCCGTTTGATGATTGC
	FSH19X_R2_R	Reverse:	GTATAACCTTACAAGCCACATGC
Block 3	FSH19X_R3_F	Forward:	TGTGAATAACAAGAGATTTTATCTGG
	FSH19X_R3_R	Reverse:	CTAGAGACACAACAAAAACAAGC
Block 4	FSH19X_R4_F	Forward:	GCAATACTAGGTTAGTTGAATTGTTC
	FSH19X_R4_R	Reverse:	CAAGTGGGAGGGGAAGGT
<i>CpXY<sup>h</sup>14X</i>			
Block 1	FSH11X_R1_F	Forward:	CATGGTGTTCCTGAAGC
	FSH11X_R1_R	Reverse:	GGGTAAACTAATAATTCATATCGTG
Block 2	FSH11X_R2_F	Forward:	CCTAAAAGCAAGACTGGATGC
	FSH11X_R2_R	Reverse:	TGCAATTTTTGGTGTAATAAGC
<i>PXCpXY<sup>h</sup>14X</i>			
Block 1	GENE11X_R1_F	Forward:	GCAAACCTTCCATTCTTGG
	GENE11X_R1_R	Reverse:	GACCGACACTGAATCACTAACAG
Block 2	GENE11X_R2_F	Forward:	GTTATTTCTGCGTTTCCAC
	GENE11X_R2_R	Reverse:	AAGGAGCCCAAGGAAAGTG
Block 3	GENE11X_R3_F	Forward:	CTAAGCTCTAAGAAATCACCTGGAC
	GENE11X_R3_R	Reverse:	CCCAAATGATATGTTTCACACC
Block 4	GENE11X_R4_F	Forward:	CAAGCAGTAGTGCTTACCTGTATTG
	GENE11X_R4_R	Reverse:	GCTTTGGTCCTTCAGTAGCC
<i>PXCpXY<sup>h</sup>6X</i>			
Block 1	EST5X_R1_F	Forward:	TTTTTAATGGCTGCTAGAGATGG
	EST5X_R1_R	Reverse:	CCTAATATGCCCGACCAAGC
Block 2	EST5X_R2_F	Forward:	ACACATTCTGTTGATTATTTTCTCC
	EST5X_R2_R	Reverse:	CAAGGTTTATGCTGCCATT
<i>PXCpXY<sup>h</sup>10X</i>			
Block 1	GENE4X_R1_F	Forward:	CGTACGCAAATCAGAGAGG
	GENE4X_R1_R	Reverse:	ATAGTGGATAGCCTTCCTTTCC
Block 2	GENE4X_R2_F	Forward:	GAGAAATTGGCATGTTAGTGG
	GENE4X_R2_R	Reverse:	CCAATTTTCTGTTATCCTCTCC

**Table S6.** Continued

<b>Locus name</b>	<b>Primer Name</b>	<b>Primer Direction</b>	<b>Primer (5'-3')</b>
Block 3	GENE4X_R3_F	Forward:	AAATTCTACTCAATGCTCTTGTACC
	GENE4X_R3_R	Reverse:	CAAGTTATCCAGAAGACAATACAAGG
Block 4	GENE4X_R4_F	Forward:	GGTCACATTGTTTATGATGATGG
	GENE4X_R4_R	Reverse:	AGGCCGGTCTAGGAAACC

**Table S7.** List of functional gene pairs

<b>Locus</b>	<b>CDS length predicted<sup>a</sup> (bp)</b>	<b>CDS length acquired<sup>b</sup> (bp)</b>	<b>Percentage<sup>c</sup> (%)</b>	<b>CDS length w/ monoica<sup>d</sup> (bp)</b>	<b>Percentage<sup>e</sup> (%)</b>
<i>Stratum 1</i>					
<i>CpXY<sup>h</sup>1</i>	2265	2265	100	522	23.0
<i>CpXY<sup>h</sup>2</i>	1884	1884	100	1713	90.9
<i>CpXY<sup>h</sup>3</i>	2268	1322	58.3	465	20.5
<i>CpXY<sup>h</sup>4</i>	293	293	100	0	0
<i>CpXY<sup>h</sup>5</i>	807	807	100	624	77.3
<i>CpXY<sup>h</sup>6</i>	3336	1803	54.0	1188	35.6
<i>CpXY<sup>h</sup>7</i>	2271	2271	100	1566	69.0
<i>CpXY<sup>h</sup>8</i>	3612	3612	100	3319	91.9
<i>CpXY<sup>h</sup>9</i>	1161	1161	100	747	64.3
<i>CpXY<sup>h</sup>10</i>	525	525	100	177	33.7
<i>CpXY<sup>h</sup>11</i>	3111	3111	100	2391	76.9
<i>CpXY<sup>h</sup>12</i>	552	552	100	279	50.5
<i>CpXY<sup>h</sup>13</i>	579	579	100	452	78.1
<i>CpXY<sup>h</sup>14</i>	1437	1437	100	1109	77.2
<b>Total 1</b>	<b>24101</b>	<b>21622</b>	<b>89.7</b>	<b>14552</b>	<b>60.4</b>
<i>Stratum 2</i>					
<i>CpXY<sup>h</sup>15</i>	3804	3050	80.2	2048	53.8
<i>CpXY<sup>h</sup>16</i>	237	237	100	204	86.1
<i>CpXY<sup>h</sup>17</i>	1347	1347	100	1182	87.8
<i>CpXY<sup>h</sup>18</i>	744	573	77.0	330	44.4
<i>CpXY<sup>h</sup>19</i>	819	819	100	255	31.1
<i>CpXY<sup>h</sup>20</i>	6609	3680	55.7	2786	42.2
<i>CpXY<sup>h</sup>21</i>	1272	1272	100	927	72.9
<i>CpXY<sup>h</sup>22</i>	243	243	100	0	0
<i>CpXY<sup>h</sup>23</i>	1011	1011	100	723	71.5
<i>CpXY<sup>h</sup>24</i>	1779	1779	100	1077	60.5
<i>CpXY<sup>h</sup>25</i>	549	549	100	225	41.0
<i>CpXY<sup>h</sup>26</i>	2703	2290	84.7	1536	56.8
<i>CpXY<sup>h</sup>27</i>	2088	1230	58.9	920	44.1
<i>CpXY<sup>h</sup>28</i>	264	249	94.3	85	32.2
<i>CpXY<sup>h</sup>29</i>	1686	1585	94.0	1280	75.9

**Table S7.** Continued

<b>Locus</b>	<b>CDS length predicted<sup>a</sup> (bp)</b>	<b>CDS length acquired<sup>b</sup> (bp)</b>	<b>Percentage<sup>c</sup> (%)</b>	<b>CDS length w/ monoica<sup>d</sup> (bp)</b>	<b>Percentage<sup>e</sup> (%)</b>
<i>CpXY<sup>h</sup>31</i>	618	618	100	180	29.1
<i>CpXY<sup>h</sup>32</i>	747	747	100	453	60.6
<i>CpXY<sup>h</sup>33</i>	504	504	100	344	68.3
<b>Total 2</b>	<b>27336</b>	<b>22095</b>	<b>80.8</b>	<b>14555</b>	<b>53.2</b>
<i>Collinear</i>					
<i>CpXY<sup>h</sup>34</i>	1413	1413	100	814	57.6
<i>CpXY<sup>h</sup>35</i>	819	819	100	428	52.3
<i>CpXY<sup>h</sup>36</i>	1149	1149	100	822	71.5
<i>CpXY<sup>h</sup>37</i>	1299	1299	100	876	67.4
<i>CpXY<sup>h</sup>38</i>	903	903	100	252	27.9
<i>CpXY<sup>h</sup>39</i>	1017	1017	100	420	41.3
<i>CpXY<sup>h</sup>40</i>	1845	1845	100	1131	61.3
<i>CpXY<sup>h</sup>41</i>	1149	1149	100	741	64.5
<i>CpXY<sup>h</sup>42</i>	1956	1835	93.8	996	50.9
<i>CpXY<sup>h</sup>43</i>	1287	566	44.0	448	34.8
<i>CpXY<sup>h</sup>44</i>	897	280	31.2	231	25.8
<i>CpXY<sup>h</sup>45</i>	1419	1089	76.7	712	50.2
<i>CpXY<sup>h</sup>46</i>	498	498	100	129	25.9
<i>CpXY<sup>h</sup>47</i>	3615	3268	90.4	2701	74.7
<i>CpXY<sup>h</sup>48</i>	174	174	100	174	100
<i>CpXY<sup>h</sup>49</i>	3354	1770	52.8	1500	44.7
<i>CpXY<sup>h</sup>50</i>	309	309	100	0	0
<b>Total 3</b>	<b>23103</b>	<b>19383</b>	<b>83.9</b>	<b>12375</b>	<b>53.6</b>
<i>Entire SDR</i>					
<b>Total</b>	<b>74540</b>	<b>63100</b>	<b>84.7</b>	<b>41482</b>	<b>55.7</b>

<sup>a</sup>CDS length predicted is length of coding sequences (CDS) of each gene based on the information listed in the Table S9 of supplementary materials from Wang et al. (2012); <sup>b</sup>CDS length acquired is the actual length of CDS of each gene we acquired for this research; <sup>c</sup>Percentage = CDS length acquired ÷ CDS length predicted × 100%; <sup>d</sup>CDS length w/ monoica is the actual length of CDS of each gene we acquired when *V. monoica* sequences were added and gaps of homologous sequences were deleted; <sup>e</sup>Percentage = CDS length w/ monoica ÷ CDS length predicted × 100%

**Table S8.** List of gene pairs of which Y-linked gene is pseudogene

<b>Locus</b>	<b>CDS length predicted<sup>a</sup> (bps)</b>	<b>CDS length acquired<sup>b</sup> (bps)</b>	<b>Percentage<sup>c</sup> (%)</b>	<b>CDS length w/ monoica<sup>d</sup> (bps)</b>	<b>Percentage<sup>e</sup> (%)</b>
<i>Stratum 1</i>					
<i>PY<sup>h</sup>CpXY<sup>h</sup>2</i>	369	369	100	0	0
<i>PY<sup>h</sup>CpXY<sup>h</sup>3</i>	600	600	100	0	0
<i>PY<sup>h</sup>CpXY<sup>h</sup>5</i>	261	261	100	96	36.8
<i>PY<sup>h</sup>CpXY<sup>h</sup>7</i>	813	813	100	282	34.7
<i>PY<sup>h</sup>CpXY<sup>h</sup>22</i>	3240	2217	68.4	1580	48.8
<i>PY<sup>h</sup>CpXY<sup>h</sup>23</i>	339	339	100	0	0
<b>Total 1</b>	<b>5622</b>	<b>4599</b>	<b>81.8</b>	<b>1958</b>	<b>34.8</b>
<i>Stratum 2</i>					
<i>PY<sup>h</sup>CpXY<sup>h</sup>27</i>	252	252	100	0	0
<i>PY<sup>h</sup>CpXY<sup>h</sup>28</i>	1029	1029	100	497	48.3
<i>PY<sup>h</sup>CpXY<sup>h</sup>30</i>	684	684	100	0	0
<i>PY<sup>h</sup>CpXY<sup>h</sup>31</i>	465	465	100	289	62.2
<b>Total 2</b>	<b>2430</b>	<b>2430</b>	<b>100</b>	<b>786</b>	<b>32.3</b>
<i>Entire SDR</i>					
<b>Total</b>	<b>8052</b>	<b>7029</b>	<b>87.3</b>	<b>2744</b>	<b>34.1</b>

<sup>a</sup>CDS length predicted is length of coding sequences (CDS) of each gene based on the information listed in the Table S10 of supplementary materials from Wang et al. (2012); <sup>b</sup>CDS length acquired is the actual length of CDS of each gene we acquired for this research; <sup>c</sup>Percentage = CDS length acquired ÷ CDS length predicted × 100%; <sup>d</sup>CDS length w/ monoica is the actual length of CDS of each gene we acquired when *V. monoica* sequences were added and gaps of homologous sequences were deleted; <sup>e</sup>Percentage = CDS length w/ monoica ÷ CDS length predicted × 100%

**Table S9.** List of gene pairs of which X-linked gene is pseudogene

<b>Locus</b>	<b>CDS length predicted<sup>a</sup> (bp)</b>	<b>CDS length acquired<sup>b</sup> (bp)</b>	<b>Percentage<sup>c</sup> (%)</b>	<b>CDS length w/ monoica<sup>d</sup> (bp)</b>	<b>Percentage<sup>e</sup> (%)</b>
<i>Stratum 1</i>					
<i>PXCpXY<sup>h</sup>6</i>	1467	1467	100	883	60.2
<i>PXCpXY<sup>h</sup>10</i>	1834	1834	100	1041	56.8
<i>PXCpXY<sup>h</sup>14</i>	2736	2736	100	1796	65.6
<i>PXCpXY<sup>h</sup>15</i>	965	0	0	0	0
<b>Total 1</b>	<b>7002</b>	<b>6037</b>	<b>86.2</b>	<b>3720</b>	<b>53.1</b>
<i>Stratum 2</i>					
<i>PXCpXY<sup>h</sup>21</i>	363	363	100	0	0
<i>PXCpXY<sup>h</sup>22</i>	432	0	0	0	0
<b>Total 2</b>	<b>795</b>	<b>363</b>	<b>45.7</b>	<b>0</b>	<b>0</b>
<i>Entire SDR</i>					
<b>Total</b>	<b>7797</b>	<b>6400</b>	<b>82.1</b>	<b>3720</b>	<b>47.7</b>

<sup>a</sup>CDS length predicted is length of coding sequences (CDS) of each gene based on the information listed in the Table S11 of supplementary materials from Wang et al. (2012); <sup>b</sup>CDS length acquired is the actual length of CDS of each gene we acquired for this research; <sup>c</sup>Percentage = CDS length acquired ÷ CDS length predicted × 100%; <sup>d</sup>CDS length w/ monoica is the actual length of CDS of each gene we acquired when *V. monoica* sequences were added and gaps of homologous sequences were deleted; <sup>e</sup>Percentage = CDS length w/ monoica ÷ CDS length predicted × 100%

**Table S10.** List of gene pairs of which X-linked and Y-linked gene are both pseudogenes

Locus	CDS length predicted <sup>a</sup> (bp)	CDS length acquired <sup>b</sup> (bp)	Percentage <sup>c</sup> (%)	CDS length w/ monoica <sup>d</sup> (bp)	Percentage <sup>e</sup> (%)
<i>Stratum 1</i>					
<i>PXY<sup>h</sup>CpXY<sup>h</sup>1</i>	840	528	62.9	410	48.8
<b>Total 1</b>	<b>840</b>	<b>528</b>	<b>62.9</b>	<b>410</b>	<b>48.8</b>
<i>Stratum 2</i>					
<i>PXY<sup>h</sup>CpXY<sup>h</sup>2</i>	306	0	0	0	0
<i>PXY<sup>h</sup>CpXY<sup>h</sup>3</i>	652	0	0	0	0
<b>Total 2</b>	<b>958</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<i>Collinear</i>					
<i>PXY<sup>h</sup>CpXY<sup>h</sup>4</i>	905	0	0	0	0
<b>Total 3</b>	<b>905</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<i>Entire SDR</i>					
<b>Total</b>	<b>2703</b>	<b>528</b>	<b>19.5</b>	<b>410</b>	<b>15.2</b>

<sup>a</sup>CDS length predicted is length of coding sequences (CDS) of each gene based on the information listed in the Table S9 of supplementary materials from Wang et al. (2012); <sup>b</sup>CDS length acquired is the actual length of CDS of each gene we acquired for this research; <sup>c</sup>Percentage = CDS length acquired ÷ CDS length predicted × 100%; <sup>d</sup>CDS length w/ monoica is the actual length of CDS of each gene we acquired when V. monoica sequences were added and gaps of homologous sequences were deleted; <sup>e</sup>Percentage = CDS length w/ monoica ÷ CDS length predicted × 100%

**Table S11.** PAML-based tests of rate differences between each X-MSY gene pair, using homologous genes of *V. monoica* as outgroup sequences

Locus	Number of codons	Model	LnL <sup>a</sup>	$\chi^2$	df	P	free $d_N/d_S$ <sup>b</sup>
<i>Stratum 1</i>							
<i>CpXY<sup>h1</sup></i>	174	Model = 0	-815.072	0.944	2	0.624	X: 0.0001
		Model = 1	-814.600				Y: 0.0001
<i>CpXY<sup>h2</sup></i>	571	Model = 0	-2689.056	0.584	2	0.747	X: 0.1033
		Model = 1	--2688.764				Y: 0.0434
<i>CpXY<sup>h3</sup></i>	155	Model = 0	-694.962	2.526	2	0.283	X: 0.1650
		Model = 1	-693.699				Y: 0.0001
<i>CpXY<sup>h4</sup></i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.
<i>CpXY<sup>h5</sup></i>	207	Model = 0	-1029.630	1.758	2	0.415	X: 0.3874
		Model = 1	-1028.751				Y: 0.2624
<i>CpXY<sup>h6</sup></i>	396	Model = 0	-1913.193	6.924	2	<b>0.031</b>	X: 0.1710
		Model = 1	-1909.731				Y: 0.7703
<i>CpXY<sup>h7</sup></i>	521	Model = 0	-2581.604	0.874	2	0.646	X: 0.1844
		Model = 1	-2581.167				Y: 0.3018
<i>CpXY<sup>h8</sup></i>	1099	Model = 0	-6005.801	0.357	2	0.837	X: 0.4571
		Model = 1	-6005.623				Y: 0.5577
<i>CpXY<sup>h9</sup></i>	248	Model = 0	-1255.834	0.399	2	0.819	X: 0.1205
		Model = 1	-1255.634				Y: 0.2013
<i>CpXY<sup>h10</sup></i>	59	Model = 0	-292.862	1.670	2	0.434	X: 0.0001
		Model = 1	-292.027				Y: 0.3678
<i>CpXY<sup>h11</sup></i>	797	Model = 0	-4614.639	1.142	2	0.565	X: 0.3423
		Model = 1	-4614.068				Y: 0.5376
<i>CpXY<sup>h12</sup></i>	92	Model = 0	-426.097	1.607	2	0.448	X: 0.0001
		Model = 1	-425.294				Y: 0.0722
<i>CpXY<sup>h13</sup></i>	149	Model = 0	-738.030	1.288	2	0.525	X: 0.1006
		Model = 1	-737.386				Y: 0.2651
<i>CpXY<sup>h14</sup></i>	366	Model = 0	-1705.698	0.472	2	0.790	X: 0.1418
		Model = 1	-1705.462				Y: 0.1252
<i>PXCpXY<sup>h6</sup></i>	291	Model = 0	-1588.403	2.806	2	0.246	X: 0.0957
		Model = 1	-1587.000				Y: 0.2810



**Table S11.** Continued

<b>Locus</b>	<b>Number of codons</b>	<b>Model</b>	<b>LnL<sup>a</sup></b>	<b><math>\chi^2</math></b>	<b>df</b>	<b>P</b>	<b>free d<sub>N</sub>/d<sub>S</sub><sup>b</sup></b>
<i>PXCpXY<sup>h</sup>10</i>	344	Model = 0	-1839.174	1.484	2	0.476	X: 0.1789
		Model = 1	-1838.431				Y: 0.3596
<i>PXCpXY<sup>h</sup>14</i>	596	Model = 0	-3161.233	0.089	2	0.956	X: 0.4954
		Model = 1	-3161.189				Y: 0.4545
<i>PXCpXY<sup>h</sup>15</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.
<i>PY<sup>h</sup>CpXY<sup>h</sup>2</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.
<i>PY<sup>h</sup>CpXY<sup>h</sup>3</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.
<i>PY<sup>h</sup>CpXY<sup>h</sup>5</i>	31	Model = 0	-161.088	0	2	1	X: 999
		Model = 1	-161.088				Y: 999
<i>PY<sup>h</sup>CpXY<sup>h</sup>7</i>	92	Model = 0	-490.496	1.022	2	0.600	X: 0.2276
		Model = 1	-489.985				Y: 1.0260
<i>PY<sup>h</sup>CpXY<sup>h</sup>22</i>	525	Model = 0	-2801.900	4.964	2	0.084	X: 0.2559
		Model = 1	-2799.416				Y: 0.7238
<i>PY<sup>h</sup>CpXY<sup>h</sup>23</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.
<i>PXY<sup>h</sup>CpXY<sup>h</sup>1</i>	135	Model = 0	-726.018	9.525	2	<b>0.0085</b>	X: 0.1905
		Model = 1	-721.256				Y: 4.6207
<i>Stratum 2</i>							
<i>CpXY<sup>h</sup>15</i>	679	Model = 0	-3378.387	0.368	2	0.832	X: 0.0991
		Model = 1	-3378.203				Y: 0.2309
<i>CpXY<sup>h</sup>16</i>	68	Model = 0	-308.316	2.304	2	0.316	X: 0.0001
		Model = 1	-307.164				Y: 999
<i>CpXY<sup>h</sup>17</i>	393	Model = 0	-1894.299	2.261	2	0.323	X: 0.6619
		Model = 1	-1893.169				Y: 0.5020
<i>CpXY<sup>h</sup>18</i>	109	Model = 0	-543.900	1.148	2	0.563	X: 6.7885
		Model = 1	-543.326				Y: 999
<i>CpXY<sup>h</sup>19</i>	85	Model = 0	-397.413	2.710	2	0.258	X: 999
		Model = 1	-396.058				Y: 0.0001

**Table S11.** Continued

<b>Locus</b>	<b>Number of codons</b>	<b>Model</b>	<b>LnL<sup>a</sup></b>	<b><math>\chi^2</math></b>	<b>df</b>	<b>P</b>	<b>free d<sub>N</sub>/d<sub>S</sub><sup>b</sup></b>
<i>CpXY<sup>h</sup>20</i>	928	Model = 0	-4670.270	0.627	2	0.731	X: 0.2319
		Model = 1	-4669.957				Y: 0.4259
<i>CpXY<sup>h</sup>21</i>	309	Model = 0	-1397.853	2.777	2	0.249	X: 0.3479
		Model = 1	-1396.464				Y: 0.0001
<i>CpXY<sup>h</sup>22</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.
<i>CpXY<sup>h</sup>23</i>	240	Model = 0	-1167.039	11.841	2	<b>0.0027</b>	X: 999
		Model = 1	-1161.119				Y: 0.0001
<i>CpXY<sup>h</sup>24</i>	359	Model = 0	-1793.057	3.134	2	0.209	X: 0.6913
		Model = 1	-1791.491				Y: 0.6099
<i>CpXY<sup>h</sup>25</i>	74	Model = 0	-391.573	1.100	2	0.575	X: 0.0001
		Model = 1	-391.023				Y: 0.0001
<i>CpXY<sup>h</sup>26</i>	508	Model = 0	-2441.034	0.859	2	0.651	X: 0.1149
		Model = 1	-2440.604				Y: 0.1510
<i>CpXY<sup>h</sup>27</i>	305	Model = 0	-1419.431	0.528	2	0.768	X: 0.0001
		Model = 1	-1419.168				Y: 0.1000
<i>CpXY<sup>h</sup>28</i>	26	Model = 0	-107.056	0	2	1	X: 77.520
		Model = 1	-107.056				Y: 125.95
<i>CpXY<sup>h</sup>29</i>	426	Model = 0	-2161.536	5.068	2	0.079	X: 0.0506
		Model = 1	-2159.002				Y: 0.7241
<i>CpXY<sup>h</sup>30</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.
<i>CpXY<sup>h</sup>31</i>	59	Model = 0	-285.068	6.798	2	0.033	X: 999
		Model = 1	-281.669				Y: 0.0001
<i>CpXY<sup>h</sup>32</i>	150	Model = 0	-834.590	1.631	2	0.442	X: 0.8902
		Model = 1	833.775				Y: 0.9517
<i>CpXY<sup>h</sup>33</i>	114	Model = 0	-523.214	0.985	2	0.611	X: 0.0001
		Model = 1	-522.722				Y: 0.3168
<i>PXCpXY<sup>h</sup>21</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.
<i>PXCpXY<sup>h</sup>22</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.

**Table S11.** Continued

<b>Locus</b>	<b>Number of codons</b>	<b>Model</b>	<b>LnL<sup>a</sup></b>	<b><math>\chi^2</math></b>	<b>df</b>	<b>P</b>	<b>free d<sub>N</sub>/d<sub>S</sub><sup>b</sup></b>
<i>PY<sup>hi</sup>CpXY<sup>hi</sup>27</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.
<i>PY<sup>hi</sup>CpXY<sup>hi</sup>28</i>	165	Model = 0	-713.834	0.133	2	0.936	X: 2.0960
		Model = 1	-713.768				Y: 0.0001
<i>PY<sup>hi</sup>CpXY<sup>hi</sup>30</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.
<i>PY<sup>hi</sup>CpXY<sup>hi</sup>31</i>	95	Model = 0	-546.709	1.435	2	0.488	X: 0.3126
		Model = 1	-545.992				Y: 1.0126
<i>PXY<sup>hi</sup>CpXY<sup>hi</sup>2</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.
<i>PXY<sup>hi</sup>CpXY<sup>hi</sup>3</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.
<i>Collinear</i>							
<i>CpXY<sup>hi</sup>34</i>	270	Model = 0	-1281.725	0.027	2	0.987	X: 0.3496
		Model = 1	-1281.711				Y: 0.0001
<i>CpXY<sup>hi</sup>35</i>	142	Model = 0	-704.325	0	2	1	X: 3.0418
		Model = 1	-704.325				Y: 3.0418
<i>CpXY<sup>hi</sup>36</i>	273	Model = 0	-1320.355	0	2	1	X: 6.7233
		Model = 1	-1320.355				Y: 5.0099
<i>CpXY<sup>hi</sup>37</i>	292	Model = 0	-1417.728	3.822	2	0.148	X: 999
		Model = 1	-1415.817				Y: 999
<i>CpXY<sup>hi</sup>38</i>	84	Model = 0	-366.312	0	2	1	X: 1.2615
		Model = 1	-366.312				Y: 0.0001
<i>CpXY<sup>hi</sup>39</i>	140	Model = 0	-643.787	0	2	1	X: 6.0363
		Model = 1	-643.787				Y: 0.2951
<i>CpXY<sup>hi</sup>40</i>	377	Model = 0	-1802.625	2.646	2	0.266	X: 999
		Model = 1	-1801.301				Y: 0.0001
<i>CpXY<sup>hi</sup>41</i>	247	Model = 0	-1188.777	1.182	2	0.554	X: 999
		Model = 1	-1188.007				Y: 0.0001
<i>CpXY<sup>hi</sup>42</i>	332	Model = 0	-1542.212	0	2	1	X: 0.0001
		Model = 1	-1542.212				Y: 0.0001

**Table S11.** Continued

<b>Locus</b>	<b>Number of codons</b>	<b>Model</b>	<b>LnL<sup>a</sup></b>	<b><math>\chi^2</math></b>	<b><i>df</i></b>	<b><i>P</i></b>	<b>free <math>d_N/d_S</math><sup>b</sup></b>
<i>CpXY<sup>h</sup>43</i>	148	Model = 0	-650.084	0	2	1	X: 3.7435
		Model = 1	-650.084				Y: 5.5687
<i>CpXY<sup>h</sup>44</i>	77	Model = 0	-347.411	1.704	2	0.427	X: 0.0001
		Model = 1	-346.560				Y: 0.0001
<i>CpXY<sup>h</sup>45</i>	237	Model = 0	-1138.093	0.930	2	0.628	X: 2.7199
		Model = 1	-1137.628				Y: 0.0001
<i>CpXY<sup>h</sup>46</i>	43	Model = 0	-237.445	0	2	1	X: 7.5883
		Model = 1	-237.445				Y: 14.863
<i>CpXY<sup>h</sup>47</i>	896	Model = 0	-4192.464	0.842	2	0.656	X: 0.0001
		Model = 1	-4192.042				Y: 0.1510
<i>CpXY<sup>h</sup>48</i>	58	Model = 0	-270.720	0.511	2	0.775	X: 0.0001
		Model = 1	-270.465				Y: 0.0001
<i>CpXY<sup>h</sup>49</i>	500	Model = 0	-2494.026	1.786	2	0.409	X: 0.3195
		Model = 1	-2493.133				Y: 999
<i>CpXY<sup>h</sup>50</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.
<i>PXY<sup>h</sup>CpXY<sup>h</sup>4</i>	0	Model = 0	n.a.	n.a.	2	n.a.	X: n.a.
		Model = 1	n.a.				Y: n.a.

<sup>a</sup>LnL means log likelihood; <sup>b</sup>free  $d_N/d_S$  represent the ratio of nonsynonymous/synonymous substitution rates in each branch under model 2

**Table S12.** Sequencing coverage

Loci	n <sup>a</sup>	Total sites	Seq-CDS <sup>b</sup>	Pre-CDS <sup>c</sup>	Percentage <sup>d</sup>	S <sup>e</sup>	S <sub>wm</sub> <sup>f</sup>	Percentage <sup>g</sup>
<i>Y-linked loci</i>								
<i>CpXY<sup>h</sup>8Y<sup>h</sup></i>	42	4569	3018	3612	83.6%	7	6	85.7%
<i>CpXY<sup>h</sup>14Y<sup>h</sup></i>	42	2343	1158	1437	80.6%	6	5	83.3%
<i>PXCpXY<sup>h</sup>14Y<sup>h</sup></i>	42	3838	1827	2736	66.8%	8	4	50.0%
<i>PXCpXY<sup>h</sup>10Y<sup>h</sup></i>	42	3477	1161	1834	63.3%	10	3	30.0%
<i>CpXY<sup>h</sup>13Y<sup>h</sup></i>	42	888	309	579	53.4%	6	5	83.3%
<i>PXCpXY<sup>h</sup>6Y<sup>h</sup></i>	42	2692	969	1467	66.1%	6	2	33.4%
<b>Total</b>	<b>42</b>	<b>17807</b>	<b>8433</b>	<b>11665</b>	<b>72.3%</b>	<b>43</b>	<b>25</b>	<b>58.1%</b>
<i>X-linked loci</i>								
<i>CpXY<sup>h</sup>8X</i>	48	4476	2868	3618	79.3%	17	16	94.1%
<i>CpXY<sup>h</sup>14X</i>	48	2345	1179	1437	82.0%	1	0	0.00%
<i>PXCpXY<sup>h</sup>14X</i>	48	3658	1956	2420	80.8%	15	10	67.7%
<i>PXCpXY<sup>h</sup>10X</i>	48	3668	1122	1495	75.1%	29	12	41.4%
<i>CpXY<sup>h</sup>13X</i>	48	0	0	579	0	n.d.	n.d.	n.d.
<i>PXCpXY<sup>h</sup>6X</i>	48	1710	609	1186	51.3%	0	0	n.a.
<b>Total</b>	<b>48</b>	<b>15857</b>	<b>7734</b>	<b>10735</b>	<b>72.0%</b>	<b>62</b>	<b>38</b>	<b>61.3%</b>

<sup>a</sup>n = number of sampled alleles; <sup>b</sup>Seq-CDS represents length of coding sequences (CDS) by sequencing; <sup>c</sup>Pre-CDS represents length of coding sequences (CDS) of each gene based on the information listed in the Table S9 of supplementary materials from Wang et al. (2012); <sup>d</sup>Percentage = Seq-CDS ÷ Pre-CDS; <sup>e</sup>S = number of segregating sites; <sup>f</sup>S<sub>wm</sub> = number of segregating sites when *V. monoica*'s sequences are added into alignment; <sup>g</sup>Percentage = S<sub>wm</sub> ÷ S

**Table S13.** Maximum-likelihood *HKA* analysis of silent polymorphism of papaya sex-linked loci against 25 autosomal loci.

Loci	Model	ML <sup>a</sup>	$\chi^2$	df	P	K <sup>b</sup> (gene)
<i>Y-linked loci</i>						
<i>CpXY<sup>h</sup>8Y<sup>h</sup></i>	Neutral	-151.122	0.654	1	0.4187	1
	Selection	-150.795				1.62
<i>CpXY<sup>h</sup>14Y<sup>h</sup></i>	Neutral	-151.834	2.324	1	0.1274	1
	Selection	-151.774				2.72
<i>PXCpXY<sup>h</sup>14Y<sup>h</sup></i>	Neutral	-151.193	1.162	1	0.2811	1
	Selection	-151.034				2.01
<i>PXCpXY<sup>h</sup>10Y<sup>h</sup></i>	Neutral	-152.577	1.85	1	0.1738	1
	Selection	-151.652				2.41
<i>CpXY<sup>h</sup>13Y<sup>h</sup></i>	Neutral	-155.891	10.272	1	0.0014**	1
	Selection	-150.754				7.32
<i>PXCpXY<sup>h</sup>6Y<sup>h</sup></i>	Neutral	-152.830	3.452	1	0.063	1
	Selection	-151.104				2.71
All Y loci	Neutral	-157.173	8.022	1	0.0046**	1
	Selection	-153.162				2.79
<i>X-linked loci</i>						
<i>CpXY<sup>h</sup>8X</i>	Neutral	-152.324	0.398	1	0.53	1
	Selection	-152.125				1.514
<i>CpXY<sup>h</sup>14X</i>	Neutral	-150.401	-1.392	1	1	1
	Selection	-151.097				0.240
<i>PXCpXY<sup>h</sup>14X</i>	Neutral	-151.513	-1.056	1	1	1
	Selection	-152.041				1.267
<i>PXCpXY<sup>h</sup>10X</i>	Neutral	-157.564	9.362	1	0.0022**	1
	Selection	-152.883				3.006
<i>CpXY<sup>h</sup>13X</i>	Neutral	n.d.	n.d.	n.d.	n.d.	1.
	Selection	n.d.				n.d.
<i>PXCpXY<sup>h</sup>6X</i>	Neutral	n.a.	n.a.	n.a.	n.a.	1
	Selection	n.a.				n.a.
All X loci	Neutral	-154.550	1.328	1	0.25	1
	Selection	-153.886				1.70

<sup>a</sup>ML = maximum likelihood; <sup>b</sup>K (gene) = selection parameter for target gene, neutral models set K = 1 for all loci. \*P<0.05, \*\*P<0.01

**Table S14.** Distribution of functional genes and pseudogenes in the X/Y-specific regions

<b>Region</b>	<b>Chromosome</b>	<b>Functional genes</b>	<b>Pseudogenes</b>	<b><i>P</i> value</b>
Entire SDR	X	84	14	<b>0.044</b>
	Y	72	24	
Stratum 1	X	40	11	0.134
	Y	34	17	
Stratum 2	X	27	2	0.094
	Y	20	6	
Collinear	X	17	1	0.743
	Y	18	1	