J Mol Evol (2015) 80:265–277
DOI 10.1007/s00239-015-9680-1

ORIGINAL ARTICLE

The Evolutionary Tempo of Sex Chromosome Degradation in *Carica papaya*

Meng Wu[1,2] · **Richard C. Moore**[1]

Received: 5 August 2014 / Accepted: 2 May 2015 / Published online: 19 May 2015

Abstract Genes on non-recombining heterogametic sex chromosomes may degrade over time through the irreversible accumulation of deleterious mutations. In papaya, the non-recombining male-specific region of the Y (MSY) consists of two evolutionary strata corresponding to chromosomal inversions occurring approximately 7.0 and 1.9 MYA. The step-wise recombination suppression between the papaya X and Y allows for a temporal examination of the degeneration progress of the young Y chromosome. Comparative evolutionary analyses of 55 X/Y gene pairs showed that Y-linked genes have more unfavorable substitutions than X-linked genes. However, this asymmetric evolutionary pattern is confined to the oldest stratum, and is only observed when recently evolved pseudogenes are included in the analysis, indicating a slow degeneration tempo of the papaya Y chromosome. Population genetic analyses of coding sequence variation of six Y-linked focal loci in the oldest evolutionary stratum detected an excess of nonsynonymous polymorphism and reduced codon bias relative to autosomal loci. However, this pattern was also observed for corresponding X-linked loci. Both the MSY and its corresponding X-specific region are pericentromeric where recombination has been shown to be greatly reduced. Like the MSY region, overall selective efficacy on the X-specific region may be reduced due to the interference of selective forces between highly linked loci, or the Hill–Robertson effect, that is accentuated in regions of low or suppressed recombination. Thus, a pattern of gene decay on the X-specific region may be explained by relaxed purifying selection and widespread genetic hitchhiking due to its pericentromeric location.

Keywords Plant sex chromosomes · Molecular evolution · Y degradation · Hill–Robertson effect

Introduction

Sex chromosomes are defined by the presence of a sex-determining region (SDR) containing sex-determining factor(s) that is recombinationally suppressed in the heterogametic sex (Charlesworth et al. 2005; Bachtrog et al. 2011; Gschwend et al. 2012a; Charlesworth 2013). Natural selection favors the suppression of recombination in the SDR by preventing the creation of allelic combinations of the sex-determining factors that would otherwise result in sexual neuters and/or hermaphrodites (Charlesworth et al. 2005; Bergero and Charlesworth 2009; Charlesworth and Mank 2010). In the absence of recombination, however, adaptive evolution in the non-recombining SDR of the sex-limited chromosome (the Y in XY systems or W in ZW system; here, we will focus on the Y chromosome) is impaired, leading to the accumulation of weakly deleterious mutations and/or repetitive elements. Ultimately, this can lead to gene loss and chromosomal degeneration (Bachtrog and Charlesworth 2002; Marais et al. 2008; Zhou and Bachtrog 2012a; b; Bachtrog 2013). Y chromosome degeneration is thought to be driven by the Hill–

Electronic supplementary material The online version of this article (doi:10.1007/s00239-015-9680-1) contains supplementary material, which is available to authorized users.

✉ Richard C. Moore
moorerc@miamiOH.edu

[1] Botany Graduate Program, Department of Biology, Miami University, Oxford, OH 45056, USA

[2] Present Address: Evolution, Ecology & Behavior Graduate Program, Department of Biology, Indiana University, Bloomington, IN 47405, USA

Robertson effect caused by genetic interference of selection at linked sites (Hill and Robertson 1966; McVean and Charlesworth 2000; Charlesworth 2012). The nature of the Hill–Robertson effect can depend on the evolutionary age and gene composition of the Y chromosome (see Bachtrog 2008, 2013).

The evolutionary tempo of Y chromosomal degeneration is best studied in young sex chromosome systems. For example, the neo-Y chromosome of *Drosophila miranda* has lost over 40 % of its gene content since it formed ~1 MYA from a Y-autosomal fusion (Bachtrog et al. 2008; Zhou and Bachtrog 2012b). The gene-rich, translocated autosomal region of the neo-Y became recombinationally isolated once it fused to the Y and rapidly degenerated, predominantly due to the effects of Muller's ratchet (Kaiser and Charlesworth 2010). However, a different evolutionary tempo of Y degeneration has been reported for the young sex chromosomes of the dioecious plant, *Silene latifolia*. Unlike the neo-sex chromosomes of *D. miranda*, the ~10-MY-old sex chromosomes in *S. latifolia* evolved *de novo* from a pair of autosomes that harbored a pair of sex-determining loci, a dominant male-fertility locus and a dominant female-sterility locus. While the *S. latifolia* Y chromosome shows evidence of degeneration, including an accumulation of repetitive and transposable elements, only ~20 % of the Y genes discovered possesses loss-of-function mutations (Bergero and Charlesworth 2011; Chibalina and Filatov 2011). It has been hypothesized that the tempo of plant Y chromosome degeneration may be slowed relative to animal Y chromosome systems due to enhanced purifying selection on deleterious mutations in the haploid phase of the plant life cycle (Bergero and Charlesworth 2011; Chibalina and Filatov 2011). Haploid selection is intensified in plants because thousands of plant genes are expressed in the male gametophyte (62 % of *Arabidopsis thaliana* genes [Honys and Twell 2004] and 95 % of surveyed *S. latifolia* genes [Chibalina and Filatov 2011]). Indeed, 99 % of sex-linked genes in *S. latifolia* are expressed in pollen. In contrast, haploid sperm in animals have limited gene expression (e.g., ~10 % of surveyed ESTs are expressed in human sperm [Ostermeier et al. 2002]) and have transcriptionally inert sex chromosomes (Handel et al. 1994).

Carica papaya is a polygamous tropical plant in the family *Caricaceae* with evolutionary young sex chromosomes (Yu et al. 2008a; Wang et al. 2012). Papaya has an active-Y sex determination system, with two slightly different Y chromosomes that differentiate males (XY) from hermaphrodites (XY^h). The male Y and hermaphroditic Y^h share 99 % DNA identity (Yu et al. 2008b) and we refer to both as the papaya Y in this paper when the distinction is not important. The non-recombining SDR of the papaya Y, also called the male-specific region of the Y (MSY), is a ~8.1 Mbp region comprising ~13 % of the Y chromosome and is somewhat larger than its X corresponding region [~5.3 Mbp; (Yu et al. 2009; Wang et al. 2012)]. Sequence comparisons between the MSY and its corresponding X region (the X-specific region) have revealed two evolutionary strata in the MSY of papaya defined by chromosomal inversions: stratum 1 has an estimated age of ~7.0 MY, stratum 2 has an estimated age of ~1.9 MY (Wang et al. 2012). In addition, there is a collinear region on the MSY that was not formed by chromosomal inversion but was incorporated in the MSY recently, potentially through the recruitment of sexually antagonistic mutations (Wang et al. 2012; Kirkpatrick and Guerrero 2014).

Cytogenetic analysis supports a pericentromeric location of the papaya MSY and the X-specific region, while combined physical and genetic mapping supports reduced recombination in the X-specific region, a characteristic of pericentromeric regions (Choo 1998; Chen et al. 2007; Zhang et al. 2008; Yu et al. 2009). Though the papaya X-specific region may recombine in females, it exhibits a much lower recombination rate than most of the genome due to its location (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 recombination associated with the Hill–Robertson interference effect, including reduced efficacy of purifying selection, elevated genetic hitchhiking, and reduced 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 showed an accumulation of repetitive sequences on the X-specific region compared to the homologous autosomal region in related taxa *Vasconcellea monoica*, which lacks sex chromosomes (Gschwend et al. 2012b; Wang et al. 2012). The inability to effectively purge repetitive elements may be due to Hill–Roberson interference effects on the X. Greatly reduced genetic diversity has also been observed at X alleles, consistent with selective sweeps acting on loci in the X-specific region (Weingartner and Moore 2012). Reduced recombination in the X-specific region would also enhance genetic hitchhiking effects leading to the accumulation of linked deleterious mutations on the X.

The ancestral sex-determining region in papaya, if pericentromeric, would have had reduced recombination relative to other chromosomal regions. The question we are addressing is whether Hill–Robertson effects were enhanced in the male-specific Y region following the inversion event leading recombination suppression in oldest evolutionary stratum resulting in Y-specific degeneration.

We took two approaches to study the tempo of papaya Y chromosome degeneration. First, we performed a comparative evolutionary analysis of X- and Y-linked genes and determined whether the signature of reduced adaptive evolution on the Y exists and whether it differs among differently aged evolutionary strata. Second, we used molecular population genetics to analyze the levels and patterns of coding sequence polymorphism of six focal X- and Y-linked loci found in a 0.8 Mb region of the MSY in the oldest evolutionary stratum 1 and its 1.4 Mb X region. We found a significant signature of Y chromosomal degeneration confined to stratum 1, but only when we include recent pseudogenes in the analyses; functional X/Y gene pairs do not show significant asymmetric protein evolution in any stratum. This suggests asymmetric evolution of functional Y-linked genes, even in the oldest evolutionary stratum, is undetectable or non-existent. We did find an accumulation of nonsynonymous polymorphism on Y-linked stratum 1 loci relative to autosomes consistent with relaxed purifying selection in this region; however, we found a similar, if attenuated, signature of degeneration on the corresponding X region. This may be explained by weaker purifying selection on the pericentromeric X due to enhanced Hill–Roberson interference effects in a region with low recombination.

Materials and Methods

Comparative Evolutionary Analyses of X- and Y-Linked Loci

The initial data set of X/Y gene pairs consisted of 70 X/Y gene pairs in the MSY including 50 functional X/Y gene pairs [their expression validated by RT-PCR (Wang et al. 2012)]. Comparative evolutionary analyses were performed only for gene pairs with sufficient coverage in the outgroup, *V. monoica*. After removing gene pairs without sufficient outgroup sequence, including four functional gene pairs and 11 pseudogene pairs, the dataset consisted of 55 X/Y gene pairs, with 46 functional gene pairs, 5 Y-pseudogene pairs (functional X-allele + Y-pseudogene), 3 X-pseudogene pairs (functional Y-allele + X-pseudogene), and 1 X/Y-pseudogene pair. It was inferred that the pseudogenes only recently lost their function, as their sequences are otherwise intact and comparable (Wang et al. 2012); as such, we included pseudogenes in our analyses in addition to conducting analyses using only functional X/Y gene pairs. 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 coding regions were identified from previously published gene annotations (Wang et al. 2012). X/Y gene coding sequences were ascertained by BLAST analysis against the papaya expressed sequence tag dataset (EST; http://www.ncbi.nlm.nih.gov/nucest) and 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 % (77,057 bp) 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 % (48,356 bp) of the predicted X/Y paired coding sequence including pseudogene pairs. Complete outgroup sequence was difficult to obtain due to the many gaps and ambiguous sequences in the whole-genome shotgun sequence of *V. monoica*.

The codon substitution model (CODEML) in PAML 4.7 (Yang 1997, 2007) was used to discern asymmetric protein evolutionary rates in individual X/Y gene pairs along the MSY as well as concatenated dataset of genes by stratum. An unrooted phylogenetic tree with three branches consisting of X/Y gene pairs and their homologous sequences from *V. monoica* was used. A model (model = 2) that assumes X and Y branches have the same ratio of nonsynonymous to synonymous substitutions ($\omega_{Vm} \neq \omega_X = \omega_Y$) was compared to a free-ratio model (model = 1) in which every branch has a unique ratio ($\omega_{Vm} \neq \omega_X \neq \omega_Y$). Other than model specification, default parameters in CODEML for branch model comparisons were used. Significant differences between the two models were assessed using a likelihood ratio test (M1 vs. M2).

The number of chromosome-specific mutations was determined using DnaSP v5 (Librado and Rozas 2009) by comparing the number of synonymous or nonsynonymous substitutions between X/Y-linked locus (X + Y) and the numbers of substitution of each X/Y-linked locus relative to the outgroup *V. monoica* (X + V, Y + V). The number of substitutions specific to the X was 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 substitutions was calculated by: Y = [(X + Y) + (Y + V) − (X + V)]/2. Numbers of synonymous or nonsynonymous substitutions between X- and Y-linked alleles for each evolutionary stratum were compared using a one-sided paired *t* test in JMP Pro 10 (SAS Institute Inc.) Although PAML estimates of d*N***N* and d*S***S* are similar to those we determined here, we chose to use these values, as PAML estimates are dependent on the model used and we had low power to distinguish between models for individual loci due to low divergence between loci. DnaSP v5 was also used to estimate synonymous

nucleotide divergence (K_s) and nonsynonymous nucleotide divergence (K_a) per site relative to the outgroup *V. monoica* for concatenated datasets of all X/Y-linked genes on stratum 1, stratum 2, and the collinear region.

Population Genetic Analyses of Focal X- and Y-linked Stratum 1 Loci

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 S1). The focal Y-linked genes are located in a 0.8 Mb region of stratum 1. The X homologs, among which three are predicted pseudogenes, are distributed across a 1.4 Mb region very near the corresponding region of Y-specific heterochromatic knob 4 (Fig. 1b). Assignment of X- and Y-linked alleles was based on the published sequences of the X and Y^h (Wang et al. 2012).

Forty male and two hermaphrodite papayas were used for population genetic analyses of focal Y and autosomal loci and 24 female papaya individuals were used for focal X-loci amplification in the population analyses (Table S2). Individuals were evenly sampled from natural populations found in five geographically dispersed regions of Costa Rica (Brown et al. 2012). These populations are not strongly genetically differentiated based on microsatellite analysis (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. *Jacaratia dolichaula* was used as outgroup for some autosomal loci for which there was no homologous sequence in the *V. monoica* sequence database. Tajima's relative rate test (Tajima 1993) indicated no significant differences in evolutionary rates between *V. monoica* and *J. dolichaula* relative to *C. papaya* (Table S3).

Genomic DNA was isolated from leaf tissues using the DNeasy Plant Mini Kit (Qiagen, Valencia, CA). The published sequences of the X and Y^h (Wang et al. 2012) were used to design X-specific and Y-specific PCR primers adjacent to the boundary of target exon regions using Primer3 (Rozen and Skaletsky 2000). For polymorphism analyses,

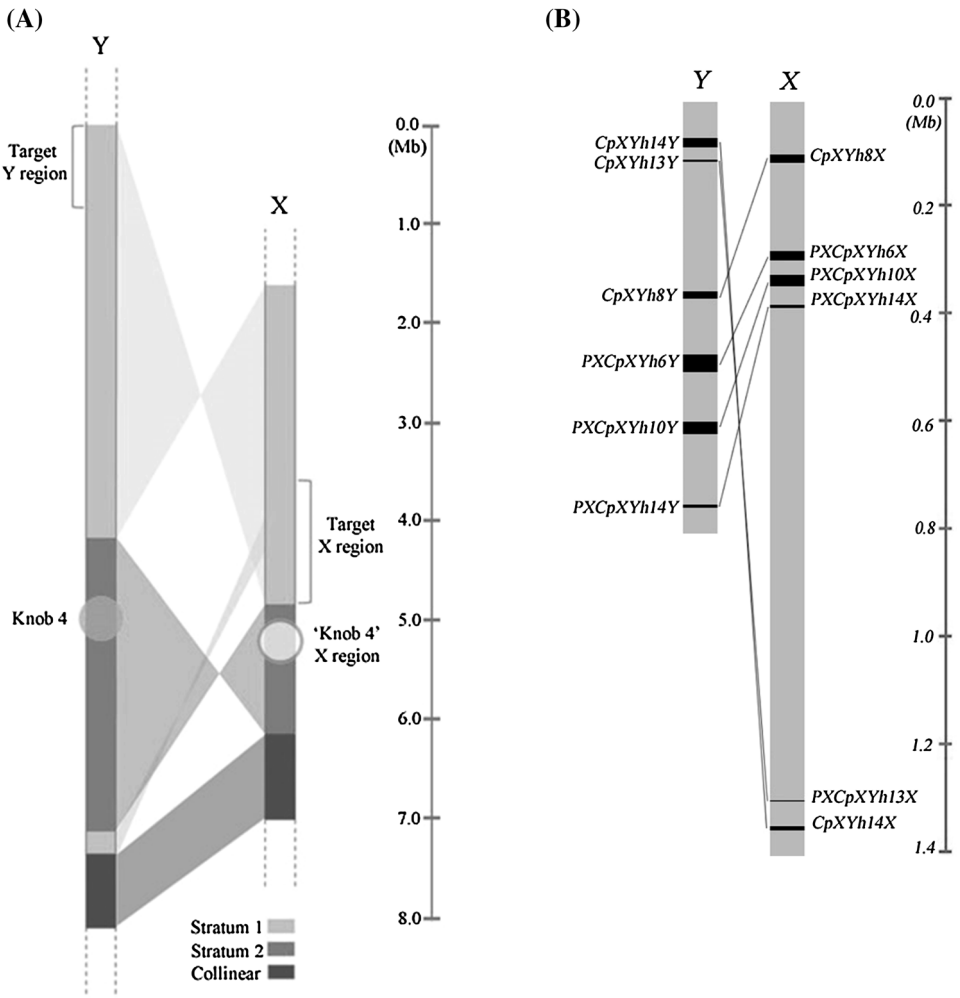

Fig. 1 Diagram of the sex chromosome regions targeted in this study. **a** The X/Y gene pairs used for comparative analyses are dispersed in the three evolutionary strata of the SDR on the X and Y chromosomes of papaya. The fourth heterochromatic knob in the Y region is indicated by a *solid circle* and its estimated corresponding position in the X region is indicated by an *empty circle*. The centromere is in or near Knob 4 (Zhang et al. 2008). The target X/Y regions for population genetic analyses are indicated by *brackets*. **b** Relative positions of the six focal loci in stratum 1 are used for population genetic analyses. Location of each target X/Y gene pairs is indicated by *black bars*. The width of each bar corresponds to gene size. The *solid black lines* link X/Y gene pairs

Y-linked alleles were amplified from males using Y-specific primers, while X-linked alleles were amplified from females using X-specific primers (Table S4 and S5). Due to the limited size range of cycle sequencing (< 1200 bp), large focal loci were divided into smaller blocks for amplification. PCR was performed with GoTaq® Colorless Master Mix (Promega, Madison, WI). The PCR reaction program consisted of an initial denaturation time of 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 a final extension for 5 min at 72 °C. Single-band products were cleaned using a combination of 5 U exonuclease I and 0.5 U shrimp alkaline phosphatase at 37 °C for 40 min followed by enzyme inactivation 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) and followed by a post-sequencing clean-up step using ethanol precipitation. Precipitated sequencing reactions were resuspended in Hi-Di formamide before loading them into an ABI 3130xl or 3730 Genetic Analyzer (Applied Biosystems). Both forward and reverse strands were sequenced and sequencing results were assembled and aligned using BioEdit and BioLign software (Hall 1999). Polymorphic sites were confirmed against the original ABI chromatogram file. Homologous outgroup sequences and coding sequence assignments were subsequently added to alignments. Final sequence alignments may be obtained from GenBank (Accession # KR698941–KR699604).

The number of synonymous and nonsynonymous polymorphic and divergent substitutions in the exon regions of the focal loci was determined using DnaSP v5 (Librado and Rozas 2009). The ratio of nonsynonymous (P_A) to synonymous (P_S) polymorphism and the ratio of amino acid (D_A) to synonymous (D_S) divergence were calculated for both individual loci and for concatenated protein coding sequences of the six focal X/Y gene pairs. Polymorphism and divergence data were determined from individual and concatenated coding sequences from a set of 38 autosomal genes dispersed across the genome of papaya (Table S6).

Nucleotide diversity at silent sites (π_{sil} and θ_{sil}) of the six focal X/Y gene pairs was estimated using DnaSP v5 (Librado and Rozas 2009; Watterson 1975). Neutrality 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. The compound tests *DH* and *DHEW* combining individual tests based on site frequency spectra (*D* and *H*) are less sensitive to underlying demography and were implemented to discern signatures of positive selection (Zeng et al. 2007). Significance of site frequency tests was determined using standard coalescence simulations without recombination. Linkage disequilibrium (LD) across the X-specific or Y-specific regions was determined by concatenating aligned X- or Y-specific alleles for individual loci and using DnaSP v5 to calculate r^2 between pairs of parsimoniously informative sites with significance determined by Fisher's exact test. GC content and three estimates of codon usage bias, the scaled $\chi 2$, 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 datasets.

Polymorphism patterns of silent sites on the six focal X/Y gene pairs were compared to 25 autosomal loci (a subset of the 38 autosomal loci used above and noted in Table S6) using the Maximum Likelihood Hudson, Kreitman, and Aguade (MLHKA) test (Wright and Charlesworth 2004). Neutrality tests (Tajima's *D* and/or Fay and Wu's *H*) based on the site frequency spectrum detected no signatures of selection on any of these 25 autosomal genes (Table S7) which were defined as neutrally evolving ($k = 1$, where k is the selection parameter; Wright and Charlesworth 2004) for the all MLHKA models tested. We compared models of selection at individual and concatenated X- and Y-linked loci to a neutral model where all loci were considered neutrally evolving ($k = 1$). We used a starting value of 100 for *T* (divergence time of the two species in 2Ne generations) and a Markov chain length of 100,000. The selection parameter *k* in the test measures the excess ($k > 1$) or deficiency ($k < 1$) of polymorphism relative to divergence. 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.

Results

Faster Protein Sequence Evolution on the Y Versus the X

Few individual X/Y gene pairs, including those with recently evolved pseudogenes, showed a significant asymmetric evolutionary pattern using the CODEML program in PAML (Table S8). Lack of asymmetric protein evolution by tests on individual X/Y gene pairs may be due to a lack of differential selective forces acting on individual X/Y genes, or a paucity of fixed differences between X and Y chromosomes due to their recent divergence and to the limited numbers of codons in the individual genes analyzed. To address this concern, all functional and pseudogene sequences from each stratum of the SDR (including stratum 1, stratum 2, and the collinear region) were concatenated and patterns of asymmetric evolution were analyzed by stratum (Table 1). The concatenated stratum 1 dataset significantly fits the free-ratio model over a model

where X and Y rates are constrained to be equal ($P = 0.009$), showing an elevated ω on the Y ($\omega = 0.4078$) relative to its X counterpart ($\omega = 0.2640$). 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 ($P = 0.74$) or the collinear region ($P = 0.39$). If the analysis is restricted to concatenated data sets of only functional X/Y gene pairs, significant asymmetric protein evolution is no longer detected in Stratum 1 (Table S9).

To confirm that it is an excess of nonsynonymous mutations and not a deficit of synonymous substitutions on the Y that drives the higher ω on the Y relative to the X for the X/Y gene pair dataset with pseudogenes, the number of synonymous and nonsynonymous substitutions at Y-linked loci versus X-linked loci were compared using a matched-pair analysis. Across the entire SDR region, equivalent numbers of synonymous mutations were found between X- and Y-linked loci ($P = 0.1244$; Fig. 2a). However, there is a significant excess of nonsynonymous mutations at Y-linked loci compared to X-linked loci ($P = 0.0016$; Fig. 2b). The excess of nonsynonymous substitutions is localized to 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). Similar analysis restricted to functional X/Y gene pairs also detects significantly elevated nonsynonymous mutations on the Y, and specifically in stratum 1. However, synonymous substitutions are also significantly Y-biased in this stratum, consistent with the lack of elevated d*N*/d*S* for functional Y-loci determined in the PAML analysis (Fig S1).

In order to assess the evolutionary tempo of degeneration after recombination suppression, the ratio of nonsynonymous to synonymous divergence (K_a/K_s) for X- and Y-linked loci relative to the outgroup *V. monoica* was estimated from the strata-specific concatenated datasets (Fig. 3). K_a/K_s increased by age of stratum for both X- and Y-linked loci when strata-specific genes are analyzed as a group, the only difference being a higher K_a/K_s for the Y in the oldest evolutionary stratum 1. Thus functional constraint, as inferred from K_a/K_s, is reduced for both X- and Y-linked loci in older evolutionary strata, with the greatest difference between X and Y in stratum 1. The same pattern is also observed when we restrict our analysis to the concatenated data sets of only functional X/Y gene pairs (Fig S2).

>Finally, the distribution of previously reported pseudogenes was re-evaluated for X and Y chromosomes according to evolutionary strata (Table 2; data from Wang et al. 2012). There are 24 predicted pseudogenes on the MSY (25 % of 96) and 14 predicted pseudogenes in the corresponding X region (14 % of 98). Across the SDR, there are slightly significantly higher proportions of pseudogenes on the Y than on the X (Fisher's exact test; $P = 0.044$). This difference cannot be attributed to a single stratum; however, it can be explained by an excess of total Y pseudogenes in the combined strata 1 and 2 ($P = 0.033$). Furthermore, there is a significantly greater proportion of pseudogenes (both X- and Y-linked combined) in the oldest stratum 1 than in either stratum 2 ($P = 0.048$) or the collinear region ($P = 0.003$).

Elevated Nonsynonymous Polymorphism and Divergence in Both the Y- and X-Specific Regions of Stratum 1 Relative to Autosomal Loci

Significant LD exists across the entire 1.4 Mb region of the X chromosome and across the 0.8 Mb target region of Y, consistent with reduced (X) and/or suppressed (Y) recombination in these regions (Fig. S3). To test whether there is reduced efficacy of purifying selection on segregating polymorphism in X- and/or Y-linked loci in stratum 1, the total number of polymorphic (P) and divergent (D) nonsynonymous and synonymous mutations were compared for the concatenated datasets of the six focal X/Y gene pairs and autosomal loci (Table 3). An excess of nonsynonymous

Table 1 PAML-based tests of evolutionary rate differences between concatenated datasets of 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[a]	χ^2	d*f*	*P*	Free ω[b]
SDR	15,178	Model = 2	−77,080.264	5.500	1	0.019*	X: 0.2958
		Model = 1	−77,077.514				Y: 0.4222
1	6833	Model = 2	−35,900.841	6.883	1	0.009**	X: 0.2640
		Model = 1	−35,897.399				Y: 0.4078
2	5092	Model = 2	−25,242.898	0.111	1	0.739	X: 0.3101
		Model = 1	−25,242.842				Y: 0.3489
Collinear	4115	Model = 2	−19,781.612	0.738	1	0.390	X: 0.6176
		Model = 1	−19,781.243				Y: 0.2414

[a] LnL = log likelihood

[b] Free ω = the ratio of nonsynonymous/synonymous substitution rates (d_N/d_S) in each branch under model 2

* $P < 0.05$; ** $P < 0.01$

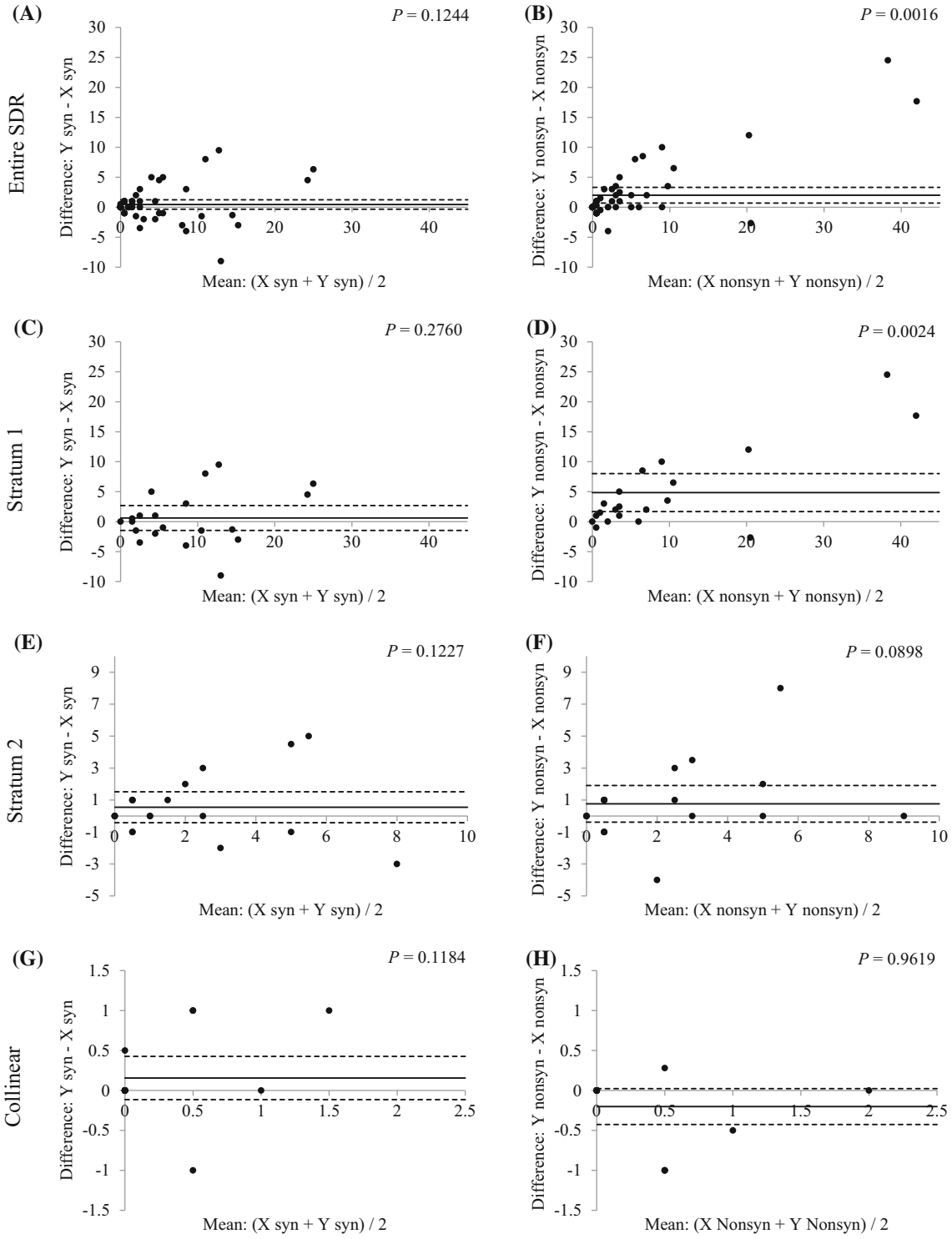

Fig. 2 Matched-pair analysis comparing the difference in synonymous or nonsynonymous mutations between X and Y alleles versus mean number of differences between X and Y alleles for **a–b** the entire SDR region, **c–d** stratum 1, **e–f** stratum 2, or **g–h** the collinear region. *Solid horizontal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

polymorphic and divergent mutations was found at the focal X/Y-linked loci relative to the autosomal loci (Table 3). The D_A/D_S ratio on the Y (1.36) and the X (1.29) is much greater than that on the autosomes (0.60). The three X predicted pseudogenes have a slightly higher D_A/D_S ratio (1.31) than the three functional X-loci (1.27). The P_A/P_S is similarly

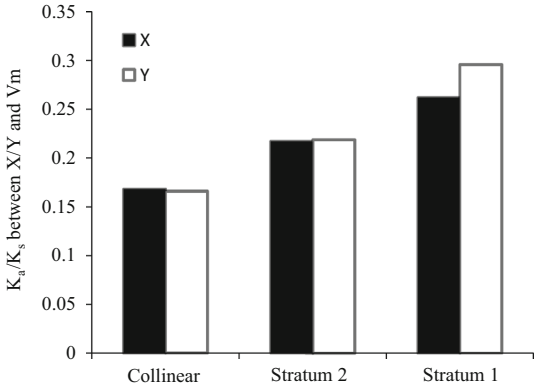

Fig. 3 Plot of K_a/K_s of concatenated sequences grouped by evolutionary strata. Divergence values used to determine K_a/K_s are relative to orthologous sequences in *V. monoica*

elevated on the Y (1.75) and the X (2.00) relative to autosomal loci (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). The neutrality index (P_A/D_A:P_S/D_S) is greater than 1 for both X- and Y-loci and is less than 1 for autosomal loci (0.74). A neutrality index of 1 is consistent with a model of strict neutral evolution, while a value greater than 1 is consistent with an excess of intraspecific nonsynonymous variation (Rand and Kann 1996).

As relaxed purifying selection can also lead to reduced codon bias, three measures of codon bias, the scaled $\chi 2$, the ENC, and the CBI, were estimated from the concatenated datasets (2801 Y-linked codons, 2618 X-linked codons, and 1298 autosomal codons). All three measures have the same pattern, with the signatures of codon bias greatest for autosomal loci (as measured by a greater scaled $\chi 2$, a low ENC and a CBI closer to 1), intermediate for focal X-linked loci, and lowest for focal Y-linked loci (Table 4). Recombination suppression can suppress gene conversion leading to reduced GC bias and a signature of reduced codon bias (Marais 2003). Focal X- and Y-linked loci have similar GC content to autosomal loci in coding sequences overall and at third codon positions, suggesting that the differences in codon bias may be due to reduced purifying

Table 2 Distribution of functional genes and pseudogenes in the X/Y-specific regions

Region	Chromosome	Functional genes[a]	Pseudogenes[a]	$P(Y\psi > X\psi)$
Entire SDR	X	84	14	0.044
	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	
Stratum 1 &	X	67	13	0.033
Stratum 2	Y	54	23	

[a] Numbers of functional genes and pseudogenes from Wang et al. (2013)

Table 3 Coding sequence polymorphism and divergence in X/Y-linked loci

Genes	Class	*A*	*S*	*A*/*S*	P_A/D_A:P_S/D_S
All 5 X-linked	Polymorphism	10	5	2	1.55
	Divergence[a]	226	175	1.29	
2 functional X-linked	Polymorphism	5	3	1.67	1.31
	Divergence[a]	117	92	1.27	
3 pseudo X-linked	Polymorphism	5	2	2.5	1.91
	Divergence[a]	109	83	1.31	
All 6 Y-linked	Polymorphism	7	4	1.75	1.28
	Divergence[a]	288	211	1.36	
Autosomal ($n = 34$)	Polymorphism	5	11	0.45	0.74
	Divergence[b]	135	220	0.60	

A number of nonsynonymous mutations, *S* number of synonymous mutations, P_A/D_A:P_S/D_S ratio of nonsynonymous polymorphism and divergence to synonymous polymorphism and divergence, also referred to as the neutrality index (Rand and Kann 1996)

[a] Outgroup *V. monoica*

[b] Outgroup *V. monoica* or *J. dolichaula*

selection against unpreferred codons and not due to suppressed gene conversion.

A Signature of Hitchhiking on the Target X Region

As with a previous study (Weingartner and Moore 2012), contrasting patterns of nucleotide diversity were found on X- and Y-linked loci (Table 5). The average π_{sil} is 0.00091 ± 0.00033 (SE) for the Y-linked loci and 0.00034 ± 0.00015 (SE) for the X-linked loci (Table 5). The average θ_w values at silent sites are 0.00084 ± 0.00027 (SE) for the Y-linked loci and 0.00093 ± 0.00042 (SE) for the X-linked loci (Table 5). Both π and θ_w show elevated polymorphism on the Y. The Y:X ratio of average π_{sil} is 2.64 and the Y:X ratio of average θ_{sil} is 0.90, while the neutral expectation of Y:X polymorphism is 0.33 based on the relative effective population sizes of the sex chromosomes.

Furthermore, the site frequency spectrum is skewed for X- versus Y-linked loci. All X-linked alleles have negative Tajima's *D* and Fay and Wu's *H* values, consistent with an excess of rare alleles and high-frequency derived alleles, respectively. According to the compound *DH* test, the X-linked loci vary significantly from neutral expectations, consistent with positive selection. The signature of a selective sweep was found in nearly all of the six focal X-loci across the target X region. However, for all the Y-linked loci, both *D* and *H* values tend to be near zero, indicating no 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 S10). For all the Y-linked genes, the selection parameter *k* is larger than 1 and this pattern is significant for $CpXY^h13Y^h$ ($P = 0.0014$) 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, $PXCpXY^h10X$, which shows a significant elevated polymorphism ($P = 0.0022$). The concatenated dataset of Y has a significant *k* value of 2.79 ($P = 0.0046$) while the X-concatenated dataset has a non-significant *k* of 1.70 ($P = 0.25$). The significant large *k* value on the Y supports an increased nucleotide variation relative to divergence on the Y than expected.

Discussion

The Evolutionary Tempo of Y Chromosomal Degradation in Papaya

Non-recombining Y chromosomes tend to accumulate deleterious mutations and have a reduced adaptive evolutionary rate compared to the recombining regions of the genome (Bachtrog 2013). In animal species, like *D. miranda*, the degeneration of the neo-Y chromosome happened very rapidly. The 1-MY-old *D. miranda* 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). Partial Y degeneration has been documented in the young Y chromosomes of plant species, as described in *S. latifolia* (Marais et al. 2008); however, degeneration is not as extensive as it is found in *D. miranda*. At most 20 % of the genes on this 10-MY-old Silene Y chromosome are impaired or have lost their functions (Bergero and Charlesworth 2011; Chibalina and Filatov 2011). Considering the shorter generation time in *D. miranda*, a more accurate evaluation of the age of the Y chromosome should be based on rates of synonymous protein divergence between X- and Y-linked genes (Bachtrog 2011). The synonymous protein X/Y divergence in *S. latifolia* is 2.5–10 % and for *D. miranda* is 1–7 %; thus, the Y chromosome in *S. latifolia* is only slightly older than that in *D. miranda* (Bachtrog 2011). It was proposed that the slower pace of 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 (Bergero and Charlesworth 2011; Chibalina and Filatov 2011).

The non-recombining SDR of papaya 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 ± 0.7 %, respectively) and one collinear region (Avg. $K_s = 0.7 \pm 0.2$ %; Wang et al. 2012). According to the synonymous protein X/Y divergence, the age of the sex chromosomes in papaya is comparable to that of *S. latifolia*. We found evidence of a faster nonsynonymous substitution rate consistent with reduced efficacy of purifying selection in the oldest

Table 4 Condon usage bias estimates for concatenated gene datasets

Gene dataset	n^a	Sites	*S*	No. codons	Scaled χ^{2b}	ENC	CBI	G+C total	G+C3
Autosomal	84	3859	15	1298.143	0.239	50.587	0.321	0.429	0.327
X	48	7824	19	2617.5	0.205	51.041	0.29	0.417	0.323
Y	42	8425	16	2800.928	0.187	51.587	0.273	0.424	0.323

ENC effective number of codons, CBI codon bias index (0 = no bias; 1 = complete bias), GC content at coding positions, GC content at 3rd codon positions

[a] Number of alleles

[b] Scaled χ^2, a measure of the divergence from equal codon usage (0 = no divergence)

Table 5 Tests of selection based on the site frequency spectrum (*D* and *H*), haplotype frequency spectrum, or combinations thereof (*DH*)

Locus	*n*	*S*	π_{sil}	θ_{sil}	*D*	H^a	*DH* (*P*)
Y-linked focal loci							
CpXY^{h}8Y^h	42	7	0.00051	0.00042	1.07	0.57	0.77
CpXY^{h}14Y^h	42	6	0.00048	0.00063	−0.67	−0.56	0.13
PXCpXY^{h}14Y^h	42	8	0.00035	0.00048	−0.80	0.43	0.42
PXCpXY^{h}10Y^h	42	10	0.00059	0.00063	0.12	0.63	0.39
CpXY^{h}13Y^h	42	6	0.00251	0.00217	0.41	−0.71	0.30
PXCpXY^{h}6Y^h	42	6	0.00097	0.00072	0.93	0.47	0.93
X-linked focal loci							
CpXY^{h}8X	48	17	0.00044	0.00120	−2.04*	−7.05**	≪0.001
CpXY^{h}14X	48	1	0.00006	0.00015	−0.87	n.a.	n.a.
PXCpXY^{h}14X	48	15	0.00035	0.00095	−1.97*	−8.74**	≪0.001
PXCpXY^{h}10X	48	29	0.00084	0.00235	−2.15*	−6.59**	≪0.001
CpXY^{h}13X	48	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
PXCpXY^{h}6X	48	0	0	0	n.a.	n.a.	n.a.

n.d. not determined (*CpXY^{h}13X* fails to be amplified), *n.a.* not applicable data (no polymorphisms), *n* number of alleles, *S* number of segregating sites, *D* Tajima's *D* at silent sites, *H* Fay & Wu's *H* at silent sites

[a] *P* value estimated on coalescent simulations

* $P < 0.05$; ** $P < 0.01$

evolutionary stratum of papaya, but only when we include recently evolved pseudogenes in our analysis. Thus, the pattern of asymmetric protein evolution on the Y in stratum 1 is influenced most strongly by Y-linked pseudogenes and not functional gene pairs. Our analysis of the distribution of pseudogenes across the MSY strata fails to detect a strong temporal gradient of Y-specific gene loss, as a significant elevation in the number of Y pseudogenes is observed only in the combined datasets of stratum 1 and stratum 2. Furthermore, the proportion of Y pseudogenes (25 %) on the papaya MSY is comparable to what is observed in *S. latifolia* (~20 %). While we do see a trend of increasing K_a/K_s in Y-linked alleles in older evolutionary strata (seen both with and without pseudogenes), a similar trend is observed for X-linked alleles, though to a lesser extent in the oldest evolutionary stratum. Therefore, we found that the evolutionary tempo of Y chromosomal degeneration relative to the X in papaya is slow, as even the oldest MSY stratum shows only a weak signature of asymmetric evolution when pseudogenes are included in the analysis and no signature when only functional X/Y gene pairs are considered. The tempo of Y degeneration of the papaya Y is on par, if not slower than, that observed for *S. latifolia*, consistent with the hypothesis that haploid purifying selection may constrain the extent of degeneration of plant sex chromosomes.

The slow degeneration tempo of the Y chromosome in papaya may be also explained by other factors. For example, on the neo-Y of *D. miranda*, selection for masculinizing, beneficial mutations on the Y simultaneously fixes linked deleterious mutations (Zhou and Bachtrog 2012b). However, we found no evidence showing a selective sweep on the MSY region (also see Weingartner and Moore 2012). The tempo of Y chromosome degeneration is also negatively correlated with the number of functional genes (Bachtrog 2008, 2013). Compared to the large gene-rich neo-Y chromosome in *D. miranda*, the MSY of papaya is confined to a small pericentromeric region, a region that is characterized by reduced recombination rate and low gene density to start. Thus, an alternative hypothesis to explain the slower tempo of degeneration of the young papaya Y chromosome is that it may be behave as other old Y chromosomes that contain a small number of active genes and that degenerate at a slow pace.

It may also be possible that Y degeneration in papaya has occurred through reduced expression of Y-linked genes, which is not reflected in our studies of protein sequence evolution. 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 nonsynonymous mutations (Zhou and Bachtrog 2012b). Formation of heterochromatin and accumulation of mutations in promoter or intronic regulatory regions can 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.

Reduced Efficacy of Purifying Selection on Both the MSY and X-Specific Regions of the Papaya Sex Chromosomes

The centromere is a recombination cold spot in a wide range of eukaryotic genomes (Choo 1998). The centromeric regions of all nine major linkage groups in papaya have reduced recombination, including the pericentromeric region containing the X-specific region of the sex chromosomes (Chen et al. 2007; Yu et al. 2009). Furthermore, recombination frequency of the X-specific region is theoretically two-thirds that of autosomes, as recombination suppression between the X and Y forces crossovers in the X-specific region to be restricted to XX females. The efficacy of natural selection may therefore be reduced in the X-specific region as it is typically found in genomic regions with low recombination (Betancourt et al. 2009). For example, in *Drosophila melanogaster*, low recombining regions near the centromeres, telomeres, or in the dot chromosome are characterized by reduced codon bias, interpreted as relaxed selection for optimized translation (Kliman and Hey 1993). Because of its small size, the dot chromosome in *D. melanogaster* lies near the recombination-suppressing centromere and suffers an enhanced Hill–Robertson interference effect, contributing to the accumulation of deleterious mutations (Haddrill et al. 2007; Charlesworth et al. 2009; Arguello et al. 2010).

Our observations of elevated LD in both the MSY and X region are evidence of suppressed and/or reduced recombination in these regions and both the MSY and X-specific region exhibit signatures of relaxed purifying selection. First, there are elevated A/S ratios at focal Y- and X-linked loci in stratum 1 relative to autosomal loci. The neutrality index (P_A/D_A:P_S/D_S) is greater than 1 for functional focal X-linked loci, similar to the functional focal Y-linked loci, suggesting an excess of intraspecific nonsynonymous variation due to reduced efficacy of purifying selection for both X- and Y-linked loci (Rand and Kann 1996). Second, we found a positive correlation between the relaxation of functional constraint, as evidence by K_a/K_s, and the age of strata for both the MSY and X-specific region (Fig. 3). Finally, there is reduced codon bias of focal X- and Y-linked loci relative to the autosomal loci. Codon bias seems most relaxed for Y-linked loci, and this does not appear to be due to the lack of gene conversion in a region of suppressed recombination as GC content is similar between X, Y and autosomal focal loci. However, further analyses of the papaya transcriptome are needed to define optimal codons and assess the level of purifying selection acting on codons in papaya in order to fully confirm this observation.

The signature of degeneration on the X-specific region is a unique feature of the papaya sex chromosomes. It is also possible that degeneration on the X may be masking the signature of degeneration on the MSY in comparative analyses. For example, the lack of a strong signature of asymmetric protein evolution on the Y in even the oldest evolutionary strata relative to the X may be attenuated by the elevated K_a/K_s ratios found in X-linked alleles in this region. Regardless, the signatures of reduced purifying selection are still stronger on the recombinationally suppressed Y relative to the X. Although there is a relatively high percentage of pseudogenes on the X-specific region, particularly in stratum 1 and 2 inversion regions (19 % pseudogenes), there are proportionally more Y-linked pseudogenes (43 % pseudogenes in stratum 1 and 2). Also, the X-specific region has expanded in size by transposable element accumulation compared with the orthologous autosomal region in the related monoecious taxa *V. monoica* (Gschwend et al. 2012b). The accumulation of transposable elements is more profound on the non-recombining Y (Yu et al. 2007); nonetheless, it is intriguing to consider that the proliferation of transposable elements on the X is affected by reduced efficacy of purifying selection due to the reduced recombination rate in this region.

The reduced recombination rate on the X region is evidenced by the broad LD interval we observed at our X-linked loci. Lower recombination rates on the X would enhance the hitchhiking effect of positive selection that is evident across the 1.4 Mbp focal X region. The accumulation of nonsynonymous mutations at focal X-linked loci may therefore be affected by the selective sweep detected in this region (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 Y chromosomes (Charlesworth 1996; Bachtrog 2004; Zhou and Bachtrog 2012b). However, in papaya, the signature of a selective sweep is limited to the X, whereas the Y has the extended signature of significantly elevated polymorphism relative to divergence, possibly due to balancing or frequency-dependent selection, though this conclusion needs further investigation. The selective sweep in papaya spans the entire 1.4 Mbp X region containing our focal loci, which includes more than 20 genes. Thus, many detrimental mutations are probably fixed because of the widespread hitchhiking effect, which explains the degeneration pattern at this targeted X region.

Acknowledgments We thank two anonymous reviewers for their helpful comments in preparing this manuscript. This work was supported by the National Science Foundation (DBI-0922545) to RCM and by the Department of Botany at Miami University through Academic Challenge grants to MW. We are greatly indebted to Oscar Rocha at Kent State University for his help during the specimen collecting process in Costa Rica. We would also like to thank our collaborators Ray Ming (University of Illinois at Urbana-Champaign) and Qingyi Yu (Texas A&M University) for providing papaya sequence information and cultivar tissues.

References

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

Bachtrog D (2004) Evidence that positive selection drives Y-chromosome degeneration in *Drosophila miranda*. Nat Genet 36:518–522

Bachtrog D (2008) The temporal dynamics of processes underlying Y chromosome degeneration. Genetics 179:1513–1525

Bachtrog D (2011) Plant sex chromosomes: a non-degenerated Y? Curr Biol 21:R685–R688

Bachtrog D (2013) Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat Rev Genet 14:113–124

Bachtrog D, Charlesworth B (2002) Reduced adaptation of a non-recombining neo-Y chromosome. Nature 416:323–326

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, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice WR, Valenzuela N (2011) Are all sex chromosomes created equal? Trends Genet 27:350–357

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 CW Jr, Walsh JB (1988) Effects of linkage on rates of molecular evolution. Proc Natl Acad Sci USA 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

Charlesworth B (1996) The evolution of chromosomal sex determination and dosage compensation. Curr Biol 6:149–162

Charlesworth B (2012) The effects of deleterious mutations on evolution at linked sites. Genetics 190:5–22

Charlesworth D (2013) Plant sex chromosome evolution. J Exp Bot 64:405–420

Charlesworth D, Mank JE (2010) The birds and the bees and the flowers and the trees: lessons from genetic mapping of sex determination in plants and animals. Genetics 186:9–31

Charlesworth D, Charlesworth B, Marais G (2005) Steps in the evolution of heteromorphic sex chromosomes. Heredity 95:118–128

Charlesworth B, Betancourt AJ, Kaiser VB, Gordo I (2009) Genetic recombination and molecular evolution. Cold Spring Harb Symp Quant Biol 74:177–186

Chen CX, Yu QY, Hou SB, Li YJ, Eustice M, Skelton RL, Veatch O, Herdes RE, Diebold L, Saw J, Feng Y, Qian WB, Bynum L, Wang L, Moore PH, Paull RE, Alam M, Ming R (2007) Construction of a sequence-tagged high-density genetic map of papaya for comparative structural and evolutionary genomics in brassicales. Genetics 177:2481–2491

Chibalina MV, Filatov DA (2011) Plant Y chromosome degeneration is retarded by haploid purifying selection. Curr Biol 21:1475–1479

Choo KHA (1998) Why is the centromere so cold? Genome Res 8:81–82

Fay JC, Wu CI (2000) Hitchhiking under positive Darwinian selection. Genetics 155:1405–1413

Gschwend AR, Weingartner LA, Moore RC, Ming R (2012a) The sex-specific region of sex chromosomes in animals and plants. Chromosome Res 20:57–69

Gschwend AR, Yu QY, Tong EJ, Zeng FC, Han J, VanBuren R, Aryal R, Charlesworth D, Moore PH, Paterson AH, Ming R (2012b) Rapid divergence and expansion of the X chromosome in papaya. Proc Natl Acad Sci USA 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. Nucl Acids Symp Ser 41:95–98

Handel MA, Park C, Kot M (1994) Genetic control of sex-chromosome inactivation during male meiosis. Cytogenet Cell Genet 66:83–88

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

Honys D, Twell D (2004) Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. Genome Biol 5:R85

Kaiser VB, Charlesworth B (2010) Muller's ratchet and the degeneration of the *Drosophila miranda* neo-Y chromosome. Genetics 185:U339–U491

Kirkpatrick M, Guerrero RF (2014) Signatures of sex-antagonistic selection on recombining sex chromosomes. Genetics 197:531–541

Kliman RM, Hey J (1993) Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol Biol Evol 10:1239–1258

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Marais G (2003) Biased gene conversion: implications for genome and sex evolution. Trends Genet 19:330–338

Marais GAB, 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

McVean GAT, Charlesworth B (2000) The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. Genetics 155:929–944

Ostermeier GC, Dix DJ, Miller D, Khatri P, Krawetz SA (2002) Spermatozoal RNA profiles of normal fertile men. Lancet 360:772–777

Rand DM, Kann LM (1996) Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from *Drosophila*, mice, and humans. Mol Biol Evol 13:735–748

Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. Method Mol Biol 132:365–386

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tajima F (1993) Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135:599–607

Wang JP, Na JK, Yu QY, Gschwend AR, Han J, Zeng FC, Aryal R, VanBuren R, Murray JE, Zhang WL, Navajas-Perez R, Feltus FA, Lemke C, Tong EJ, Chen CX, Wai CM, Singh R, Wang ML, Min XJ, Alam M, Charlesworth D, Moore PH, Jiang JM, Paterson AH, Ming R (2012) Sequencing papaya X and Y^h chromosomes reveals molecular basis of incipient sex chromosome evolution. Proc Natl Acad Sci USA 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

Wright SI, Charlesworth B (2004) The HKA test revisited: a maximum-likelihood-ratio test of the standard neutral model. Genetics 168:1071–1076

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci 13:555–556

Yang ZH (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yu QY, Hou SB, Hobza R, Feltus FA, Wang X, Jin WW, Skelton RL, Blas A, Lemke C, Saw JH, Moore PH, Alam M, Jiang JM, Paterson AH, Vyskot B, Ming R (2007) Chromosomal location and gene paucity of the male specific region on papaya Y chromosome. Mol Genet Genomics 278:177–185

Yu QY, Hou S, Feltus FA, Jones MR, Murray JE, Veatch O, Lemke C, Saw JH, Moore RC, Thimmapuram J, Liu L, Moore PH, Alam M, Jiang JM, Paterson AH, Ming R (2008a) Low X/Y divergence in four pairs of papaya sex-linked genes. Plant J 53:124–132

Yu QY, Navajas-Perez R, Tong E, Robertson J, Moore PH, Paterson AH, Ming R (2008b) Recent origin of dioecious and gynodioecious chromosomes in papaya. Tropical Plant Biol 1:49–57

Yu QY, Tong E, Skelton RL, Bowers JE, Jones MR, Murray JE, Hou SB, Guan PZ, Acob RA, Luo MC, Moore PH, Alam M, Paterson AH, Ming R (2009) A physical map of the papaya genome with integrated genetic map and genome sequence. BMC Genom 10:371–382

Zeng K, Shi S, Wut CI (2007) Compound tests for the detection of hitchhiking under positive selection. Mol Biol Evol 24:1898–1908

Zhang WL, Wang XU, Yu QY, Ming R, Jiang JM (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 (2012a) Chromosome-wide gene silencing initiates Y degeneration in *Drosophila*. Curr Biol 22:522–525

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