Mol Genet Genomics (2015) 290:1511–1522
DOI 10.1007/s00438-015-1000-3

ORIGINAL PAPER

A polymorphic pseudoautosomal boundary in the *Carica papaya* sex chromosomes

Fiona M. Lappin · Charles M. Medert · Kevin K. Hawkins · Sandra Mardonovich · Meng Wu · Richard C. Moore

Received: 6 June 2014 / Accepted: 23 January 2015 / Published online: 25 February 2015

Abstract Sex chromosomes are defined by a non-recombining sex-determining region (SDR) flanked by one or two pseudoautosomal regions (PARs). The genetic composition and evolutionary dynamics of the PAR is also influenced by its linkage to the differentiated non-recombining SDR; however, understanding the effects of this linkage requires a precise definition of the PAR boundary. Here, we took a molecular population genetic approach to further refine the location of the PAR boundary of the evolutionary young sex chromosomes of the tropical plant, *Carica papaya*. We were able to map the position of the papaya PAR boundary A to a 100-kb region between two genetic loci approximately 2 Mb upstream of the previously genetically identified PAR boundary. Furthermore, this boundary is polymorphic within natural populations of papaya, with an approximately 100–130 kb expansion of the non-recombining SDR found in 16 % of individuals surveyed. The expansion of the PAR boundary in one Y haplotype includes at least one additional gene. Homologs of this gene are involved in male gametophyte and pollen development in other plant species.

Communicated by C. Gebhardt.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-015-1000-3) contains supplementary material, which is available to authorized users.

F. M. Lappin · C. M. Medert · K. K. Hawkins · S. Mardonovich · M. Wu · R. C. Moore (✉)
Biology Department, Miami University, Oxford, OH 45056, USA
e-mail: moorerc@miamioh.edu

Present Address:
F. M. Lappin
School of Biological Sciences, University of Aberdeen, Aberdeen, UK

Keywords Plant sex chromosomes · Molecular evolution · Pseudoautosomal region · Sexual antagonism · Y polymorphism

Abbreviations

SDR	Sex-determining region
PAR	Pseudoautosomal region
QTL	Quantitative trait loci
LD	Linkage disequilibrium
MP	Maximum parsimony

Introduction

Sex chromosomes may evolve from autosomes following the suppression of recombination in a region containing sex-determining loci (Charlesworth et al. 2005; Bergero and Charlesworth 2009). This suppression of recombination divides the emergent sex chromosome into two regions: the non-recombining sex-determining region (SDR) and the recombining pseudoautosomal region (PAR). The general fate of PARs is theorized to be one of attrition as the non-recombining SDR subsumes the PAR (Graves et al. 1998; Otto et al. 2011). However, the size and genetic composition of PARs can vary greatly among independently evolved sex chromosome systems as well as among lineages sharing a common origin of sex chromosomes (Otto et al. 2011). The PAR can range in size from the relatively small PAR of eutherian mammals, which ranges from ~700 kb in some mice to ~3 Mbp in primates, to the large PARs of independently evolved young sex chromosomes in plants such as papaya (~45 Mbp) and asparagus (Graves et al. 1998; Telgmann-Rauber et al. 2007; Yu et al. 2009; White et al. 2012). And although older sex chromosomes tend to have smaller PARs than younger sex

chromosome systems, this trend is not universal. For example, the PAR of the ~120 My old sex chromosomes of ratite birds comprise over 80 % of the sex chromosomes (Pigozzi and Solari 1999; Vicoso et al. 2013), while the PARs of the comparatively young (~7 My) papaya sex chromosomes comprise 87 % of the length of the sex chromosomes (Liu et al. 2004; Yu et al. 2009; Wang et al. 2012).

Although sex chromosome research has largely focused on the origins and evolutionary dynamics of the SDR, the PAR should not be overlooked in terms of its evolutionary and functional significance. The PAR is necessary for proper chromosomal pairing of the sex chromosomes in most species, with notable exceptions in the achiasmatic sex chromosomes of marsupials and insect groups including Dipterans and Lepidopterans which lack recombination in the heterogametic sex (reviewed in Otto et al. 2011). Because the recombination between the sex chromosomes occurs only in the PAR, the PAR is typically characterized by elevated rates of recombination relative to other autosomes (Otto et al. 2011). This increase in recombination rate has been thought to have led to the elevated levels of polymorphism in mammalian PAR genes due to the mutagenic effects of recombination (Filatov 2004; Bussell et al. 2006). This in turn leads to an accelerated evolutionary rate of PAR-linked genes (Galtier 2004).

The evolutionary dynamics of the PAR is also influenced by its linkage to the differentiated non-recombining SDR. For example, PAR attrition is in part thought to be driven by selection for SDR expansion via the acquisition of sexually antagonistic loci, or those loci that increase fitness preferentially in one sex (Rice 1984; Jordan and Charlesworth 2012). Expansion of the non-recombining SDR appears to evolve in a step-wise matter, as evidenced by the footprint of evolutionary strata in diverse species such as humans and the dioecious plant *Silene latifolia* (Lahn and Page 1999; Nicolas et al. 2005; Skaletsky et al. 2003; Bergero et al. 2007). These strata typically form through genomic rearrangements such as chromosomal inversions that incorporate large chromosomal regions into the non-recombining SDR (Lahn and Page 1999; Bergero et al. 2013). However, the distinctiveness of these strata is not completely clear (Chibalina and Filatov 2011; Sayres and Makova 2013). Furthermore, direct evidence that it is the acquisition of sexually antagonistic loci that drives this expansion is lacking (Bergero and Charlesworth 2009; Otto et al. 2011).

Genes regulating sexually selected traits may also be maintained on the PAR through the maintenance of sex-linked polymorphism (Otto et al. 2011). For example, QTL controlling sexually dimorphic traits in *S. latifolia* disproportionally map to the PAR (Delph et al. 2010). Consistent with this observation, PAR loci in this species exhibit the molecular population genetic signature of balanced polymorphism possibly due to the maintenance of sex-specific alleles (Qiu et al. 2013). This signature of elevated polymorphism is predicted to be strongest for PAR genes closely linked to the SDR (Otto et al. 2011; Jordan and Charlesworth 2012; Charlesworth et al. 2014; Kirkpatrick and Guerrero 2014), although the signature of balanced polymorphism in the *S. latifolia* PAR extends to loci as distant as 21 cM from the SDR (Qiu et al. 2013), and sex-biased gene expression in emus appears to have evolved throughout the PAR, which represents most of the emu ZW pair (Viscoso et al. 2013). Patterns of sex-biased gene expression, however, were not found for 80 PAR genes in the sexually dimorphic three-spine stickleback (Leder et al. 2010).

Understanding the linkage effects of the SDR on PAR genes requires distinguishing between the PAR and the non-recombining SDR by identifying the PAR boundary. In many mammals the PAR boundaries are precisely mapped by identifying abrupt transitions of X/Y sequence similarity to distinct patterns of X/Y divergence within genes that straddle the PAR boundary (Ellis et al. 1989; Palmer et al. 1997; Van Laere et al. 2008; Skinner et al. 2013). For example, the PAR boundary in apes and old world monkeys interrupts the gene coding the XG blood group antigen, while the PAR boundary of *Mus musculus domesticus* is found within intron 3 of the developmental patterning gene, *Midline1* (*Mid1*; Ellis et al. 1989, 1994; Palmer et al. 1997; White et al. 2012). In contrast, in the few plants with genetic markers mapped on sex chromosomes, the genetic and physical locations of the boundaries are not precisely known (Telgmann-Rauber et al. 2007; Yin et al. 2008; Blavet et al. 2012; Bergero et al. 2013).

Papaya has both genetic mapping and physical sequence data allowing for investigations into the nature of the PAR boundaries (Ming et al. 2011, 2012; Wang et al. 2012). Papaya has three genders, including males, females, and hermaphrodites. Gender is determined by an active Y sex determination system; males are XY and females are XX. Hermaphrodites are heterogametic like males and have a hermaphrodite-specific Y chromosome (Y^h). The male Y and hermaphrodite Y^h are over 99 % identical at the nucleotide level suggesting they share a recent common ancestor; however, the genotypic differences giving rise to the differences in sex expression have not been identified (Yu et al. 2008). Cultivars are predominantly gynodioecious (females and hermaphrodites), while natural populations are dioecious (males and females; Manshardt and Zee 1994; Brown et al. 2012).

The relatively small papaya non-recombining SDR spans a centromere and contains a number of heterochromatic chromosomal knobs. Four of these knobs are Y^h-specific, with the centromere located at or near knob 4 (Zhang et al. 2008). While genetic mapping of PAR boundaries places

them in the regions that flank the five chromosomal knobs, knob 1 is shared between the X and Y^h (Zhang et al. 2008), and sequencing of this region suggests the boundary nearest to the genetically determined PAR boundary A, is more centromere proximal, between knobs 1 and 2 (Zhang et al. 2008; Na et al. 2012). Sequence divergence analyses between the non-recombining region of the Y and the X homologous region typically do not include the proximal ~2 Mbp region from PAR Border A that contains knob 1, because there is a large gap in the Y^h physical map in this region due to a lack of identifiable Y^h-specific sequence in this region (Na et al. 2012; Wang et al. 2012). As the specific location of the PAR boundary in this region is unknown, we took a molecular population genetic approach to refine its position between the genetically mapped Border A and the first currently annotated fully sex-linked locus, *CpXY^{h}1*.

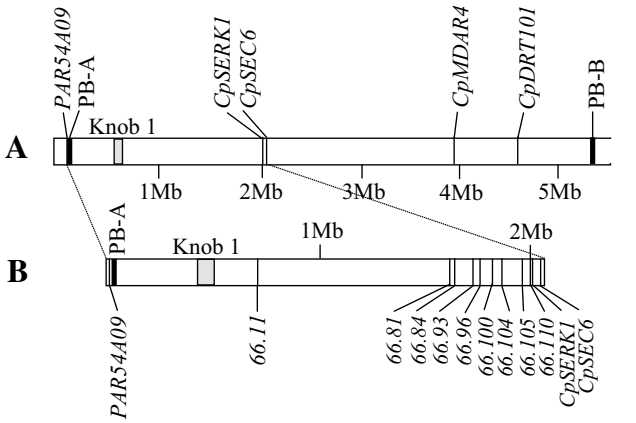

Fig. 1 Diagram of the positions of **a** the genetically determined PAR boundaries (PB-A and PB-B) and loci from a previous population genetic analysis of sex-linked loci (*CpSERK1*, *CpSEC6*, *CpMDAR4*, *CpDRT101*; Weingartner and Moore 2012) on the sequenced X-pseudomolecule (Wang et al. 2012), and **b** the focal loci from this study, focusing on 2Mbp region between the genetically determined PB-A and the first annotated sex-linked locus, *66.110*. The gene poor heterochromatic knob 1, which is shared between the papaya X and Y, is positioned relative to its placement in Na et al. (2012)

Materials and methods

Resequencing of the PAR boundary and adjacent sex-linked loci

Genomic DNA was isolated from 43 predominantly male papaya individuals from five regional populations in Costa Rica (Table S1; Brown et al. 2012) and one cultivar, SunUp. Only individuals which share <15 % genetic identity with cultivated papaya were used as determined by a Bayesian model-based clustering in the program STRUCTURE (Pritchard et al. 2000; Brown et al. 2012). The cultivar sample, SunUp, used in population genetic analyses was provided by Dr. Qingyi Yu (Texas A&M, AgriLife Research Center at Weslaco).

The region of interest is overlapped by the 1.3-Mbp supercontig 66 of the assembled papaya genome sequence, with the annotated gene *66.110* corresponding to X-linked allele of *CpXY^{h}1* (Fig. 1). Target loci were chosen from the annotated genes on supercontig 66, a region that spans knob 1 (Table 1, Table S2). We sequenced randomly selected loci progressively closer to *66.110*, assessing for patterns of polymorphism consistent with autosomal or sex-linkage (Fig S1). Primers were designed to reside in exons and amplify ~1–1.2 kb of mostly intron sequence using Primer3 (Rozen and Skaletsky 2000). For the previously identified sex-linked loci, such as *66.110*, X- or Y-specific primers were designed to distinguish between the alleles. For all other loci, primers were designed based on the papaya genome coding annotations. Genomic positions of the loci are relative to the published X-pseudomolecule sequence (Wang et al. 2012).

Loci were resequenced in either 43 individuals (86 alleles) or a core set of 7 individuals (14 alleles). The core set of 7 individuals are representatives from each of the regional populations from Costa Rica sampled in Brown et al. (2012) plus the cultivar SunUp (Table S1). Resequencing was initially performed for all 43 individuals for four loci: *PAR5A09*, *66.11*, *66.81*, and *66.110*. For these loci, heterozygosity and diversity from the core individuals are similar to those from all 43 individuals (Figs S2, S3); therefore, for subsequent loci we sequenced only from the seven core individuals. We subsequently resequenced locus *66.104* in the larger panel of individuals following the discovery of evidence for complete sex-linkage in two of the core individuals.

PCR was performed using GoTaq colorless master mix (Promega, Madison, WI). PCR conditions were 94 °C for 2 m followed by 30 cycles of 94 °C for 30 s, 56 °C for 30 s, and 72 °C for 30 s, followed by 72 °C for 5 m. The PCR product was cleaned using a combination of 5 U exonuclease I and 0.5 U shrimp alkaline phosphatase for 40 m at 37 °C followed by 15 m at 80 °C. Purified PCR products were then directly sequenced using BigDye Terminator v3.1 (Applied Biosystems, Foster City, CA) on a 3130*xl* or 3730 Genetic Analyzer (Applied Biosystems). Sequences were assembled and aligned using BioLign and BioEdit software (Hall 1999). Polymorphic sites, including heterozygous sites in individuals, were visually confirmed in Biolign and heterozygous sites denoted using the IUPAC nucleotide code for ambiguous bases. Indel polymorphisms in heterozygous individuals were sequenced from both ends to obtain the sequence flanking the indel. Final sequence alignments may be obtained from GenBank (Accession # KP780477–KP780804).

Table 1 Summary of nucleotide diversity and Tajima's *D* values for focal loci sampled from core papaya individuals

Locus	Position on X pseudo-molecule[a]	No. alleles	Total sites	S[b]	π[c]	θ_W[d]	Tajima's *D*
PAR54A09	1870	14	978	1	0.0002	0.0003	−1.15
66.11	706471	14	1,051	3	0.0010	0.0010	0.469
66.84	1626588	14	887	6	0.0028	0.0028	1.11
66.84	1641853	14	1,205	6	0.0019	0.0016	0.864
66.93	1734828	14	1,089	10	0.0030	0.0030	−0.272
66.96	1763207	14	679	12	0.0062	0.0056	0.482
66.100	1828698.5	14	478	4	0.0037	0.0026	1.31
66.104	1862489.5	14	1,351	12	0.0028	0.0028	−0.004
66.105	1963398	14	547	22	0.0197	0.0127	2.33***
66.110	2002867	14	1,033	22	0.0114	0.0068	2.83***
66.111 (CpSERK1 X/Y[e]*)*	2009452	14	1,315	45	0.0178	0.0108	2.85***
66.112 (CpSEC6X/Y[e]*)*	2052096.5	14	4,274	211	0.0253	0.0157	2.74***

*** Proportion (*P*) of coalescent simulations with larger *D* than the observed value is <0.001

[a] Midpostion of sequenced region on the X-pseudomolecule

[b] S = No. segregating sites

[c] π = Nucleotide diversity

[d] θ_W = Watterson's estimator of diversity

[e] Previously sequenced sex-linked loci (Weingartner and Moore 2012)

Analysis of patterns of heterozygosity

Patterns of heterozygosity were analyzed on unphased alleles. Sex-linked loci are expected to show patterns of heterozygosity at most or all variant sites (i.e., all X/Y individuals are heterozygous), whereas PAR-linked loci will have a mixture of homozygous and heterozygous sites (Fig. S1). Heterozygosity patterns were determined for all loci for the core set of seven male/hermaphrodite individuals. Heterozygosity patterns were analyzed in an additional 36 individuals for Loci *66.11*, *66.81*, *66.104*, and *66.110* and are consistent with the patterns found in the core 7 individuals (Fig. S2). Phased alleles (see below) from these individuals were grouped by Y haplotype (see below) and an exact test of Hardy–Weinberg expectations (Guo and Thompson 1992) was performed using Genepop (Raymond and Rousset 1995; Rousset 2008). A significant result rejects the null hypothesis of random union of gametes. F_{IS}, a measure of deviation from Hardy–Weinberg proportions within subpopulations, was also estimated in Genepop using the methods of Weir and Cokerham (1984) and Robertston and Hill (1984). A negative F_{IS} indicates an excess of heterozygotes.

Molecular population genetic analyses

Alleles for each individual were phased using the program PHASE as implemented in DnaSP v5.10 with output probability threshold for genotypes and haplotypes set to 0.9 (Librado and Rozas 2009). Nucleotide diversity was estimated for all loci using DnaSP v5.10 (Librado and Rozas 2009). We calculated two estimates of the population mutation parameter theta ($\theta = 4N_e\mu$): π (Tajima 1989), and Watterson's estimator, θ_w, which is based on the number of segregating sites (Watterson 1975). Tajima's *D*, a site frequency spectrum test statistic, was calculated based on the alignment of all alleles at each site; thus, for known sex-linked loci, both X and Y alleles were combined in the analysis. Tajima's *D* values based on the core set of seven individuals were comparable to Tajima's *D* values based on larger sample sizes (Fig. S3). In one exception, Tajima's *D* values for the sex-linked locus *66.110*, and the previously sequenced loci *CpSERK1X/Y* (*66.111*) and *CpSEC6X/Y* (*66.112*), were lower when calculated using all 43 individuals than with just the core set of seven individuals; however, *D* was still significantly positive for these loci (Table S3; Fig. S3). Alignments of phased alleles for individual loci were concatenated and LD across the loci, r^2, was calculated between parsimoniously informative sites with the significance determined by Fisher's exact test as implemented in DnaSP v5.10 (Librado and Rozas 2009).

Maximum Parsimony trees were constructed using the PARS program of the Phylip software package as implemented on the Trex online analysis server (http://www.trex.uqam.ca). For each locus we generated a single most parsimonious unrooted tree for the 14 alleles sampled from our 7 individuals. Trees were viewed using Dendroscope (Huson and Scornavacca 2012) and rooted to separate the two most divergent allele classes. Haplotype networks were assembled using statistical parsimony as implemented in the TCS program (Clement et al. 2000).

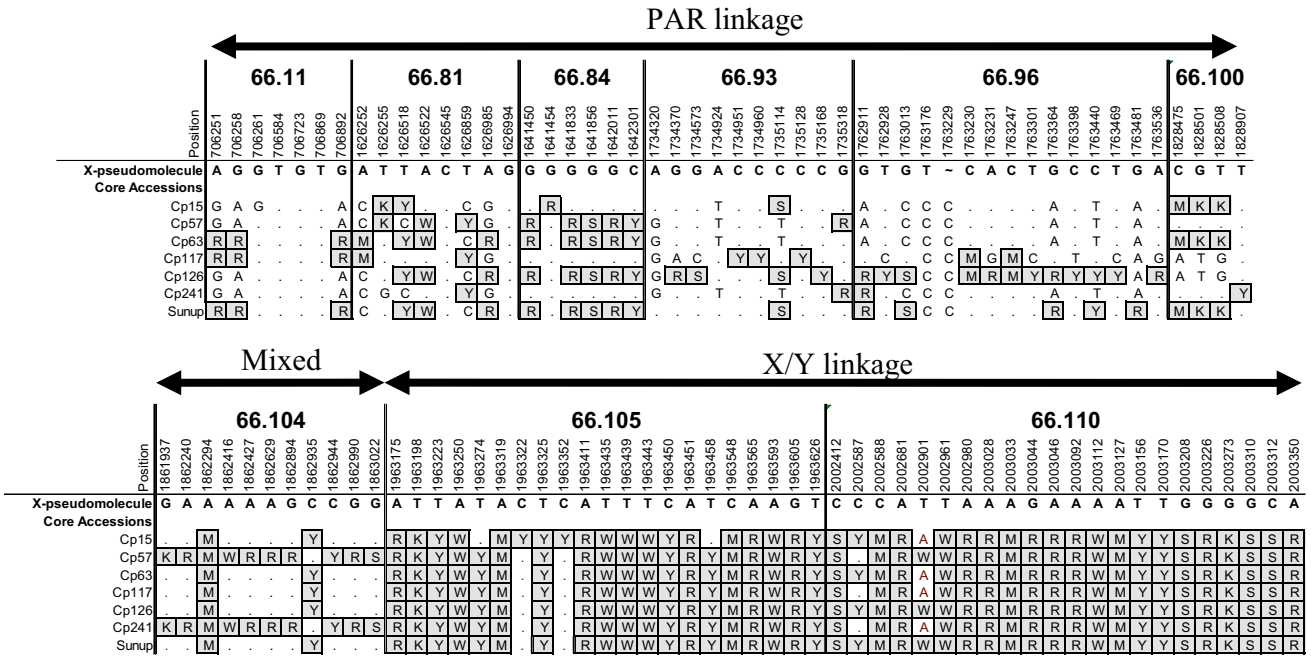

Fig. 2 Patterns of heterozygosity at focal loci in the seven core papaya male individuals. Polymorphic sites of unphased alleles are aligned to the standard X-pseudomolecule sequence (Wang et al. 2012). Gene identity and the positions of each polymorphic site on the X-pseudomolecule are indicated above each position. Homozygous sites are represented by either their nucleotide or a *Dot* indicating sequence identity with the standard. Heterozygous sites are boxed and shaded. Standard IUPAC codes for nucleotides are used to indicate the nucleotide or nucleotides at each position. Loci where all individuals are heterozygous at most or all variable sites are labeled as having "X/Y linkage," while loci without this signature are identified as having "PAR linkage." Two individuals, Cp241 and Cp57, exhibit heterozygosity patterns consistent with sex-linkage at locus *66.104*; thus, this locus is denoted as having "Mixed" linkage

Results

We surveyed the levels and patterns of sequence polymorphism at gene loci found within the ~2Mbp region between the genetically mapped Border A and the currently annotated X/Y locus, *CpXY*^{h}*1,* in a core panel of five male and two hermaphrodite individuals (see "Material and methods") and assessed each locus for the evolutionary genetic signature of sex-linkage (see Qiu et al. 2010; Lawson-Handley 2004).

Fully sex-linked loci should be heterozygous at most variable sites in XY individuals, unlike autosomal or PAR loci (Fig. S1). In loci *66.105* and *66.110*, all 7 individuals were heterozygous at almost all variable sites, consistent with complete sex-linkage (Fig. 2). None of the other loci studied showed this pattern. However, two individuals, Cp57 and Cp241, are heterozygous for almost all variant sites at locus *66.104*. Further resequencing of *66.104* in 43 individuals revealed seven individuals with this pattern (16.3 %; Fig. S2).

Fully sex-linked loci will also exhibit distinct polymorphism signatures if sequences from males and hermaphrodites are analyzed as a single population. Because fully sex-linked alleles will be more diverged than autosomal alleles, sex-linked loci will exhibit increased nucleotide diversity than autosomal or PAR loci, due to the presence of two diverged alleles per XY individual. They should also have positive Tajima's *D* values due to an excess of intermediate frequency X/Y alleles. In our analyses, we also included previously sequenced sex-linked loci adjacent to *66.110*: *CpSERK1X/Y* (*66.111*) and *CpSEC6X/Y* (*66.112*) and an additional locus, *PAR54A09* (supercontig locus *177.19*), located at the genetically identified Border A (Weingartner and Moore 2012).

Diversity estimates increase sharply between loci *66.104* and *66.105*; π increases 7-fold and θ_W almost 4-fold (Table 1; Fig. 3). Tajima's *D* for the PAR loci becomes increasingly positive near the putative SDR border and is significantly positive at *66.105* and centromere-proximal loci, consistent with complete sex-linkage. Only one putatively PAR locus, *66.96*, has a θ_W value slightly greater than the fully sex-linked locus *66.110* ($\theta_W = 0.00556$ for *66.96* vs. $\theta_W = 0.00426$ for *66.110*), though its π value is lower than that of the fully sex-linked loci and its Tajima's *D* is not significantly different from zero.

A third indicator of complete sex-linkage is linkage disequilibrium (LD) among polymorphic sites within loci, and also between different loci, due to the suppression of recombination between the X- and Y-linked chromosomal regions. Indeed, LD is highly significant between *66.105* and *66.110* and breaks down abruptly between *66.104*

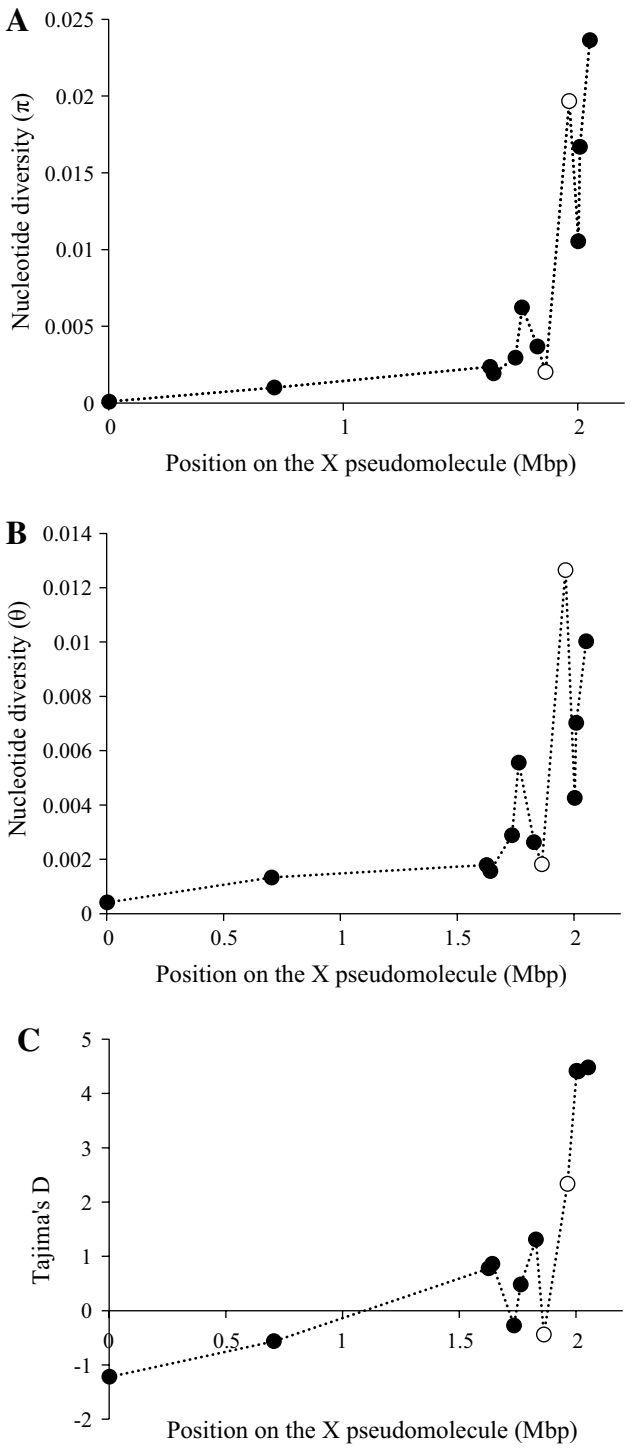

Fig. 3 Graphs of nucleotide diversity estimates **a** π, **b** θ_W, and **c** Tajima's *D* values for focal loci sampled from seven core papaya male individuals versus their position on the X-pseudomolecule. *Open circles* indicate Loci *66.104* and *66.105*. Diversity estimates (**a**, **b**) and Tajima's *D* (**c**) rise sharply between *66.104* and *66.105*, suggesting that these two loci flank the PAR boundary

and *66.105* (Fig. 4). Upstream of *66.105*, significant LD is mainly confined to sites within loci, except for a single site in *66.104* that is in significant LD with sites in *66.105* and *66.110* and significant LD between *66.96* and *66.93* (Fig. 4).

Last, phylogenies of fully sex-linked alleles in a region of suppressed recombination will yield a tree with two distinct clades corresponding to the X- and Y-linked alleles, unlike trees for loci in recombining regions (the autosomes and PAR). Both *66.105* and *66.110* phylogenies behave as expected for complete sex-linkage. The MP trees of most other loci typically produced clades with many fewer mutational steps and clades with alleles from the same individuals, consistent with pseudo-autosomal linkage (Fig. 5). There were two exceptions. First, gene *66.96* has two divergent clades separated by nine mutational steps; this is consistent with the elevated θ_W reported above. However, the alleles of the male Cp117 are grouped together in the same divergent clade, inconsistent with complete sex-linkage. Gene *66.104* also has two divergent clades separated by 9 mutational steps. Unlike *66.96*, however, one of these clades contains a single divergent allele found in only two males, Cp57 and Cp241 (Fig. 5), suggesting complete sex-linkage of this gene in these individuals.

The polymorphism we observe at *66.104* is associated with the male Y, and the region remains partially sex-linked in all the Y^h individuals and in some Y individuals. Previous analyses identified three major Y haplotypes in natural papaya populations (A, B, and C in Weingartner and Moore 2012). Individuals sampled from each of three Y chromosome haplotypes have a significant excess of heterozygotes at *66.104* ($P < 0.05$, exact test of Hardy–Weinberg) and negative F_{IS} estimates (Table S4); however, only male individuals that carry the A haplotype Y chromosome possess divergent alleles consistent with complete sex-linkage. Haplotype network analysis of *66.104* identified a single highly divergent allele unique to Y haplotype A, similar to the pattern of divergent sex-linked alleles found at the completely sex-linked locus *66.110* (Fig. 6). The silent site divergence (K_{sil}) between the divergent, potentially sex-linked, alleles at *66.104* in the A haplotype Y chromosome ($K_{sil} = 0.0135$) is similar to the average K_{sil} for the more recent stratum 2 (average $K_{sil} = 0.0152 \pm 0.003$) and collinear (PAR) regions (average $K_{sil} = 0.008 \pm 0.003$; Wang et al. 2012). In contrast, mean divergence between alleles at *66.104* in the B ($K_{sil} = 0.0034$) and C haplotype groups ($K_{sil} = 0.0000$) is similar to the mean allelic divergence at

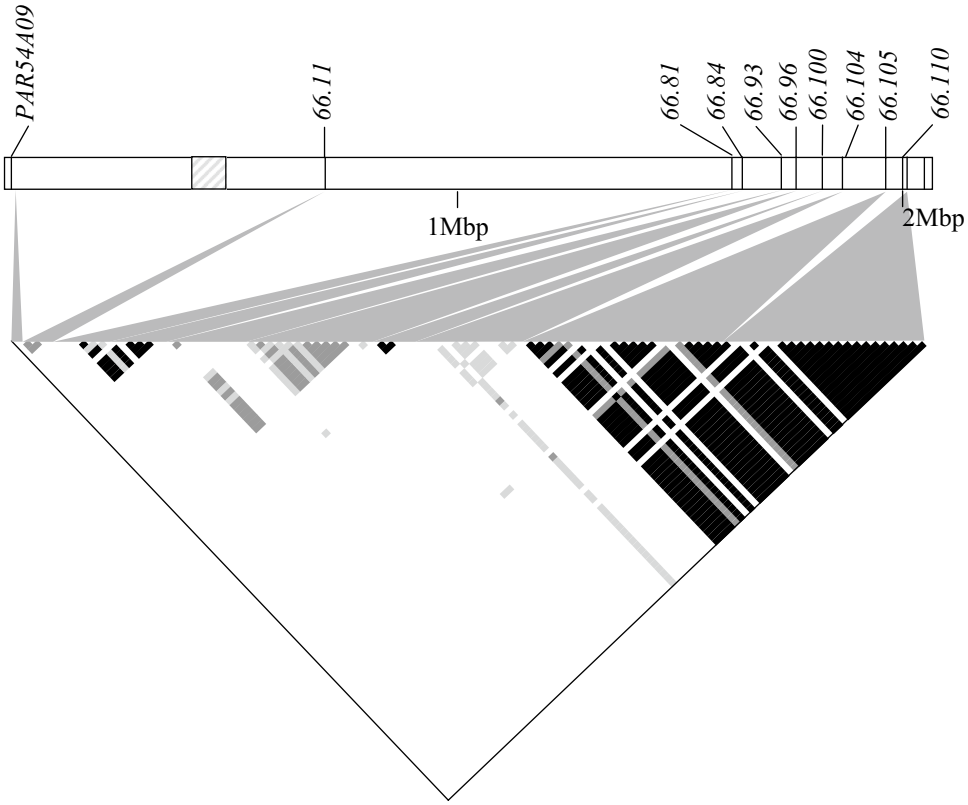

Fig. 4 Linkage disequilibrium plot of parsimoniously informative sites for the focal loci. Relative positions of polymorphic sites within each locus and the approximate distance between loci are indicated above the plot. Significant LD among sites is indicated by *shading* ($P < 0.001$, *black*; $P < 0.01$, *medium gray*; $P < 0.05$, *light gray*). Loci *66.105* and *66.110* exhibit strong LD consistent with sex-linkage; one site in *66.104* is in significant ($P < 0.05$) LD with *66.105* and *66.110*

pseudo-autosomal loci *66.11* ($K_{sil} = 0.0017–0.0034$) and *66.81* ($K_{sil} = 0.0046–0.0057$; Table S5).

Discussion

Our data support previous inferences that the region surrounding the heterochromatic knob 1 is not fully sex-linked, but rather it is a component of the PAR (Na et al. 2012; Wang et al. 2012). Specifically, our analyses place the PAR boundary in a 100-kb region between loci *66.104* and *66.105* in the majority of individuals (Fig. 7). However, in 7 of 43 sampled individuals, *66.104* appears complete sex-linked. Also, *66.104* is in moderate LD with variants in the fully sex-linked loci studied and the phylogeny and haplotype network for *66.104* suggests a mixture of PAR alleles and fully sex-linked alleles. The individuals that exhibit the signature of complete sex-linkage at *66.104* also contain just one of the three Y chromosomes (haplotype A) that have been identified in natural papaya populations (Weingartner and Moore 2012). The A haplotype Y chromosome is found in moderate frequency (16–37 %) in natural Costa Rican populations in the Caribbean, central Pacific coast, and southern Pacific coast, but is absent from northwest Guanacaste and Nicoya Peninsula (Weingartner and Moore 2012). The signature of partial-sex linkage at *66.104* in males and hermaphrodites with haplotype B or C chromosomes (e.g., excess of heterozygotes) is expected in a PAR region near the non-recombining SDR (Kirkpatrick et al. 2010). These results suggest that the PAR boundary for Y haplotype A chromosome resides in the 33-kb region between *66.100* and *66.104* (Fig. 7).

Interspecific variation in PAR boundaries has been documented in other sex chromosome systems, though intraspecific polymorphism has not often been documented (or often tested for). PAR boundary locations vary among eutherian mammalian sex chromosomes, though they tend to be conserved in closely related species (Ellis et al. 1990; Mangs and Morris 2007; Van Laere et al. 2008; Skinner et al. 2013; White et al. 2012). For example, the PAR boundary in primates arose before the divergence of the Old World monkeys and great ape lineages (25–45 My), though subsequent incorporation of an *Alu* element in the great ape lineage distinguished them from the rest of primates (Ellis et al. 1989, 1990). However, the PAR boundary in artiodactyls has been more or less conserved for the last ~50 My, though there are lineage-specific expansions of the PAR, such as in swine (Van Laere et al. 2008; Skinner et al. 2013).

In mice, which have the smallest characterized PAR of eutherian mammals (~700 kb; Perry et al. 2001), the PAR boundary varies between species and subspecies (White et al. 2012). For example, the PAR boundary is found in the *Mid1* gene in the domesticated mouse *Mus musculus*

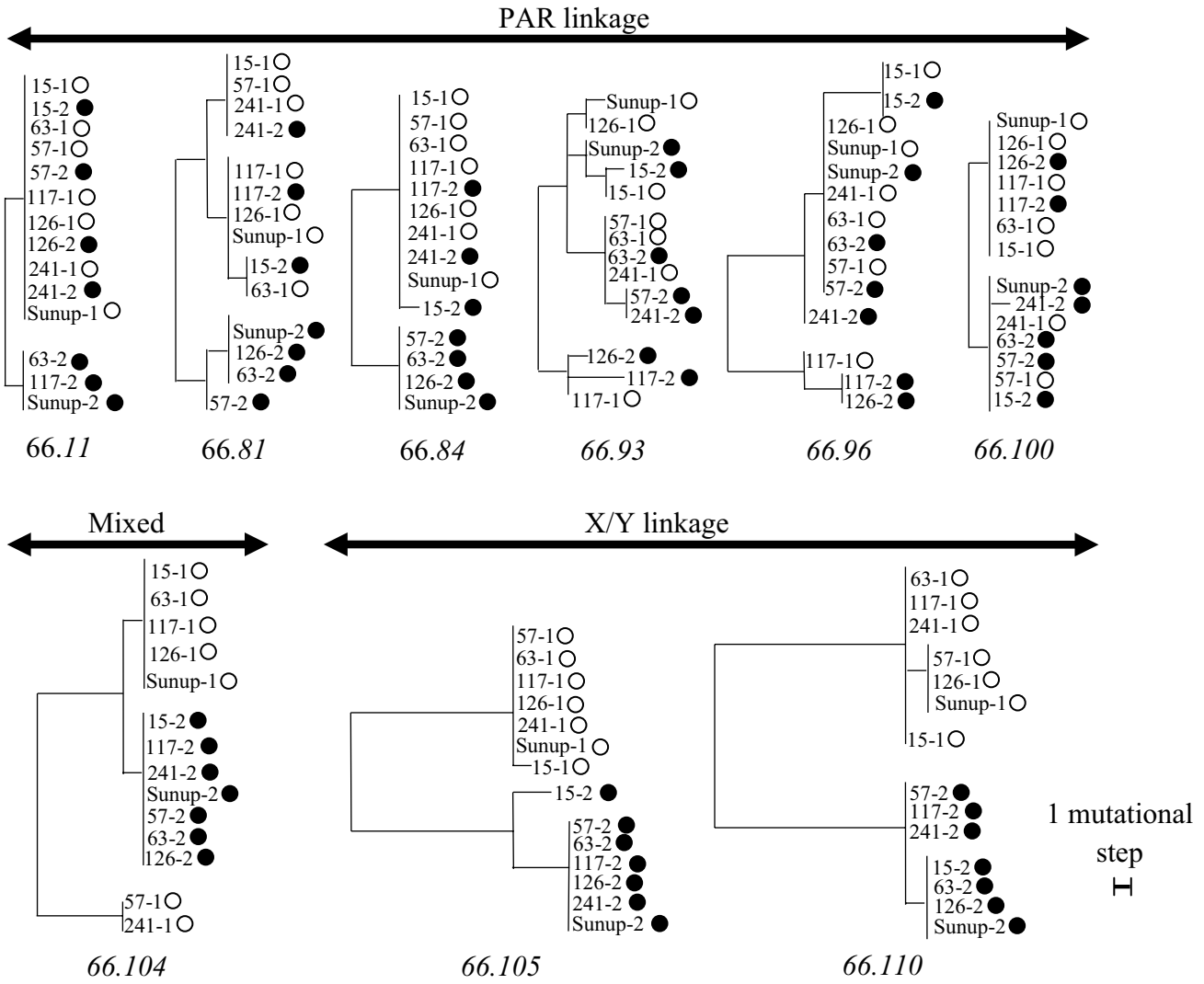

Fig. 5 Maximum Parsimony trees illustrating the evolutionary relationships between phased alleles at each focal locus in the seven core male and hermaphrodite individuals. The alleles from each individual are numbered (−1 and −2) and represented by *empty* and *filled circles*. The *scale bar* indicates a single mutational step. Trees are rooted to differentiate among the most divergent alleles. PAR-linked loci are characterized by shallow trees and clades comprised of both allele classes due to an increased homozygosity at these loci (PAR linkage). Sex-linked loci are characterized by deeply rooted trees with two classes of highly divergent alleles (X/Y linkage). Locus *66.104* is unique in that the alleles for two individuals, Cp57 and Cp241, are highly divergent, consistent with sex-linkage (mixed)

domesticus and has shifted proximally from its location in *M. spretus* within the last 2–4 my since their species divergence (Palmer et al. 1997). In contrast, the PAR boundary in a subspecies of domesticated mouse, *M. domesticus castaneous*, has shifted proximally ~430 kb and *Mid1* is found wholly within the PAR (White et al. 2012). Similar to the situation in papaya, there is intraspecific variation (on the order of hundreds of kb) in the PAR boundary of inbred strains of *M. musculus domesiticus* that is associated with repetitive DNA elements (Kipling et al. 1996a, b). In papaya, the non-recombining SDR has expanded in male haplotype A and is smaller in male haplotypes B and C. This expansion could be facilitated by the abundance of repetitive sequences in the papaya PAR and SDR through illegitimate recombination increasing the size of the SDR (Wang et al. 2012; Charlesworth 2013; VanBuren and Ming 2013).

The presence of sexually antagonistic (SA) polymorphisms in PAR loci could also affect the expansion of the non-recombining SDR (reviewed in Otto et al. 2011). The 100–130 kb expansion of the SDR on the A haplotype Y chromosome contains only a few genes (*66.101–66.104*), and only *66.104* shares significant similarity with a gene in *Arabidopsis thaliana*. The putative ortholog of gene *66.104, FORMS APLOID AND BINUCLEATE CELLS 1D* (*FAB1D* or AT1G34260; Table S2) is broadly expressed,

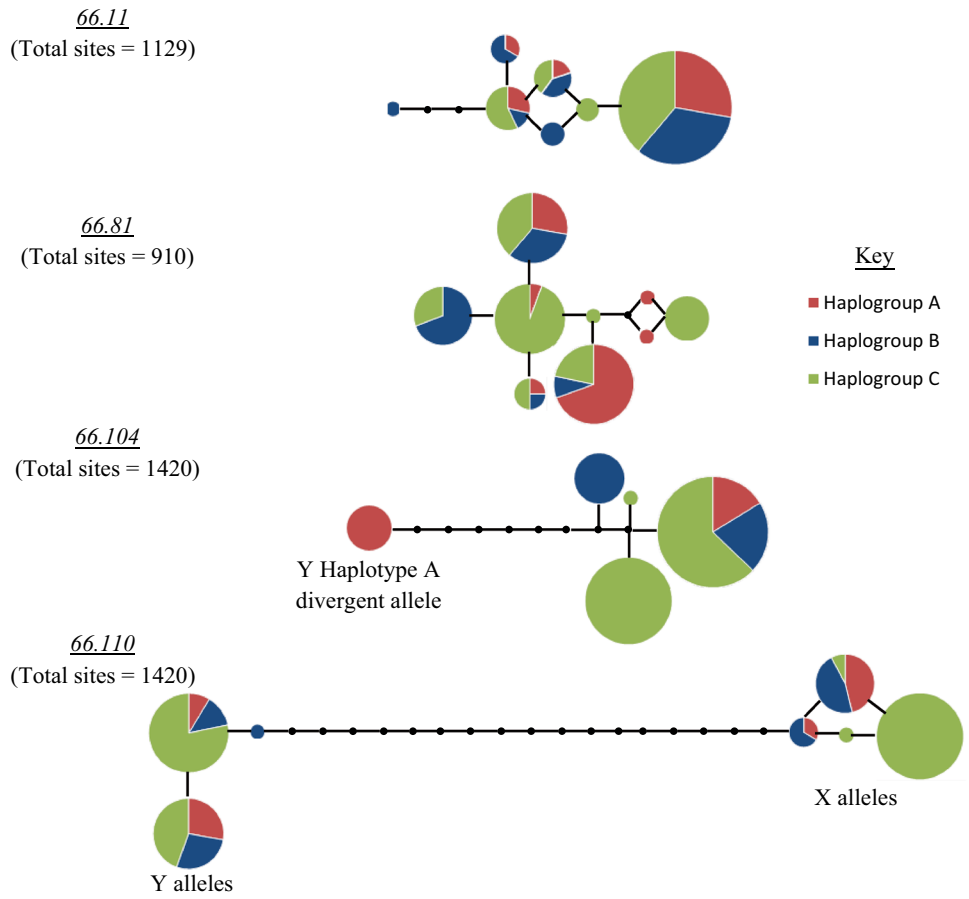

Fig. 6 Haplotype networks of loci *66.11*, *66.81*, *66.104*, and *66.110*. Haplotype *circles* are proportional to number of represented accessions. Sections within each haplotype circle are color coded to indicate Y haplotype (A = *red*, B = *blue*, C = *light green*). A mutational step is indicated by a *line* connecting haplotypes, while hypothesized haplotypes are indicated by *small*, *black circles* (Color figure online)

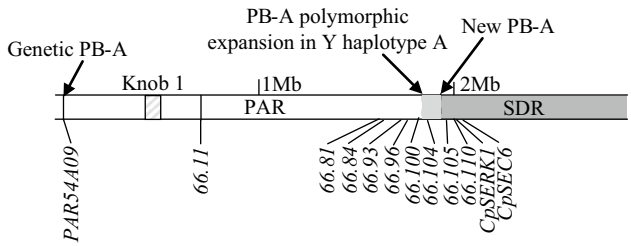

Fig. 7 Diagram of the positions of the newly defined Par Boundary A (New PB-A) between loci *66.104* and *66.105* and its polymorphic expansion between loci *66.100* and *66.104* relative to the position of the genetically determined PB-A and heterochromatic knob 1

but is also highly expressed in pollen tubes and sperm cells (Wang et al. 2008; Whitley et al. 2009; Serrazina et al. 2014). *FAB1D* is a member of a small gene family (also including *FAB1A*, *FAB1B*, *FAB1C*) of phosphatidylinositol 3,5-bisphosphate kinases that regulates endomembrane homeostasis that act redundantly and/or additively to regulate pollen development and male fertility (Hirano et al. 2011). Although single gene knockouts in *FAB1A*, *FAB1B*, and *FAB1D* are functionally normal, double knockouts of *FAB1A* and *FAB1B* are male gametophytic lethal in Arabidopsis and pollen tubes of *FAB1B* and *FAB1D* double knockouts have abnormal growth and are prone to apical bursting and growth arrest (Hirano et al. 2011; Serrazina et al. 2014). Functional analysis of the *66.104* gene in papaya is needed, however, to determine where and when it is expressed and whether it is expressed in females, which is necessary for sexually antagonistic selection to occur.

Our work corroborates previous estimates of the size of the X homologous region of the SDR as ~3.5 Mbp and extends this region by a minimum of 40 kb to include additional fully sex-linked loci (*66.105–66.109*) as well as a further 100–130 kb to include loci in the polymorphic boundary (between *66.101* and *66.104*). The X homologous region has diverged and expanded in size relative to the corresponding autosomal region in *V. monoica* (which is not dioecious, and so the region is autosomal) due to the accumulation of repetitive elements (Gschwend et al. 2012). This expansion has also occurred in the PAR region around knob 1. *V. monoica* has the lowest repetitive element content in this region and two of the five syntenic blocks showing X-specific expansion (Block 1, 32.8 % expansion and Block 2 with 57.1 % expansion) are pseudoautosomal (Gschwend et al. 2012). Accumulation of repetitive elements on the papaya X is predicted because repetitive elements are less effectively purged from

homogametic sex chromosomes (Z and X) as they spend one-third of their time in a non-recombining heterogametic state (Charlesworth et al. 2005; Bergero and Charlesworth 2009; Bellott et al. 2010). The expansion and retention of repetitive elements in the papaya PAR could be in part promoted by its pericentromeric location and reduced recombination rate relative to non-pericentromeric autosomal regions (Yu et al. 2009). However, if the papaya PAR region is pericentromeric in *V. monoica* as the evidence suggests, it is likely that the genomic expansion in this autosomal region of papaya is influenced predominantly by its proximity to the X-homologous region of the SDR.

We do not, as yet know whether there exists an abrupt PAR boundary transition in the regions we have identified, as found in many mammalian sex chromosomes, or whether there is a gradual cessation of recombination within these regions. Distinguishing between these two hypotheses will benefit from a genomic analysis of polymorphism in different Y haplotypes in the entire 133 kb region between *66.100*, which is PAR-linked, and *66.105*, which is fully sex-linked. As well, we do not know if the PAR boundary is evolutionarily conserved across the predominantly dioecious members of the Caricaceae family. It is unlikely that this border would be shared across taxa, given the hypothesized recent independent origins of recombination suppression in the SDR in this family (Yu et al. 2008). As well, the position of the SDR appears to be evolutionarily labile in other plant species, such as found for dioecious and subdioecious species of strawberry (Goldberg et al. 2010). Alternatively, it is possible that a shared, core SDR region exists across dioecious taxa within the Caricaceae, with differential expansions into the surrounding PAR in different lineages leading to the appearance of independent origins of sex chromosomes in these species. Only future comparative genomic analyses of the oldest evolutionary stratum across dioecious taxa in the Caricaceae will be able to address this question (Gschwend et al. 2013).

Acknowledgments 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 FML. We are extremely grateful to Deborah Charlesworth and one anonymous reviewer for their critical examination of the manuscript and helpful advice. 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

Bellott DW, Skaletsky H, Pyntikova T, Mardis ER, Graves T, Kremitzki C, Brown LG, Rozen S, Warren WC, Wilson RK et al (2010) Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. Nature 466:612–613

Bergero R, Charlesworth D (2009) The evolution of restricted recombination in sex chromosomes. Trends Ecol Evol 24:94–102

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, Qiu S, Forrest A, Borthwick H, Charlesworth D (2013) Expansion of the pseudo-autosomal region and ongoing recombination suppression in the *Silene latifolia* sex chromosomes. Genetics 194:673–686

Blavet N, Blavet H, Cegan R, Zemp N, Zdanska J, Janousek B, Hobza R, Widmer A (2012) Comparative analysis of a plant pseudoautosomal region (PAR) in *Silene latifolia* with the corresponding *S. vulgaris* autosome. BMC Genom 13:226–235

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

Bussell JJ, Pearson NM, Kanda R, Filatov DA, Lahn BT (2006) Human polymorphism and human-chimpanzee divergence in pseudoautosomal region correlate with local recombination rate. Gene 368:94–100

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

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

Charlesworth B, Jordan CY, Charlesworth D (2014) The evolutionary dynamics of sexually antagonistic mutations in pseudoautosomal regions of sex chromosomes. Evolution 68:1339–1350

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

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9:1657–1659

Delph LF, Arntz AM, Scotti-Saintagne C, Scotti I (2010) The genomic architecture of sexual dimorphism in the dioecious plant *Silene latifolia*. Evolution 64:2873–2886

Ellis NA, Goodfellow PJ, Pym B, Smith M, Palmer M, Frischauf AM, Goodfellow PN (1989) The pseudoautosomal boundary in man is defined by an *Alu* repeat sequence inserted on the Y chromosome. Nature 337:81–84

Ellis N, Yen P, Neiswanger K, Shapiro LJ, Goodfellow PN (1990) Evolution of the pseudoautosomal boundary in old-world monkeys and great apes. Cell 63:977–986

Ellis NA, Ye TZ, Patton S, German J, Goodfellow PN, Weller P (1994) Cloning of *Pbdx*, an *Mic2*-related gene that spans the pseudoautosomal boundary on chromosome Xp. Nat Genet 6:394–400

Filatov DA (2004) A gradient of silent substitution rate in the human pseudoautosomal region. Mol Biol Evol 21:410–417

Galtier N (2004) Recombination, GC-content and the human pseudoautosomal boundary paradox. Trends Gene 20:347–349

Goldberg MT, Spigler RB, Ashman TL (2010) Comparative genetic mapping points to different sex chromosomes in sibling species of wild strawberry (*Fragaria*). Genetics 186:1425–1433

Graves JAM, Wakefield MJ, Toder R (1998) The origin and evolution of the pseudoautosomal regions of human sex chromosomes. Hum Mol Genet 7:1991–1996

Gschwend AR, Yu QY, Tong EJ, Zeng FC, Han J, VanBuren R, Aryal R, Charlesworth D, Moore PH, Paterson AH et al (2012) Rapid divergence and expansion of the X chromosome in papaya. P Natl Acad Sci USA 109:13716–13721

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

Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. Biometrics 48:361–372

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program from Windows 95/98/NT. Nucl Acid Symp Series 41:95–98

Hirano T, Matsuzawa T, Takegawa K, Sato MH (2011) Loss-of-function and gain-of-function mutations in *FAB1A/B* impair endomembrane homeostasis, conferring pleiotropic developmental abnormalities in Arabidopsis. Plant Physiol 155:797–807

Huson DH, Scornavacca C (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst Biol 61:1061–1067

Jordan CY, Charlesworth D (2012) The potential for sexually antagonistic polymorphism in different genome regions. Evolution 66:505–516

Kipling D, Salido EC, Shapiro LJ, Cooke HJ (1996a) High frequency *de novo* alterations in the long-range genomic structure of the mouse pseudoautosomal region. Nat Genet 13:78–82

Kipling D, Wilson HE, Thomson EJ, Lee M, Perry J, Palmer S, Ashworth A, Cooke HJ (1996b) Structural variation of the pseudoautosomal region between and within inbred mouse strains. Proc Natl Acad Sci USA 93:171–175

Kirkpatrick M, Guerrero Scarpino (2010) Patterns of neutral genetic variation on recombining sex chromosomes. Genetics 184:1141–1152

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

Lahn BT, Page DC (1999) Four evolutionary strata on the human X chromosome. Science 286:964–967

Lawson-Handley L, Ceplitis H, Ellegren H (2004) Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. Genetics 167:367–376

Leder EH, Cano JM, Leinonen T, O'Hara RB, Nikinmaa M, Primmer CR, Merila J (2010) Female-biased expression on the X chromosome as a key step in sex chromosome evolution in threespine sticklebacks. Mol Biol Evol 27:1495–1503

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

Liu ZY, Moore PH, Ma H, Ackerman CM, Ragiba M, Yu QY, Pearl HM, Kim MS, Charlton JW, Stiles JI et al (2004) A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature 427:348–352

Mangs AH, Morris BJ (2007) The human pseudoautosomal region (PAR): origin, function and future. Curr Genomics 8:129–136

Manshardt RM, Zee FTP (1994) Papaya germplasm and breeding in Hawaii. Fruit Varieties J 48:146–152

Ming R, Bendahmane A, Renner SS (2011) Sex chromosomes in land plants. Annu Rev Plant Biol 62:485–514

Ming R, Yu QY, Moore PH, Paull RE, Chen NJ, Wang ML, Zhu YJ, Schuler MA, Jiang JM, Paterson AH (2012) Genome of papaya, a fast growing tropical fruit tree. Tree Genet Genome 8:445–462

Na JK, Wang J, Murray JE, Gschwend AR, Zhang W, Yu Q, Navajas-Perez R, Feltus FA, Chen C, Kubat Z et al (2012) Construction of physical maps for the sex-specific regions of papaya sex chromosomes. BMC Genom 13:176

Nicolas M, Marais G, Hykelova V, Janousek B, Laporte V, Vyskot B, Mouchiroud D, Negrutiu I, Charlesworth D, Moneger F (2005) A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. PLoS Biol 3(1):e4

Otto SP, Pannell JR, Peichel CL, Ashman TL, Charlesworth D, Chippindale AK, Delph LF, Guerrero RF, Scarpino SV, McAllister BF (2011) About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. Trends Genet 27:358–367

Palmer S, Perry J, Kipling D, Ashworth A (1997) A gene spans the pseudoautosomal boundary in mice. P Natl Acad Sci USA 94:12030–12035

Perry J, Palmer S, Gabriel A, Ashworth A (2001) A short pseudoautosomal region in laboratory mice. Genome Res 11:1826–1832

Pigozzi MI, Solari AJ (1999) The ZW pairs of two paleognath birds from two orders show transitional stages of sex chromosome differentiation. Chromosome Res 7:541–551

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Qiu S, Bergero R, Forrest A, Kaiser VB, Charlesworth D (2010) Nucleotide diversity in *Silene latifolia* autosomal and sex-linked genes. P Roy Soc B Biol Sci 277:3283–3290

Qiu S, Bergero R, Charlesworth D (2013) Testing for the footprint of sexually antagonistic polymorphisms in the pseudoautosomal region of a plant sex chromosome pair. Genetics 194:663–672

Raymond M, Rousset F (1995) Genepop (version-1.2)—population-genetics software for exact tests and ecumenicism. J Hered 86:248–249

Rice WR (1984) Sex-chromosomes and the evolution of sexual dimorphism. Evolution 38:735–742

Robertson A, Hill WG (1984) Deviations from Hardy-Weinberg proportions—sampling variances and use in estimation of inbreeding coefficients. Genetics 107:703–718

Rousset F (2008) Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Resour 8:103–106

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

Sayres MAW, Makova KD (2013) Gene survival and death on the human Y chromosome. Mol Biol Evol 30:781–787

Serrazina S, Dias FV, Malho R (2014) Characterization of FAB1 phosphatidylinositol kinases in Arabidopsis pollen tube growth and fertilization. New Phytol 203:784–793

Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T et al (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 423:825–837

Skinner BM, Lachani K, Sargent CA, Affara NA (2013) Regions of XY homology in the pig X chromosome and the boundary of the pseudoautosomal region. BMC Genet 14:1–7

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

Telgmann-Rauber A, Jamsari A, Kinney MS, Pires JC, Jung C (2007) Genetic and physical maps around the sex-determining M-locus of the dioecious plant asparagus. Mol Genet Genomics 278:221–234

Van Laere AS, Coppieters W, Georges M (2008) Characterization of the bovine pseudoautosomal boundary: documenting the evolutionary history of mammalian sex chromosomes. Genome Res 18:1884–1895

VanBuren R, Ming R (2013) Dynamic transposable element accumulation in the nascent sex chromosomes of papaya. Mob Genet Elem 3:e23462

Vicoso B, Kaiser VB, Bachtrog D (2013) Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. Proc Natl Acad Sci USA 110:6453–6458

Wang Y, Zhang WZ, Song LF, Zou JJ, Su Z, Wu WH (2008) Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. Plant Physiol 148:1201–1211

Wang JP, Na JK, Yu QY, Gschwend AR, Han J, Zeng FC, Aryal R, VanBuren R, Murray JE, Zhang WL et al (2012) Sequencing papaya X and Yh 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

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population-structure. Evolution 38:1358–1370

White MA, Ikeda A, Payseur BA (2012) A pronounced evolutionary shift of the pseudoautosomal region boundary in house mice. Mamm Genome 23:454–466

Whitley P, Hinz S, Doughty J (2009) Arabidopsis FAB1/PIKfyve proteins are essential for development of viable pollen. Plant Physiol 151:1812–1822

Yin T, DiFazio SP, Gunter LE, Zhang X, Sewell MM, Woolbright SA, Allan GJ, Kelleher CT, Douglas CJ, Wang M et al (2008) Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. Genome Res 18:422–430

Yu QY, Navajas-Perez R, Tong E, Robertson J, Moore PH, Paterson AH, Ming R (2008) 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 et al (2009) A physical map of the papaya genome with integrated genetic map and genome sequence. BMC Genom 10:371–382

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