Rapid divergence and expansion of the X chromosome in papaya

Andrea R. Gschwend[a], Qingyi Yu[b,c,1], Eric J. Tong[c], Fanchang Zeng[a], Jennifer Han[a], Robert VanBuren[a], Rishi Aryal[a], Deborah Charlesworth[d], Paul H. Moore[c], Andrew H. Paterson[e], and Ray Ming[a,1]

[a]Department of Plant Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [b]Texas AgriLife Research Center, Department of Plant Pathology and Microbiology, Texas A & M University, Weslaco, TX 78596; [c]Hawaii Agriculture Research Center, Kunia, HI 96759; [d]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, United Kingdom; and [e]Plant Genome Mapping Laboratory, University of Georgia, Athens, GA 30606

X chromosomes have long been thought to conserve the structure and gene content of the ancestral autosome from which the sex chromosomes evolved. We compared the recently evolved papaya sex chromosomes with a homologous autosome of a close relative, the monoecious *Vasconcellea monoica*, to infer changes since recombination stopped between the papaya sex chromosomes. We sequenced 12 *V. monoica* bacterial artificial chromosomes, 11 corresponding to the papaya X-specific region, and 1 to a papaya autosomal region. The combined *V. monoica* X-orthologous sequences are much shorter (1.10 Mb) than the corresponding papaya region (2.56 Mb). Given that the *V. monoica* genome is 41% larger than that of papaya, this finding suggests considerable expansion of the papaya X; expansion is supported by a higher repetitive sequence content of the X compared with the papaya autosomal sequence. The alignable regions include 27 transcript-encoding sequences, only 6 of which are functional X/*V. monoica* gene pairs. Sequence divergence from the *V. monoica* orthologs is almost identical for papaya X and Y alleles; the *Carica-Vasconcellea* split therefore occurred before the papaya sex chromosomes stopped recombining, making *V. monoica* a suitable outgroup for inferring changes in papaya sex chromosomes. The papaya X and the hermaphrodite-specific region of the Y[h] chromosome and *V. monoica* have all gained and lost genes, including a surprising amount of changes in the X.

Carica papaya | gene gains and losses | sex chromosome evolution | suppression of recombination | centromere of X chromosome

Papaya (*Carica papaya* L.) is a trioecious tropical fruit crop that has a nascent XY sex chromosome system, in which the sex determining region occupies a small fraction of the X/Y chromosome pair (1–3). Papaya has two slightly different Y chromosomes that diverged about 73,000 y ago, Y in males and Y[h] in hermaphrodites (4). All genotypes without X chromosomes (YY, YY[h], and Y[h]Y[h]) are lethal in early development, resulting in 25% aborted seeds in selfed hermaphrodites and in crosses between hermaphrodites and males (5), indicating that both Y types have lost at least one gene essential for development.

Papaya belongs to the family Caricaceae, with six genera and 35 species, 32 of which are dioecious, two are trioecious (with male, female, and hermaphrodite individuals) and one, *Vasconcellea monoica*, is monoecious (with male and female flowers on a single plant). The predominance of dioecious species suggests that dioecy is ancestral in this family and that the trioecious and monoecious species evolved recently. *V. monoica* has no sex chromosomes, because there is no sexual dimorphism among individuals, and there is a single sequence corresponding to each of the several papaya X/Y[h] gene pairs tested, whereas distinct X and Y alleles were detected in several dioecious *Vasconcellea* species (6), which indicates that the sex chromosomes in these species are homologous with those in papaya.

V. monoica therefore provides an opportunity to compare the recently evolved sex chromosomes of papaya with an orthologous autosome. Nascent sex chromosomes in several flowering plants and fish have clearly evolved from autosomes (1, 7–10). *V. monoica* and papaya are closely related (see sequence divergence results below) and both have nine chromosome pairs, suggesting that one of the autosomal pairs in *V. monoica* corresponds to the papaya sex-chromosome pair (11). However, an important issue is whether the orthologous *V. monoica* autosome is suitable as an outgroup for inferring changes in the two sex chromosomes; to be suitable, it should have had no history of being a sex chromosome in a dioecious species, and thus be likely to reflect the ancestral state. Given that the nondioecious species probably evolved from an ancestral dioecious species, we cannot a priori assume this. The cosexual state of monoecy in *V. monoica* could have arisen via recent loss of dioecy rather than retention of the ancestral state. The evolution of monoecy could have involved a reverse mutation or modification of one of the sex-determination genes in a dioecious common ancestor with papaya, or a mutation elsewhere in the genome. Analysis of X/Y gene pairs in papaya and selected *Vasconcellea* dioecious species suggested that the sex chromosomes in dioecious *Vasconcellea* species evolved after the separation of *Carica* and *Vasconcellea*, and that papaya sex chromosomes are older (2, 6). *V. monoica* and the dioecious *Vasconcellea* species therefore probably split from papaya either before, or soon after, the papaya X/Y pair stopped recombining, when the nonrecombining region was, at most, very young. We present confirming evidence for this below. Thus, it is reasonable to use *V. monoica* to infer the state of the papaya X/Y pair's ancestral autosome, and this is the approach used in our study.

Here we compare papaya X-specific sequences with orthologous *V. monoica* bacterial artificial chromosome (BAC) sequences to study changes in the papaya X-specific region since recombination was suppressed between the small sex-determining hermaphrodite-specific region of the Y[h] chromosome (HSY) and the corresponding X region. Although other studies have compared the ancient X and chicken Z chromosomes with their orthologous autosomal regions in distantly related species (12–14), our study is unique in being a detailed direct comparison between a newly evolved X-specific chromosomal region and its nonsex chromosomal homolog in a close relative. This comparison

makes it possible to study gene losses in the X and Y chromosomes of a plant after recombination was suppressed.

Results

To identify the *V. monoica* genome region homologous with the papaya X-specific region, probes were designed from the X-specific region of papaya and were hybridized to a *V. monoica* BAC library. Eleven *V. monoica* BACs and their orthologous sequences from papaya X-specific BACs were assembled and compiled into two separate pseudomolecules, ordered according to the papaya X physical map, and aligned to one another (3) (*SI Appendix*, Fig. S1 and Table S1). The *V. monoica* pseudomolecule is 1.1 Mb, whereas the corresponding papaya X pseudomolecule is 2.56 Mb, 133% larger, and an alignment revealed three syntenic regions (Fig. 1*A* and *SI Appendix*, Figs. S2 and S3*A*).

The papaya autosomal BAC (DM105M02, 73 kb) and the corresponding *V. monoica* BAC (MN15I14, 102 kb) are collinear with one shared syntenic region, and the alignable sequences are shorter in papaya (37 kb) than in *V. monoica* (99 kb) (Fig. 1*B* and *SI Appendix*, Fig. S3*B*).

The *V. monoica* genome is estimated to be 626 Mb (*SI Appendix*), about 41% larger than the 372 Mb papaya genome, which suggests a striking expansion of the papaya X-specific region, compared with the ancestral chromosome (15, 16). The papaya HSY is 131% larger than the X region (17), making it about 5.4-times larger than the *V. monoica* orthologous region.
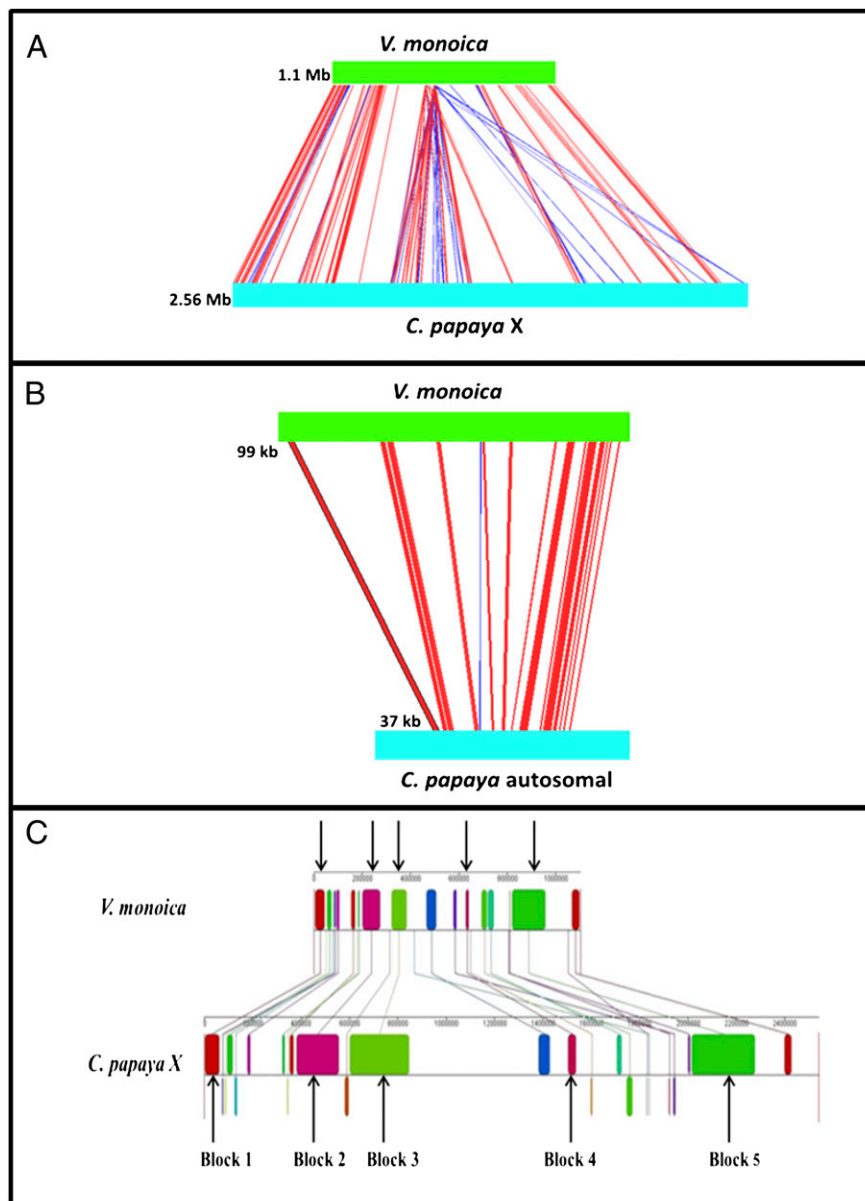
Fig. 1. Expansion of *C. papaya* X chromosome. (*A*) Alignment of the papaya X-specific sequences (2.56 Mb) with orthologous autosomal sequences in *V. monoica* (1.1 Mb). The green horizontal bar indicates the *V. monoica* sequence, and the corresponding *C. papaya* X sequence is shown as the light blue horizontal bar. Red lines indicate homologous regions in the same orientation and blue lines show homology in inverse orientations. The alignment reveals expansion of the papaya X chromosome. (*B*) Alignment of the *V. monoica* autosomal BAC sequence (99 kb) to the corresponding papaya autosomal BAC (37 kb). The *V. monoica* autosomal sequence is expanded, consistent with its larger genome size. (*C*) Five blocks (colored to distinguish the different blocks) in which the X sequence is larger than the *V. monoica* sequence (indicated by arrows) in the syntenic regions of the papaya X-specific sequence and the corresponding *V. monoica* sequence. In all five blocks, the difference is due to gypsy LTR retroelements and unknown repeats.

Gene Content and Order. We annotated genes in the *V. monoica* pseudomolecule and the corresponding papaya X pseudomolecule. In the *V. monoica* pseudomolecule, 19 transcription units were annotated, including 15 intact protein coding genes and 4 pseudogenes (Table 1 and *SI Appendix*, Table S2). Only six intact protein-coding genes are shared between the papaya X-linked region and the corresponding *V. monoica* region; nine are *V. monoica*-specific (Table 1). All four pseudogenes in *V. monoica* correspond to functional alleles in the X and HSY.

In the papaya X pseudomolecule, a total of 18 transcription units were annotated, including 16 intact protein coding genes (the 6 mentioned above that are shared with *V. monoica*, and 10 X-specific genes), and 2 pseudogenes (Table 1 and *SI Appendix*, Table S2). The two pseudogenes are X-specific, only one of which is shared with the papaya HSY.

Overall, of the 27 transcription units annotated in the combined sequences from both species, 8 are missing from *V. monoica*, including 6 X-specific genes and 2 X-specific pseudogenes, and 9 genes are missing from the X sequence but have functional orthologs in *V. monoica* (Table 1). Two X chromosome genes have become pseudogenes, whereas five of the eight X-specific genes (with no homologs in the orthologous *V. monoica* region) have either functional or pseudogene copies in the HSY (*SI Appendix*, Table S2). Only one X-specific gene has a homolog elsewhere in the draft genome of *V. monoica*. Three genes with homologs in papaya autosomes are present in the *V. monoica* sequence corresponding to the X-specific region, but are not found in the X-specific region (*SI Appendix*, Table S2).

In contrast to the sex-specific regions, all five transcripts found in the *V. monoica* autosomal BAC are also present in the papaya autosomal BAC (Table 1 and *SI Appendix*, Table S3), and neither autosomal BAC contains any pseudogenes.

The five genes aligned between the papaya and *V. monoica* autosomal BACs retained the same gene order; the order of 9 of the 10 transcription units shared between the papaya X-linked region and *V. monoica* is conserved, but the X-linked gene, *CpXY^h27*, has rearranged (Fig. 2). *CpXY^h27* is predicted to be a protein kinase and appears to have a functional copy in the papaya X and HSY, as well as in *V. monoica*, on BAC MN86B21 (*SI Appendix*, Table S2).

Gene Density. The papaya autosomal BAC has an average of one gene per 7.4 kb, compared with one gene per 19.8 kb in the homologous *V. monoica* autosomal BAC studied, consistent with the larger *V. monoica* genome (*SI Appendix*, Table S4). However, it should be noted that this papaya BAC gene density is about twice the genome-wide average (one gene per 16 kb) (16). In the sequence corresponding to the sex-determining region, there is a greater difference in gene densities between the two species, and the difference is reversed. In the *V. monoica* pseudomolecule, there is one gene per 46.9 kb versus one per 130 kb in the papaya X-linked pseudomolecule (*SI Appendix*, Table S4). Both densities are considerably lower than in the autosomal region BACs.

Repetitive Sequences. Of the excess length of the X sequence, 49.4% is within three syntenic regions, which include five large blocks of retroelements in papaya compared with fewer retroelements in the *V. monoica* orthologous region (Fig. 1*C* and *SI Appendix*, Fig. S3*A* and Table S5). There is also a region of gypsy LTR retroelement duplication located between syntenic areas 1 and 2 that further enlarge the X-specific region (Fig. 1*A* and *SI Appendix*, Fig. S3*A*). LTR elements account for 45.6% of the X pseudomolecule, whereas these elements only constitute 33.5% of the papaya draft genome (16) (*SI Appendix*, Table S6). The overall repetitive element content of the X pseudomolecule is 65.4%, again greater than the papaya genome-wide average of 52% and much higher than the average of 26.2% from the psudoautosomal-region (16, 17).

Putative Centromeric Location of the Sex-Determining Region. We observed an interesting pattern in repetitive element content of the individual papaya X region BACs and homologous *V. monoica*

Table 1. Summary of the transcription units found in the syntenic regions of the *C. papaya* (*Cp*) X-specific and autosomal BACs and the orthologous *V. monoica* (*Vm*) BACs

Types of sequences	Vm	Cp X or autosome
	Total numbers of sequences	
Combined transcription units (genes and pseudogenes for both species)	27	
Total transcription units	19	18
Transcription unit pairs (present in both species)	10	
Species specific transcription units	9	8
Combined intact protein-coding transcription units	25	
Intact protein-coding transcription units found in each species	15	16
Gene pairs present in both species*	6	
Genes specific to one species	9	10
Pseudogenes	0	4
Copy present elsewhere in the papaya whole genome	3	1
Sequences of unknown origin	6	5
Combined pseudogene transcription units	6	
Total pseudogene transcription units†	4	2
Combined pseudogene-gene pairs	4	
Pseudogene-gene pairs	4	0
Pseudogenes specific to one species	0	2
Combined transcription units‡	5	
Total transcription units (all are functional gene pairs)	5	5

*None of these have a copy elsewhere in the papaya whole genome.
†No pair of sequences is a pseudogene in both species.
‡There are no pseudogenes, and no sequences are specific to one species.

Fig. 2. Alignment of transcriptional units (genes and pseudogenes) shared between *C. papaya* and *V. monoica*. (*A*) Order of the transcriptional units shared between *V. monoica* and the papaya X-specific region. Alternating shades of blue and green blocks indicate different BACs. Red lines denote the 10 shared transcriptional units. (*B*) Alignment of the five genes shared between the papaya autosomal BAC DM105M02 and the *V. monoica* counterpart, MN15I14. Dashes indicate 20-kb intervals. Brackets denote a gene that has expanded because of increased intron size.

BACs. The papaya X-linked BACs gradually increase in repetitive element content, starting from contig Ctg11941 to Ctg08318, with 31% repetitive element content, and peaking in BACs SH65C15-AM136D11 and SH54M13, containing 82% and 87% repetitive elements, respectively. SH65C15-AM136D11 and SH54M13 are located on either side of the gap in the X physical map that could not be filled due to its repetitive nature. The repetitive element content then declines again toward the other border (border B of *SI Appendix*, Fig. S4A), where BAC SH86B15 has 49% repetitive element content (*SI Appendix*, Fig. S5 and Table S7). The *V. monoica* BACs follow the same pattern, with repetitive sequence increasing approaching the region where the X gap is located, peaking at 86% directly adjacent to the gap region, and then decreasing toward the border B region (*SI Appendix*, Fig. S5 and Table S8). These results are suggestive that the location of the centromere may be in this region in both species.

Divergence Times. As mentioned above, it is necessary to establish the suitability of *V. monoica* as an outgroup species for inferring changes in the papaya sex chromosomes. The suitability can be tested by comparing silent DNA sequence substitutions between *V. monoica* sequences, versus orthologs on the papaya X and Y^h chromosomes. If the X and Y^h had diverged long before the split of the species, *V. monoica* sequences in the region should be most similar to the X or to the Y, depending on the genetic basis of the reversion to monoecy. We therefore estimated silent site

substitutions per silent site (K_{sil}). The K_{sil} values between *V. monoica* and papaya X alleles are similar to those between *V. monoica* and the HSY alleles (the difference is not significant, $P = 0.47$), and are also similar to the values between autosomal orthologs from the two species ($P = 0.14$) (*SI Appendix*, Table S9). This similarity indicates that the split must have occurred when the X and Y still had not diverged, thus *V. monoica* is a suitable out group.

The K_{sil} values lead to an estimated divergence time between *V. monoica* and papaya of about 23.2 million y using autosomal gene pairs (*SI Appendix*, Table S9), close to the 27.5 million y previously estimated (18).

Analysis of Nonsynonymous Substitutions. We tested whether the HSY is accumulating an excess number of nonsynonymous substitutions, as predicted if the Y chromosome is degenerating. We did this test by inferring the changes in the X and Y lineages separately and comparing the numbers of per-site nonsynonymous/synonymous substitutions in each lineage. We used complete coding sequences (excluding pseudogenes, which are expected to evolve neutrally, without selective constraints, on both the X and the Y). These analyses were performed on five pairs of complete protein-coding genes that are found in the papaya X and HSY, and in *V. monoica* (the outgroup) (*SI Appendix*, Table S10). The ratios of divergence values for nonsynonymous and synonymous sites, estimated per nucleotide site (K_a/K_s) were also computed; between the *V. monoica* and papaya X genes, K_a/K_s ranges from 0.101 to 0.347, and between *V. monoica* and the papaya HSY gene copies the values were similar, from 0.101 to 0.350, indicating that both the papaya X and HSY genes are evolving under selective constraints and are thus suitable for testing whether the HSY copies are accumulating more deleterious nonsynonymous/synonymous substitutions than the X copies (Fig. 3 and *SI Appendix*, Table S10).

Of the five intact genes shared between all three of the relevant genome regions, *V. monoica*, the papaya X, and the papaya HSY, two (*CpXYh5* and *CpXYh20*) are located in the older stratum (stratum 1 of ref. 17), and three, *CpXYh34*, *CpXYh35*, and *CpXYh36*, in the adjacent region that is collinear between the X and HSY. Not surprisingly, with only two genes from stratum 1, a 2 × 2 analysis finds no significant excess of nonsynonymous substitutions in the stratum 1 HSY lineage, compared with the X lineage, taking into account the numbers of synonymous substitutions in the two lineages. Compared with the collinear genes,

Fig. 3. K_a/K_s ratios for the divergence of five genes shared by *C. papaya* and *V. monoica*. Dark solid bars represent the values when comparing *V. monoica* against the papaya X alleles, light solid bars represent the similar values from comparing the *V. monoica* and HSY alleles, and striped bars show the comparisons between the *V. monoica* genes and papaya autosomal orthologs.

GENETICS

the stratum 1 genes have much higher mean K_a/K_s since their split from their *V. monoica* orthologs, for both the X (0.297 ± 0.07) and HSY (0.312 ± 0.05), versus 0.169 ± 0.06 for the collinear X alleles and 0.171 ± 0.06 for the HSY alleles (calculations based on data from *SI Appendix*, Table S10). However, only the HSY stratum 1 genes' K_a/K_s show a marginal significance from the estimate for the three collinear region genes ($P = 0.045$). Although this finding is consistent with more nonsynonymous substitutions in the HSY sequences, no firm conclusion is possible, and testing for degeneration will require comparing a larger set of genes between *V. monoica* and both the papaya sex-linked regions.

Discussion

Most sex chromosome studies have focused on the evolution of Y chromosomes, and discussions of X chromosomes often emphasize the conservation of the ancestral autosomal gene content and structure (19, 20). Our direct comparison between the recently evolved X-specific region of the papaya X chromosome and the corresponding region of the orthologous autosome in *V. monoica* shows that the papaya X-specific region is not a simple, unaltered descendant of the ancestral autosome; it has expanded in size, genes have been lost and added, and a gene has moved. The difference is striking between this region and the unchanged papaya and *V. monoica* orthologous autosomal regions.

Particularly remarkable is the X-linked region's substantial expansion compared with the orthologous region in *V. monoica*, mostly because of retroelement insertions, in the mere ~6.7 million y since this part of the papaya sex chromosome pair stopped recombining (17). The transposable element accumulation in the X-specific region is consistent with the greater quantity of such insertions in other sex chromosomes (X of mammals and Z of chickens) (14), and can be explained by its lower recombination frequency than that of autosomes; this difference arises because recombination occurs in only females, because of lack of recombination between the X- and Y-specific region in males (or HSY in hermaphrodites), in which the X spends one-third of its time. In *Silene latifolia*, whose sex chromosomes are extremely large and heteromorphic (the Y chromosome is 570 Mb and the X is 420 Mb), these changes have evolved over ~10 million y (21). However, it is unclear if the *S. latifolia* sex chromosome pair is wholly derived from a single ancestral autosome. In papaya, where an orthologous region within a single linkage group can be compared, our results show that the rapid expansion did not involve translocations between different chromosomes.

Transposable element accumulation may lead to chromosomal rearrangements and cause loss-of-function of genes. In mammals, the gene order in the X chromosomes has been conserved across such distantly related species as human, cat, horse, cattle, and elephant, with the exception of mouse (22–26). In the human X and chicken Z, gene trafficking and sequence expansion have been detected, using comparisons with orthologous autosomal segments of other species (12–14). Given its recent origin, the differences in gene content between papaya and *V. monoica* are surprising.

We have presented evidence that the orthologous region in *V. monoica* may be located in the pericentromeric region, as is the HSY and its X counterpart in the papaya sex chromosomes. FISH analyses previously mapped the centromere of the Y^h chromosome near Knob 4 (27); the X chromosome sequence does not have this knob, but a gap in the X physical map remains, because of repetitive sequences, in a location homologous to sequences close to Knobs 3 and 5 (*SI Appendix*, Fig. S4A) (3, 27).

Do the papaya Y coding sequences show signs of genetic degeneration? Because testing for degeneration requires an outgroup to test whether Y-linked sequences are specifically accumulating an excess of deleterious substitutions, analyses are currently limited to coding sequences that can be identified in the HSY, the X,

and in the *V. monoica* outgroup, of which only six coding sequences have so far been identified. The evidence from the analysis of K_a/K_s ratios remains inconclusive. A mild degradation has been detected in the slightly older *S. latifolia* Y, where hundreds of XY gene pairs could be compared since their divergence from an outgroup (28, 29). Given the recent evolution of the collinear region and the limited number of genes in the older stratum 1, it might be difficult to detect the slow processes of sequence degradation. Papaya is, however, ideal for studying genome expansion (which is predicted to be a very early result of recombination suppression), and excellent for future assessments of gene-expression changes.

Methods

***V. monoica* BAC Library Screening and DNA Isolation.** The *V. monoica* BAC library was screened following the protocol of the DIG High Primer DNA Labeling and Detection Starter Kit II (Roche), using probes designed from papaya X BACs located throughout the X-linked region and a probe designed from a papaya autosomal BAC. Twelve positive BACs were confirmed using PCR, 11 corresponding to the X-linked region, and 1 corresponding to the papaya autosomal region (*SI Appendix*, Fig. S4). A miniprep BAC DNA isolation was performed to check the insert size of each BAC via clamped homogeneous electric field gel electrophoresis. The BAC-carrying cells were grown at 37 °C overnight using glycerol stock from a single colony and isolated using the BACMAX DNA purification kit from Epicenter Biotechnologies (cat# BMAX044).

RNA Isolation. RNA for RT-PCR (see below) was isolated from *V. monoica* leaves and flowers using the phenol/chloroform method. After testing the RNA quality using gel electrophoresis, the RNA was treated with DNase, and reverse transcribed into cDNA using ImProm-II Reverse Transcription System from Promega (Cat# A3800).

BAC Sequencing. Eleven *V. monoica* BACs (~1.10 Mb) corresponding to the X-linked region of the papaya X chromosome (~2.56 Mb) and one *V. monoica* BAC (~100 kb) corresponding to a papaya autosomal region (~72.8 kb) were sequenced, using Sanger and 454 sequencing technology, and assembled, using Roche's GS assembler, leaving only a few gaps in the sequences. BAC sequences are available through the National Center for Biotechnology Information (NCBI) (*SI Appendix*, Table S11).

***C. papaya* and *V. monoica* Alignments.** The 11 *V. monoica* BACs, as well as the corresponding 16 BACs and two contigs of the papaya X-linked region, were combined to make *V. monoica* and papaya pseudomolecules and a synteny analysis and dot-plot comparison were performed using Symap with the default settings (30). Chromosome expansion and collinearity analyses were performed using the genome alignment tool Mauve with the default settings (31). Sequence comparisons between the X-linked pseudomolecule and the corresponding *V. monoica* pseudomolecule, as well as the papaya and *V. monoica* autosomal BACs, were carried out using the Artemis Comparison Tool developed by the Sanger Institute with a 500-bp alignment length.

***C. papaya* and *V. monoica* Repeat Analysis.** To annotate repetitive sequence, a combination of TEdenovo from the REPET pipeline (32), RepeatScout (33) and Recon1.05 (34) repeat annotation software were used to identify novel *V. monoica*-specific repetitive elements. Eleven BACs from *V. monoica* corresponding to the X-linked region in papaya, as well as one autosomal BAC and whole-genome shotgun assemblies of *V. monoica* genomic DNA, were used to create a custom repeat dataset. Redundancies in the dataset were eliminated using CD-HIT software (35). *V. monoica* specific repeats were combined with Repbase (36), TIGR plant repeats (http://plantrepeats.plantbiology.msu.edu/index.html), and papaya-specific repeats (37) to generate a custom library. This nonredundant library was used with RepeatMasker to mask repeats in the 12 *V. monoica* BACS. A strict cut-off value of 350 was used to ensure that only true repetitive elements were masked. Repetitive elements for the papaya whole genome and X region were taken from Ming et al. and Wang et al., respectively (16, 17). Given the small sample size of *V. monoica* BACs and low copy number of some repeats, the reported repeat percentages are likely to underestimate the true values.

***V. monoica* Gene Prediction.** Genes were predicted in the *V. monoica* BACs using Genscan (http://genes.mit.edu/GENSCAN.html) and FGENESH (http://

linux1.softberry.com/berry.phtml), as well as homology to papaya-expressed sequence tags (ESTs) and gene models. The papaya autosomal and X BACs were previously annotated by refs. 17 and 38. *V. monoica* genes were confirmed through RT-PCR, with primers designed using primer3 (http://frodo.wi.mit.edu/primer3/) to span at least one intron, if possible. *V. monoica* leaf and flower cDNA were synthesized using Promega ImProm-II Reverse Transcription System (Cat.# A3800). The PCR products were sequenced using Sanger sequencing and manually edited in Sequencher 4.1.10 (Gene Codes Corporation, 2011). The confirmed genes were blasted to the NCBI non-redundant protein database to predict the gene structures and functions through homology. Each individual transcript was translated, and those with premature stop codons were classified as pseudogenes.

K_a/K_s and Divergence Time Analysis. *V. monoica* and papaya gene pairs were manually annotated for exon and intron regions using EST sequences, RT-PCR, and homology with genes in the NCBI database. The sequences were aligned using SeaView v4 (39) and exported into DnaSP v5 (40) to estimate synonymous substitutions per synonymous site (K_s), nonsynonymous substitutions per nonsynonymous site (K_a), and synonymous and noncoding plus synonymous substitutions were used to estimate substitutions per silent site (K_{sil}) using Nei and Gojobori's method (41). Divergence times were calculated using the K_{sil} values, calibrated with the synonymous substitution rate of 4×10^{-9} substitutions per synonymous site per year determined for *Arabidopsis*, a member of Brassicacea, the closest family to Caricaceae (42). *CpXY^h20*, *CpXY^h29*, and *CpXY^h37* were removed from the divergence time analysis because of missing sequence data (incomplete BAC sequences).

V. monoica Genome Survey Sequencing. The *V. monoica* whole genome was survey-sequenced using one lane of Illumina sequencing and the sequences are available at (http://www.life.illinois.edu/ming/LabWebPage/Downloads.html).

1. Liu Z, et al. (2004) A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* 427:348–352.
2. Yu Q, et al. (2008) Low X/Y divergence in four pairs of papaya sex-linked genes. *Plant J* 53:124–132.
3. Na J-K, et al. (2012) Construction of physical maps for the sex-specific regions of papaya sex chromosomes. *BMC Genomics*, 10.1186/1471-2164-13-176.
4. Yu Q, et al. (2008) Recent origin of dioecious and gynodiouecious Y chromosomes in papaya. *Trop Plant Biol* 1:49–57.
5. Ming R, Wang J, Moore PH, Paterson AH (2007) Sex chromosomes in flowering plants. *Am J Bot* 94:141–150.
6. Wu X, et al. (2010) The origin of the non-recombining region of sex chromosomes in *Carica* and *Vasconcellea*. *Plant J* 63:801–810.
7. Peichel CL, et al. (2004) The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Curr Biol* 14:1416–1424.
8. Filatov DA (2005) Evolutionary history of *Silene latifolia* sex chromosomes revealed by genetic mapping of four genes. *Genetics* 170:975–979.
9. Yin T, et al. (2008) Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. *Genome Res* 18:422–430.
10. Spigler RB, Lewers KS, Main DS, Ashman TL (2008) Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome. *Heredity (Edinb)* 101:507–517.
11. Damasceno Junior PC, Da Costa FR, Pereira TNS, Neto MF, Pereira MG (2009) Karyotype determination in three Caricaceae species emphasizing the cultivated form (*C. papaya* L.). *Caryologia* 62:10–15.
12. Charchar FJ, et al. (2003) Complex events in the evolution of the human pseudoautosomal region 2 (PAR2). *Genome Res* 13:281–286.
13. Ross MT, et al. (2005) The DNA sequence of the human X chromosome. *Nature* 434:325–337.
14. Bellott DW, et al. (2010) Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* 466:612–616.
15. Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. *Plant Mol Biol Rep* 9:208–218.
16. Ming R, et al. (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature* 452:991–996.
17. Wang J, et al. (2012) Sequencing papaya X and Y^h chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc Natl Acad Sci USA* 109:13710–13715.
18. Carvalho FA, Renner SS (2012) The phylogeny of the Caricaceae. *Genetics and Genomics of Papaya*, eds Ming R, Moore PH (Springer, Hyderberg).
19. Bull JJ (1983) *Evolution of Sex Determining Mechanisms* (Benjamin/Cummings, Menlo Park, CA).
20. Ohno S (1967) Sex chromosomes and sex linked genes. *Monographs on Endocrinology*, eds Labhart A, Mann T, Samuels LT (Springer, Heidelberg).
21. 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.
22. Murphy WJ, et al. (2007) A 1.5-Mb-resolution radiation hybrid map of the cat genome and comparative analysis with the canine and human genomes. *Genomics* 89:189–196.
23. Raudsepp T, et al. (2004) Exceptional conservation of horse-human gene order on X chromosome revealed by high-resolution radiation hybrid mapping. *Proc Natl Acad Sci USA* 101:2386–2391.
24. Ihara N, et al. (2004) A comprehensive genetic map of the cattle genome based on 3802 microsatellites. *Genome Res* 14(10A):1987–1998.
25. Leticia C, et al. (2009) Physical mapping of the elephant X chromosome: Conservation of gene order over 105 million years. *Chromosome Res* 17:917–926.
26. Sandstedt SA, Tucker PK (2004) Evolutionary strata on the mouse X chromosome correspond to strata on the human X chromosome. *Genome Res* 14:267–272.
27. Zhang W, Wang X, Yu Q, Ming R, Jiang J (2008) DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya. *Genome Res* 18:1938–1943.
28. 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.
29. Chibalina MV, Filatov DA (2011) Plant Y chromosome degeneration is retarded by haploid purifying selection. *Curr Biol* 21:1475–1479.
30. Soderlund C, Nelson W, Shoemaker A, Paterson A (2006) SyMAP: A system for discovering and viewing syntenic regions of FPC maps. *Genome Res* 16:1159–1168.
31. Darling AE, Mau B, Perna NT (2010) progressiveMauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* 5:e11147.
32. Flutre T, Duprat E, Feuillet C, Quesneville H (2011) Considering transposable element diversification in *de novo* annotation approaches. *PLoS ONE* 6:e16526.
33. Price AL, Jones NC, Pevzner PA (2005) De novo identification of repeat families in large genomes. *Bioinformatics* 21:i351–i358.
34. Bao Z, Eddy SR (2002) Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res* 12:1269–1276.
35. Li W, Godzik A (2006) Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22:1658–1659.
36. Jurka J, et al. (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110:462–467.
37. Nagarajan N, et al. (2008) Genome-wide analysis of repetitive elements in papaya. *Trop Plant Biol* 1:191–201.
38. Blas AL, et al. (2010) Cloning of the papaya chromoplast-specific lycopene β-cyclase, CpCYC-b, controlling fruit flesh color reveals conserved microsynteny and a recombination hot spot. *Plant Physiol* 152:2013–2022.
39. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* 27:221–224.
40. Librado P, Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25:1451–1452.
41. Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 3:418–426.
42. Beilstein MA, Nagalingum NS, Clements MD, Manchester SR, Mathews S (2010) Dated molecular phylogenies indicate a Miocene origin for *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 107:18724–18728.

GENETICS