RESEARCH ARTICLE

AMERICAN JOURNAL OF BOTANY

A wild origin of the loss-of-function lycopene beta cyclase (*CYC-b*) allele in cultivated, red-fleshed papaya (*Carica papaya*)[1]

Meng Wu[2,4], Jamicia Lewis[3], and Richard C. Moore[2,5]

PREMISE OF THE STUDY: The red flesh of some papaya cultivars is caused by a recessive loss-of-function mutation in the coding region of the chromoplast-specific lycopene beta cyclase gene (*CYC-b*). We performed an evolutionary genetic analysis of the *CYC-b* locus in wild and cultivated papaya to uncover the origin of this loss-of-function allele in cultivated papaya.

METHODS: We analyzed the levels and patterns of genetic diversity at the *CYC-b* locus and six loci in a 100-kb region flanking *CYC-b* and compared these to genetic diversity levels at neutral autosomal loci. The evolutionary relationships of *CYC-b* haplotypes were assessed using haplotype network analysis of the *CYC-b* locus and the 100-kb *CYC-b* region.

KEY RESULTS: Genetic diversity at the recessive *CYC-b* allele (*y*) was much lower relative to the dominant *Y* allele found in yellow-fleshed wild and cultivated papaya due to a strong selective sweep. Haplotype network analyses suggest the *y* allele most likely arose in the wild and was introduced into domesticated varieties after the first papaya domestication event. The shared haplotype structure between some wild, feral, and cultivated haplotypes around the *y* allele supports subsequent escape of this allele from red cultivars back into wild populations through feral intermediates.

CONCLUSIONS: Our study supports a protracted domestication process of papaya through the introgression of wild-derived traits and gene flow from cultivars to wild populations. Evidence of gene flow from cultivars to wild populations through feral intermediates has implications for the introduction of transgenic papaya into Central American countries.

KEY WORDS *Carica papaya*; Caricaceae; domestication; fruit color; gene flow; selective sweep

Many perennial crop systems are characterized by a "domestication continuum", with extant populations ranging from exploited wild plants to incipient domesticates to highly developed cultivars (Miller and Gross, 2011). The domestication process in long-lived perennials in particular may extend over hundreds or even thousands of years through the introgression wild traits into the domesticated pool (Allaby et al., 2008). In this model of protracted domestication, cultivated populations may consist of individuals from diverse geographic origins and undergo significant gene flow with wild populations. For example, the domestication history of the apple, a perennial fruit tree crop, involved introgression from multiple wild species as the crop spread from its origins in central Asia to western Europe (Cornille et al., 2012).

Gene flow in perennial crop systems may be intentional, as humans breed desirable traits of wild plants into cultivars (Kovach and McCouch, 2008; Miller and Gross, 2011). Through directed breeding between crops and their wild relatives, crop improvement alleles that may have been lost during the domestication bottleneck may be reintroduced from the wild. Alleles controlling traits such as disease resistance or improved quality and/or yield are often targets for directed breeding programs (Luby et al., 2001; Foulongne et al., 2003; Quilot et al., 2004; Hufford et al., 2013; Vázquez Calderón et al., 2014). Alternatively, gene flow may occur unintentionally from cultivars to wild relatives in areas where the two overlap. The introgression of cultivar alleles into wild relatives has been documented in a wide array of annual and perennial systems (Miller and Gross, 2011; Ellstrand et al., 2013). However, the introgression of

[1] Manuscript received 15 August 2016; revision accepted 17 November 2016.

[2] Miami University, Department of Biology, 212 Pearson Hall, Oxford, Ohio 45056 USA; and

[3] Department of Biological Sciences, Alabama State University, Montgomery, Alabama 36104 USA

[4] Present address: Department of Biology, 1001 E Third Street, Bloomington, Indiana 47405 USA

[5] Author for correspondence (e-mail: moorerc@miamiOH.edu); phone (513) 529-4278

doi:10.3732/ajb.1600301

cultivated alleles into wild populations is more difficult to detect in perennial species, due to the higher degree of shared genetic variation that typically exists between perennial species and their wild relatives.

The perennial fruit crop papaya (*Carica papaya* L., Caricaceae) is cultivated in tropical regions worldwide, though it originated in Mesoamerica and can be found there growing in disturbed habitats throughout the region. Wild papayas typically have small, round, yellow-fleshed fruit and are strictly dioecious (Manshardt and Zee, 1994; Brown et al., 2012). Cultivated papayas are characterized by large, oblong fruit with thick flesh that ranges in hue from bright yellow to deep red, depending on the cultivar. Furthermore, most cultivated papaya are gynodioecious, and fruit are derived from selfing hermaphrodites. Feral escapees, identified by their high degree of genetic similarity with exotic cultivars, may also be found growing in the Mesoamerican landscape, providing a potential genetic conduit from cultivars to wild populations (Brown et al., 2012).

Little archaeological evidence exists describing the domestication origins of papaya, though records from early Spanish explorers describe the cultivation of papaya by Mesoamerican cultures (Storey, 1976). Though papaya is a perennial fruit crop, it lacks the prolonged juvenile stage and longevity of other such crops and is sexually, not clonally, propagated (Zohary and Spiegel-Roy, 1975; Miller and Gross, 2011). As such, papaya domestication would likely have initially focused on traits that would improve its sexual propagation, such as the selection for a gynodioecious reproductive system, which allows for hermaphroditic individuals that can produce fruit via self-pollination. Indeed, molecular evolutionary analysis of the papaya sex chromosomes supports the origin of gynodioecy from wild dioecious papaya roughly 4000 yr ago (VanBuren et al., 2015). Subsequent selection for larger fruit, typical of many fruit crops, may have been facilitated in gynodioecious cultivars, as hermaphrodites have longer, elliptically shaped fruit vs. the more rounded shape of female papaya. Diversification traits that are not critical for cultivation, such as fruit shape, color, or fragrance, would have been selected for subsequent to domestication. In annual crops, diversification alleles tend to be loss-of-function alleles with strong phenotypic effects that may be readily selected for by selective breeding (Gross and Olsen, 2010).

The red flesh color of some papaya cultivars is an example of a crop diversification trait. The causative mutation for red fruit color is a recessive, loss-of-function mutation caused by a “TT” frameshift insertion in the coding region of the *CYC-b* gene for chromoplast-specific lycopene beta cyclase, that catalyzes the conversion of lycopene to beta-carotene (Blas et al., 2010; Devitt et al., 2010). It is the accumulation of lycopene that gives red-fleshed papaya, as well as other red-fleshed fruit, such as tomato, their color (Giuliano, 2014). Red fruit color is a likely target for cultivar improvement since some consumers prefer red-fleshed papaya and call them “strawberry papaya” (Blas et al., 2010).

Multiple hypotheses can explain the origin of the recessive loss-of-function *CYC-b* allele (*y*) contributing to red-fleshed cultivated papaya. First, the causative mutation for the *y* allele may have occurred in a yellow cultivar background from a functional allele (*Y*) and the resulting red-fleshed papaya selectively cultivated. Such a stark contrast in phenotype would have been an easy target for selection by papaya growers. It would have taken a single generation of selfing in a hermaphrodite to produce a homozygous, phenotypically red-fruited papaya after the emergence of such a mutation. This mode is similar to the purported origin of diversification alleles reported in many annual cereal crops, such as alleles controlling fragrance and amylose production in certain rice varieties (Olsen and Purugganan, 2002; Kovach et al., 2009). Second, the indel mutation may have originated in the wild and subsequently introgressed into cultivars. Introgression of this wild allele into cultivars may have occurred multiple ways, for example, through selective introduction of the trait from wild red-fleshed germplasm or even passive gene flow of the masked recessive allele into co-occurring crops. Finally, there may have even been a secondary domestication event of red-fleshed papaya that originated in the wild. Under the selective introgression and secondary domestication scenarios, the *y* allele mutation would likely need to rise to appreciable frequency to be noticeable in the wild. Wild dioecious papaya populations exhibit a high frequency of biparental inbreeding, which may accelerate the appearance of red-fleshed, homozygous recessive papaya (Brown et al., 2012). Active introgression of wild germplasm into cultivars by breeders is common for many perennial fruit crops, including papaya (Vázquez Calderón et al., 2014), though incipient domestication events would also favor this scenario.

We took an evolutionary genetic approach to understand the origins of the *y* allele in red papaya cultivars. First, we compared the levels and patterns of genetic diversity at the *CYC-b* locus and six loci in the surrounding 100-kb genomic region in a panel of yellow and red cultivars and in a diverse sampling of wild papaya from various regional populations in Costa Rica. From these analyses, we deduced a very recent origin of the *y* allele and detected the footprint of a large selective sweep in the 100-kb region containing the *y* allele in red cultivars. In addition, we were able to infer the origins of the *y* allele through phylogenetic and haplotype network analyses of *CYC-b* haplotypes in wild and cultivated individuals. Our analyses suggest that the *y* allele initially arose in the wild and was subsequently introgressed, either actively or passively, into cultivars, reminiscent of the domestication continuum that characterizes other Mesoamerican perennial fruit crop systems (Miller and Schaal, 2005; Hughes et al., 2007; Galindo-Tovar et al., 2008). Furthermore, we present evidence that subsequent to the introgression of the *y* allele into cultivars, the allele has been reintroduced to wild populations via feral escapees.

MATERIALS AND METHODS

Plant materials—A total of 48 wild papaya individuals were sampled for population genetic analyses from five geographically dispersed regional populations in Costa Rica: Caribbean, Northwest Pacific, Nicoya Peninsula, Central Pacific, and Southwest Pacific (6–8 individuals per regional population; Appendices S1, S2, see Supplemental Data with the online version of this article; also see Brown et al., 2012). Costa Rica is in the southern part of the range of papaya, which grows throughout Mesoamerica from southern Mexico to Panama. Thirty-three of these individuals were females, of which 15 had mature, yellow-fleshed fruit at the time of collection (Appendix S2). Six individuals (one hermaphrodite and five females) found in these populations shared high genetic similarity with cultivated papaya based on a previous analysis using the program STRUCTURE (Brown et al., 2012) and were classified as feral individuals. Only one of these feral individuals, the hermaphrodite, had mature fruit at the time of collection, and it had red-fleshed fruit (Appendix S2). One to two feral individuals were found in each regional population except the Caribbean population (Appendix S2).

In addition, 15 cultivar individuals were sampled, including seven with yellow-fleshed fruit and eight with red-fleshed fruit (online Appendix S3). These included improved cultivars, which are commercial cultivars and breeding lines, and unimproved cultivars collected from different regions in Central America, including Costa Rica and Panama (assignments based on those of Kim et al. [2002]). Improved cultivars are the result of selective breeding while unimproved cultivars represent lines without a history of selective breeding, some of which, especially those derived from Mesoamerica, were derived from individuals growing untended in the landscape (Kim et al., 2002). Cultivar tissues were provided by Dr. Qingyi Yu (Texas A&M, AgriLife Research Center at Weslaco) and Dr. Francis Zee (USDA Agricultural Research Station, Hawaii). Young leaf tissue was stored at −80°C before DNA extraction. Genomic DNA was isolated from stored leaf tissues using the DNeasy Plant Mini Kit (Qiagen, Valencia, California, USA).

Sequencing of the* CYC-b *locus and its flanking genomic segments—The *CYC-b* gene consists of a single exon of 1485 bp. The dominant functional allele coding for yellow flesh (*Y*) is found in all yellow-fleshed individuals, while the recessive allele coding for red flesh (*y*) is characterized by a loss-of-function TT dinucleotide insertion at position 830 bp (online Appendix S4). A genomic segment containing the entire coding region (1675 bp) of the lycopene beta-cyclase (*CYC-b*) gene was resequenced in each individual. In addition, segments of six flanking genes distributed across a ~100-kb region centered on *CYC-b* were also resequenced. The six adjacent genes are 50 kb upstream (586 bp), 30 kb upstream (1109 bp), 19 kb upstream (983 bp), 16 kb downstream (1012 bp), 26kb downstream (680 bp) and 53kb downstream (631 bp) relative to the *CYC-b* (Appendix S4).

PCR primers were designed using Primer3 (online Appendix S5; Rozen and Skaletsky, 2000). Due to the limited range of cycle sequencing reactions (<1200 bp), the amplified *CYC-b* region (~1800 bp) was divided into two smaller segments for amplification (Appendix S4). For some flanking loci, internal sequencing primers were designed to amplify regions flanking indel polymorphisms in heterozygotes. PCR was performed using GoTaq Colorless Master Mix (Promega, Madison, Wisconsin, USA) and consisted of initial denaturation for 2 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at 56°C, 30 s at 72°C, and final extension for 5 min at 72°C. PCR products were cleaned by using a combination of 5 U exonuclease I and 0.5 U shrimp alkaline phosphatase at 37°C for 40 min followed by enzyme inactivation at 80°C for 15 min. Purified PCR products were cycle sequenced using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, California, USA), followed by a postsequencing clean-up step using ethanol precipitation.

Sequences were assembled and aligned using BioEdit and BioLign 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 numbers KY175231–KY175901).

Molecular population genetic analyses—Alleles for each individual were statistically phased using the program PHASE as implemented in the program DnaSP v5.10 with output probability threshold for genotypes and haplotypes set to 0.9 (Librado and Rozas, 2009). After phasing, individuals were genotyped at the *CYC-b* locus. The *Y* allele lacks the TT frame-shift insertion found in the *y* allele. Individuals were either homozygous dominant (*YY*), heterozygous (*Yy*) or homozygous recessive (*yy*). Silent-site nucleotide diversity was estimated as π, or pairwise nucleotide diversity (Nei, 1987), and θ_w, or Watterson's estimator, based on the number of segregating sites (Watterson, 1975) for all loci using DnaSP v5.10 (Librado and Rozas, 2009). Silent-site nucleotide diversity was also estimated for 39 unlinked autosomal loci in wild and cultivar accessions (originally sampled by Wu and Moore [2015]). While the sampled wild accessions were the same used for this study, only a subset of yellow-fleshed (total number of alleles, $n = 8$) and red-fleshed ($n = 4$) cultivars had available sequence, as the original study from which the data are drawn focused on genetic diversity in wild accessions (Wu and Moore, 2015).

For the haplotype analysis of the 100-kb *CYC-b* genomic region, the different sequenced unlinked segments were concatenated first and then statistically phased to construct the haplotypes. To test for evidence of a selective sweep, we calculated the extended haplotype homozygosity (EHH) statistic for both red-fleshed (*yy*) and yellow-fleshed (*YY*/*Yy*) major groups across the target 100-kb region surrounding the *CYC-b* using the program SWEEP (Sabeti et al., 2002). The EHH statistic is used to quantify the probability that two chromosomes chosen at random are identical at progressively increasing distances from the core region. If a genomic region has recently been a target of positive selection, there would be increased linkage disequilibrium (LD) and reduced decay of homozygosity. The EHH value is expected to decay with increasing distance from the *CYC-b* locus (the core region in this study) as recombination breaks down haplotype structure. The haplotype decay pattern was visualized by haplotype bifurcation diagrams using SWEEP (Sabeti et al., 2002).

Haplotype network analyses—A haplotype network of all sequenced accessions was constructed by analyzing the phased alignments of the *CYC-b* locus and 100-kb *CYC-b* region using statistical parsimony as implemented in the TCS program (Clement et al., 2000). To provide an independent neutral-marker assessment of population structure, we also used TCS to construct a haplotype network for a concatenated alignment of four previously reported unlinked autosomal loci (*EST5*, *EST6*, *EST7*, and *EST14* of Weingartner and Moore [2012]) from 45 wild males from the same regions of Costa Rica as the individuals sampled for *CYC-b*. Fifteen cultivars were sampled at these autosomal loci, included four unimproved cultivars and 11 improved cultivars (Appendix S3). Of these improved cultivars, five of eight red cultivars and six of eight yellow cultivars are shared with the *CYC-b* cultivar data set.

RESULTS

Evidence of a selective sweep around the loss-of-function* CYC-b *allele—There was little genetic diversity in the 100-kb *CYC-b* region containing the recessive allele (*y*) coding for red flesh in cultivars, feral, or wild individuals. Nucleotide polymorphism was absent at the *y* allele and flanking loci in red-fruited cultivars and greatly reduced in wild and feral individuals (mean silent site diversity, π_{sil}, was 0.0006 ± 0.0003 SE in wild individuals and 0.0002 ± 0.0001 SE in feral individuals averaged over all loci; Fig. 1). In contrast, diversity

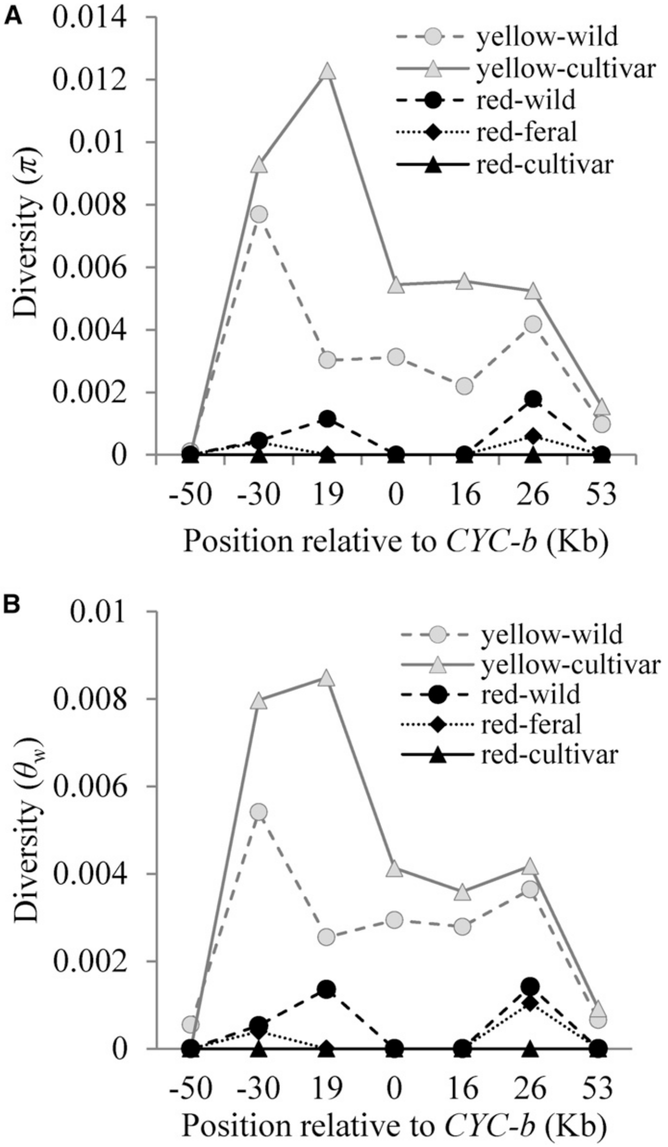

FIGURE 1 Graphs of nucleotide diversity estimates (A) π and (B) θ_w for seven genomic segments distributed across 100-kb region centered on the *CYC-b* locus. Represented are diversity estimates for wild, feral, and cultivar accessions with haplotypes containing the dominant *Y* allele coding for yellow-fleshed papaya (yellow-wild, yellow-cultivar) and haplotypes containing the recessive *y* allele coding for red-fleshed papaya (red-wild, red-feral, red-cultivar).

was higher at most loci in the 100-kb region surrounding the *Y* allele in yellow-fruited cultivars, wild and feral individuals (mean π_{sil} = 0.0056 ± 0.0018 SE in cultivars, mean π_{sil} = 0.0030 ±0.0011 SE in wild individuals, and mean π_{sil} = 0.0043 ± 0.0021 SE in feral individuals averaged over all loci; Fig. 1A). When diversity was estimated as θ_w, similar levels and patterns of diversity were observed (Fig. 1B).

The lack of diversity among cultivar, feral, and wild *y* alleles suggests a recent, shared origin of the genomic region containing the *CYC-b* loss-of-function allele. It also suggests strong selective pressure for this genomic region in red cultivars. Consistent with a selective sweep, the haplotype diversity of the 100-kb *CYC-b* region containing the *y* allele was much lower than that of containing *Y* alleles (online Appendix S6). This reduction in diversity in the genomic region containing the *y* allele in red-fleshed cultivars is a characteristic of that region. The distribution of diversity estimates (π_{sil}) at 39 unlinked autosomal loci in wild accessions was higher than that for a subset of yellow-fleshed and red-fleshed cultivars, although the distribution of diversity estimates for yellow cultivars did not significantly vary from that of red cultivars (Wilcoxon rank sums 1-way test, χ^2 = 18.14, df = 2, *P* = 0.0001; Wilcoxon paired-sample test, *P* = 0.003 [wild vs. yellow], *P* < 0.0001 [wild vs. red] and *P* = 0.34 [red vs. yellow]; online Appendix S7). In contrast, the distribution of diversity estimates for loci in the 100-kb *CYC-b* region for both wild accessions and yellow-fleshed cultivars was significantly higher than that for red-fleshed cultivars, while the distribution of diversity estimates for wild accessions and yellow-fleshed cultivars did not significantly vary (Wilcoxon rank sums 1-way test, χ^2 = 18.14, df =2, *P* = 0.0001; Wilcoxon paired-sample test, *P* = 0.003 [wild vs. yellow], *P* < 0.0001 [wild vs. red] and *P* = 0.34 [red vs. yellow]; Appendix S7). This result suggests that the greatly reduced diversity seen in the 100-kb *CYC-b* region in red cultivars compared with yellow cultivars in particular is a characteristic of that region and not of the genome as a whole.

To investigate explicitly the signature of selection around the *y* allele, we calculated the extended haplotype homozygosity (EHH) statistic for loci in the 100-kb region (Fig. 2). A recent selective sweep will elevate EHH around the selected locus to a distance that depends on the strength of selection and the rate of recombination. EHH was elevated across the entire 100-kb region surrounding the *y* allele compared with the *Y* allele, indicating high linkage disequilibrium in the genomic region containing the *y* allele driven by a strong selective sweep (Fig. 2A). This pattern is also clear when comparing the haplotype bifurcation diagrams for *y* and *Y* alleles (Fig. 2B, C). There is a core haplotype devoid of polymorphism across the 100-kb region surrounding the *y* allele, whereas the *Y* allele core haplotype breaks down immediately outside of the *CYC-b* locus.

Genealogical relationships among* CYC-b *haplotypes suggest a wild origin of the loss-of-function* y *allele—Statistical parsimony was used to construct a haplotype network of the *CYC-b* locus. There were 14 single nucleotide polymorphisms (SNPs) and 15 haplotypes found in our population genetic survey of *CYC-b* (Appendix S8). The *y* allele was associated with a single haplotype (15) and found primarily in red-fleshed cultivars and feral individuals, but also in some wild individuals. These wild individuals were found in the Caribbean, Northwest Pacific and Central Pacific regional populations of Costa Rica (Fig. 3A). Haplotype 15 was derived from haplotype (1) by a single mutation (i.e., the TT insertion mutation). Haplotype 1 contained a dominant *Y* allele and was found in Caribbean and Central Pacific regional populations of Costa Rica. Haplotypes from both unimproved and improved yellow-fleshed cultivars also stemmed from this haplotype (2, 4, 14). Haplotypes found in yellow-fleshed cultivars (2, 3, 4, 12 13, and 14) were distinguished by multiple independent mutational steps. Most of these haplotypes were one to two mutational steps from wild haplotype 1, though haplotype 12 was eight mutational steps removed from haplotype 1 and most similar to haplotype 6 consisting of wild individuals from the Caribbean and Nicoya Peninsula regions of Costa Rica.

In a larger haplotype network for the phased concatenated 100-kb region based on 93 SNPs and 60 haplotypes, the *y* allele was found

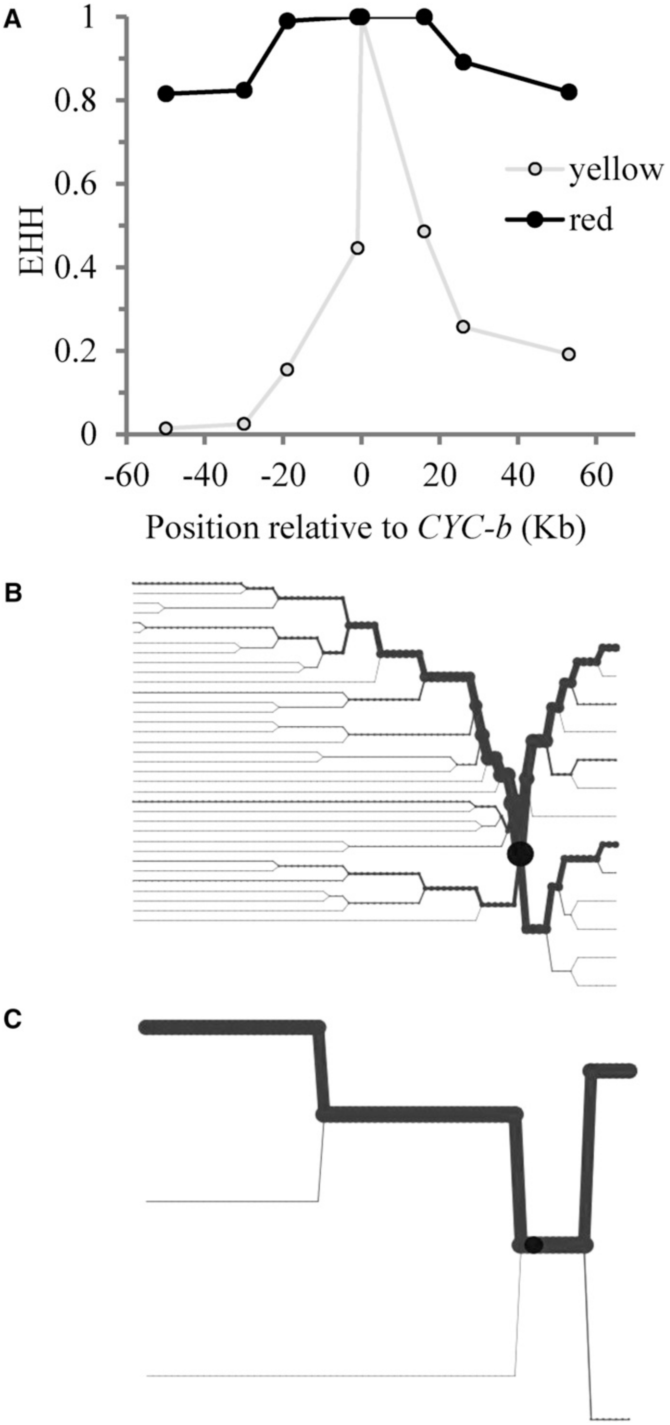

FIGURE 2 (A) Extended haplotype homozygosity (EHH) decay for the haplotypes containing the *Y* and *y* alleles as a function of distance from *CYC-b*. Haplotype bifurcation diagrams for the *Y* (B) or *y* (C) haplotypes, in which the central dark dot represents the *CYC-b* core haplotype, branches represent haplotype divergence, and the thickness of the branches is proportional to the number of chromosomes containing that haplotype.

in four haplotypes (42, 43, 48, and 50) connected by one to two mutational steps (Fig. 4). This cluster is most closely related to haplotypes prevalent in the central Pacific, northwest Pacific and Nicoya Peninsula regions of Costa Rica. The haplotype network around these haplotypes was highly reticulate, likely due to recombination between loci in this region (Fig. 4). The most common haplotype containing the *y* allele (43) was found in all red-fleshed cultivars, seven feral individuals and two individuals from the northwestern Pacific region. Two haplotypes containing the *y* allele (42 and 50) were derived from haplotype 43 by a single mutational step and found mainly in feral individuals, but also in individuals from the Caribbean and Central Pacific populations. The relationships of haplotype 48, which contains the *y* allele and was found in a Caribbean individual, to the cultivar haplotype 43 was ambiguous due to homoplasy of the loss-of-function mutation (the stars in Fig. 4). Most of the improved yellow cultivars were found in haplotypes that are many mutational steps separated from haplotypes containing the *y* allele. These haplotypes (53–57) were most similar to wild haplotypes from the Caribbean, central Pacific, and southwestern Pacific regions of Costa Rica. The unimproved cultivars from Costa Rica and Panama were dispersed throughout the network, as are two alleles from the improved cultivar Higgins (Fig. 4).

To determine whether the genealogical patterns observed for the *CYC-b* locus and 100-kb region reflect the neutral population history of red and yellow cultivars and wild populations, we constructed a haplotype network for a concatenated alignment of four previously reported autosomal loci (*EST5*, *EST6*, *EST7*, and *EST14* of Weingartner and Moore, 2012). These loci are unlinked to the *CYC-b* locus and to each other, are considered to be neutrally evolving, and are not suspected to underlie traits in domesticated papaya. There were 14 single nucleotide polymorphisms (SNPs) and 33 haplotypes found in the EST data set, similar in size to the *CYC-b* locus data set (online Appendix S9). The haplotype network for the concatenated neutral autosomal loci reflects a different genealogical pattern than observed for the *CYC-b* locus (Fig. 3B). Both red- and yellow-fleshed cultivars were found primarily in a single haplotype (haplotype 20) along with two alleles from a single feral individual. This haplotype was only three mutational steps from a major wild haplotype (13) that was found in four of the five regional populations. Alleles for the unimproved red cultivar UH918 were found in haplotype 31 that was also three mutational steps from the cultivar haplotype, while alleles for the unimproved yellow cultivar 928H, collected from Costa Rica, were five mutational steps from the cultivar haplotype and found in the second major wild haplotype 1.

In addition to this haplotype analysis of four unlinked autosomal loci, previous analyses of population structure including most of our wild and cultivar accessions based on 20 unlinked neutral simple sequence repeat (SSR) loci (Brown et al., 2012) support the separation of wild and cultivar populations into two distinct genetic populations (online Appendix S10). In this analysis, two clusters of shared genetic ancestry were identified, one belonging to wild individuals and one to cultivars. Unimproved cultivars sampled in our analysis shared 90% identity with wild individuals, suggesting a shared origin with wild papaya and consistent with the haplotype network analyses of the *EST* and *CYC-b* analyses. Interestingly, the dioecious cultivar DREW also shared high genetic ancestry with native papaya populations. This and its dioecious nature suggest a recent and separate origin of this cultivar apart from gynodioecious cultivars. Feral individuals shared between 80 and 99% genetic identity with cultivars, consistent with the close relationships we found between cultivars and ferals in the *EST* and *CYC-b* haplotype analyses.

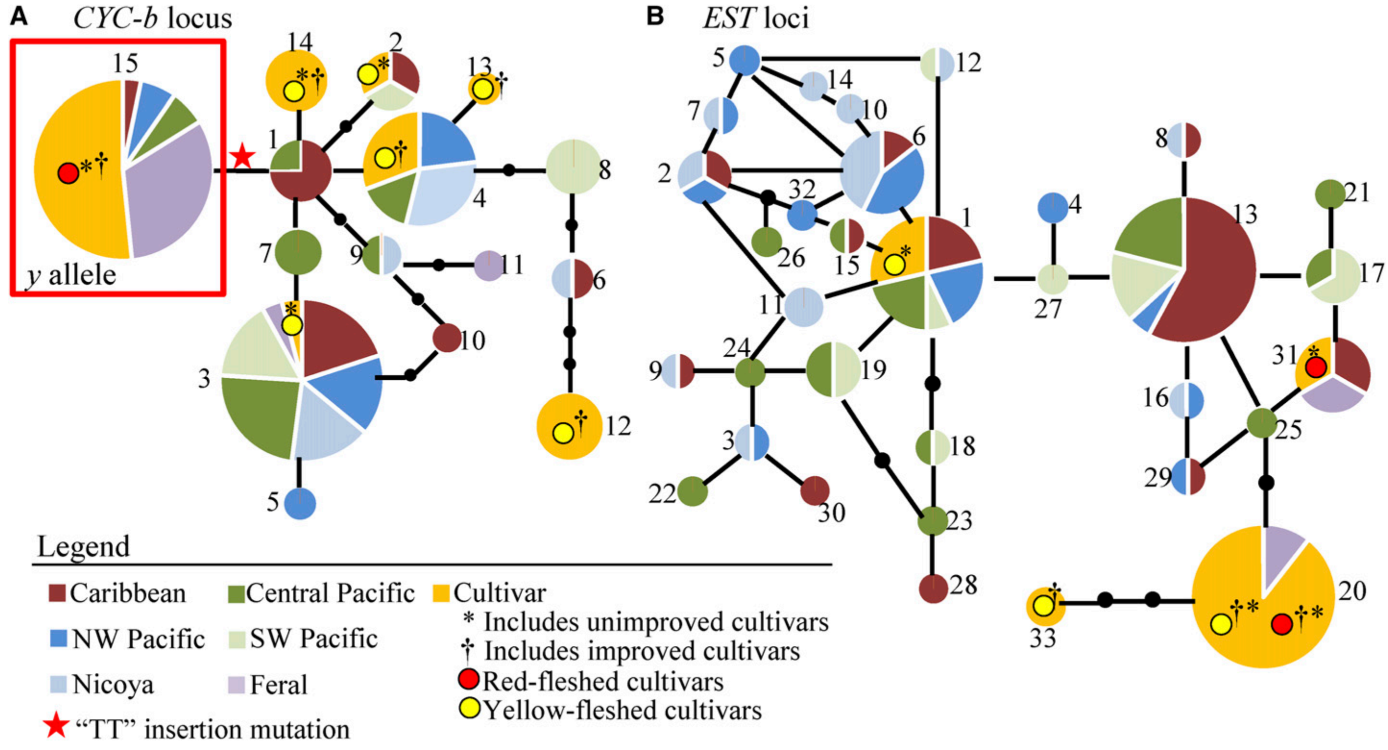

FIGURE 3 Haplotype networks of (A) *CYC-b* locus haplotypes and (B) the concatenated sequence of four neutral autosomal loci. Haplotypes are connected by mutational steps (lines) and inferred haplotypes (small black circles). Haplotype circle diameter is proportional to the number of individuals possessing that haplotype. Haplotype identification number is adjacent to each haplotype circle. Haplotypes are proportionally subdivided according to one of five Costa Rican regional population (wild), feral, or cultivar origin. Haplotypes with improved (†) and unimproved (*) cultivars are indicated. The locations of yellow-fleshed and red-fleshed cultivars are indicated by smaller yellow and red circles, respectively. In (A), the location of the causative mutation for the *y* allele is indicated by a red star and haplotype 15, which contains the loss-of-function *y* allele, is outlined by a red square.

The loss-of-function* y *allele is found in wild populations at low frequencies—Red- and yellow-fleshed cultivars were exclusively homozygous for their respective alleles. In the wild, however, the *y* allele was found at low frequency. It was found only in the Caribbean, northwestern Pacific, and central Pacific regional populations and at frequencies ranging between 8 and 20% (Fig. 5, online Appendix S11). In those populations, the *y* allele was found exclusively in a heterozygous state, and all 15 sampled wild individuals with mature fruit had yellow flesh (Brown et al., 2012). The observed genotypic frequencies in the wild were consistent with Hardy–Weinberg predictions based on the allele frequency of the *CYC-b* in the wild (P = 0.68). In contrast, the *y* allele was highly frequent (~80%) in feral individuals with a frequency. Likewise, most feral papayas were homozygous recessive (*yy*, 63%), and the rest were heterozygous (*Yy*, 37%). There were no homozygous yellow (*YY*) feral individuals genotyped (Fig. 5).

DISCUSSION

Our population genetic analyses support the recent emergence of the *y* allele and its selection in red-fruited cultivation. There was a lack of diversity in the 100-kb region surrounding the *y* allele in red-fleshed cultivars compared with yellow-fleshed cultivars and wild accessions, consistent with recent and strong selection for the genomic region containing the *y* allele. This pattern was specific to the 100-kb *CYC-b* region as comparisons of diversity estimates between red-fleshed and yellow-fleshed cultivars at unlinked autosomal loci are not statistically different. The diversity estimates for cultivars for autosomal loci were lower than that observed in the wild; however, diversity was likely underestimated for cultivars at these autosomal loci due to the small sample size used for cultivars in this analysis (Yang, 1996; Sinclair and Hobbs, 2009; Gorbachev, 2012; Fumagalli, 2013; Subramanian, 2016). Furthermore, the extended haplotype structure observed in the 100-kb region containing the *y* allele was consistent with a rapid spread of this allele in cultivation due to strong artificial selection for red-fruited cultivars.

There was also little to no diversity in the *CYC-b* 100-kb region containing the *y* allele in wild or feral individuals; the 100-kb haplotypes in red cultivars and these individuals were either identical or differ by two to four polymorphisms, consistent with a recent common origin of the haplotypes containing the *y* allele. If the *y* allele was selected from standing variation in the wild, we might expect to find a more diverse haplotype pool in the larger 100-kb region surrounding the *y* allele in wild individuals, similar to what we found in the 100-kb region surrounding the *Y* allele in wild individuals; however, we observed little variation at the *y* locus in wild individuals. This result suggested that the *y* allele might have originated so recently that it has not accrued appreciable polymorphism or undergone extensive recombination in the 100-kb region within wild *Y* haplotypes.

It is not possible for us to infer the origin of the *y* allele from our diversity analysis alone given the nearly identical *y* allele sequence in wild and cultivated papaya. However, comparison of the haplotype

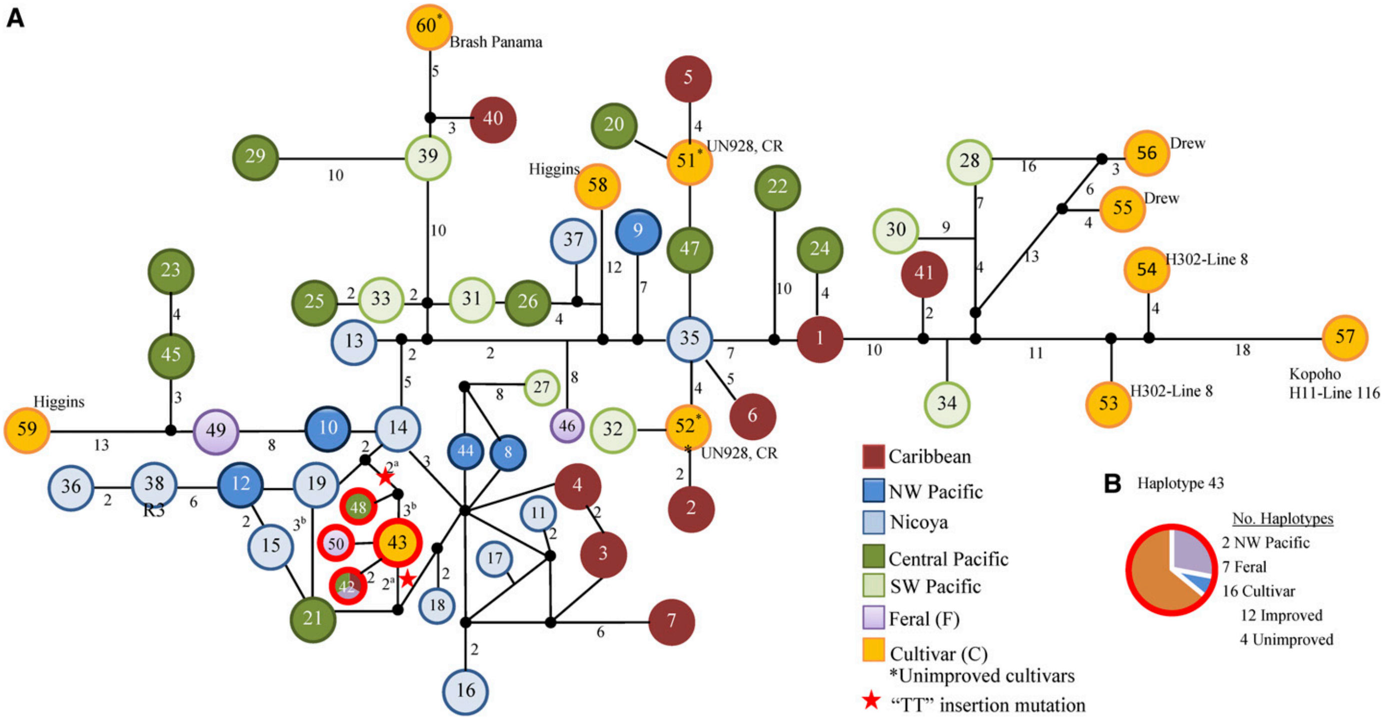

FIGURE 4 (A) Haplotype network of haplotypes from the 100-kb *CYC-b* region. Haplotypes are connected by mutational steps (lines) and inferred haplotypes (small black circles), with the number of mutational steps indicated adjacent to each line. Haplotype identification numbers are within each haplotype circle. The location of the causative mutation for the *y* allele is indicated by a star. Because of homoplasy, this location is in one of two positions. Haplotypes containing the *y* allele for red flesh are outlined in red. Haplotype diameter is proportional to the number of individuals possessing that haplotype, most ranging from 1 to 3 individuals, with the notable exception of haplotype 43, which is found in 25 individuals (see B). Haplotypes are proportionally subdivided according to one of five Costa Rican regional populations (wild), feral or cultivar origin. Haplotypes with unimproved (*) cultivars are indicated. The name of each unimproved cultivar is indicated next to the haplotype where they are found. (B) The proportion of wild (North Pacific regional population, blue), feral (light purple), and cultivar (orange) individuals in the most common haplotype (43) that contains the *y* allele for red flesh.

relationships at the *CYC-b* locus to those at unlinked, neutral autosomal loci can help us distinguish between a wild and cultivated origin of the *y* allele. If the *y* haplotype originated in cultivation from a previously yellow-fruited cultivar, red- and yellow-fleshed cultivars would share a common ancestor at neutral loci and the *y*-bearing haplotype would be derived from a *Y*-bearing haplotype found in cultivars (Fig. 6, case 1). Alternatively, if the *y* haplotype originated in the wild and was introgressed into cultivars via selective breeding or passive introgression, we would find that red and yellow cultivars share a common ancestor at neutral loci, but that the *y* haplotype would be derived from a wild *Y* haplotype (Fig. 6, case 2). A third possibility is that red cultivars were domesticated independently from yellow cultivars, in which case haplotypes for both neutral loci and the *y* allele would be derived from wild haplotypes (Fig. 6, case 3).

The phylogenetic patterns and haplotype network reconstructions supported the introgression of the *y* haplotype from the wild into cultivars (case 2, Fig. 6). The *y* haplotype in wild, feral, and red-fruited cultivars was most similar to a wild *Y* haplotype (1 in Fig. 3A), being separated by only one or two mutational steps (Fig. 3). Many yellow-fruited cultivars (both improved and unimproved) were also only one or two mutational steps from the wild haplotype 1 in the *CYC-b* locus haplotype network. As such, it was difficult to distinguish cultivar relationships based on the *CYC-b* locus haplotype network alone. However, these relationships were more clearly resolved in the haplotype network analysis of the larger 100-kb *CYC-b* region, which supported independent origins of this region in red and yellow cultivars from diverse wild haplotypes (Fig. 4). The independent origins of *CYC-b* haplotypes in cultivars were not due to separate domestication origins of red and yellow cultivars, as the haplotype networks of four, unlinked neutral loci and our STRUCTURE analysis supported a single shared ancestry of improved red- and yellow-fleshed, gynodioecious cultivars (Fig. 3B, Appendix S10). The inference of haplotype relationships from haplotype networks constructed using statistically phased haplotype inferences from concatenated gene data sets, such as what we used for the EST and *CYC-b* 100-kb regions, can give rise to increased reticulation in the networks. While the relationships among many wild, minor haplotypes were obscured due to reticulation, the relationship among red-fleshed and yellow-fleshed cultivars in these networks conformed to that presented in our proposed case 2 (Fig. 6B).

It is clear based on these haplotype networks that the *y* allele likely originated in the wild. What is less clear based on our current data set, however, is whether the introgression of the *y* allele into gynodioecious cultivars was through selective breeding of wild, red-fleshed individuals into cultivars or via passive gene flow between wild and co-occurring papaya crops. Active introgression would presumably require the discovery of a relatively rare wild homozygous *yy* individual, and we might expect red-fleshed wild papaya to be of high enough frequency to make discovery more

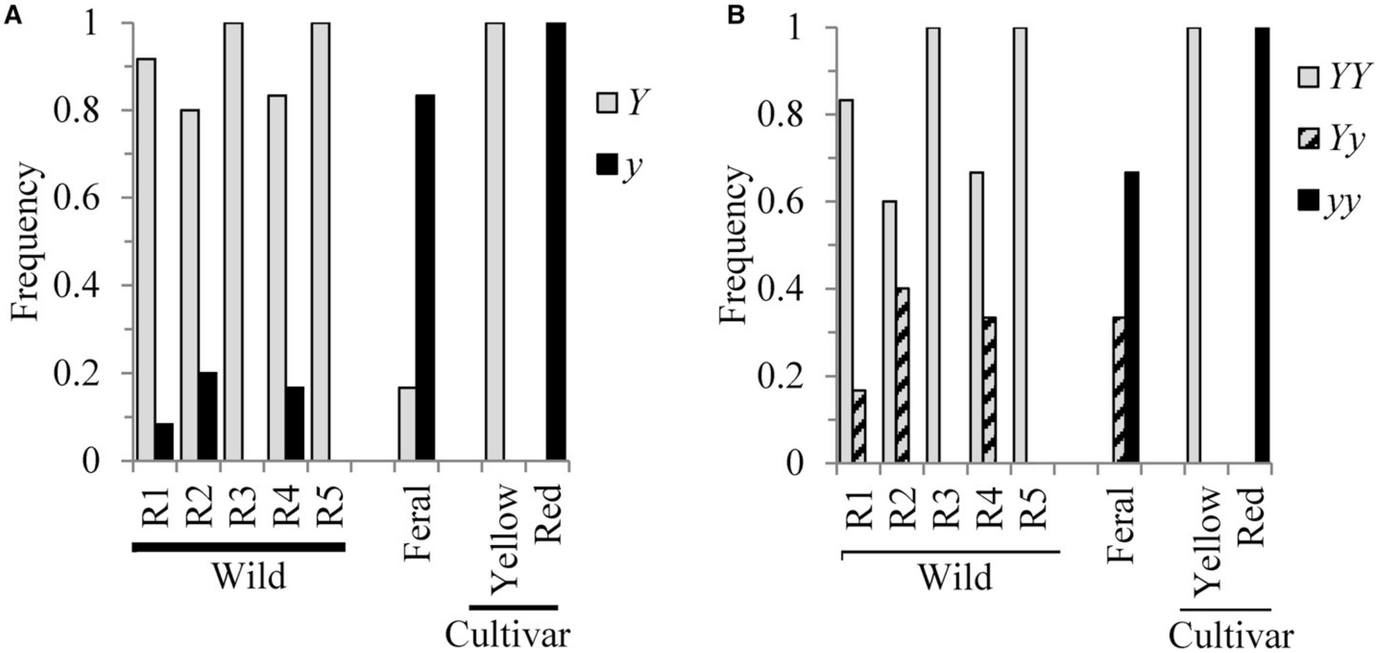

FIGURE 5 *CYC-b* allelic (A) and genotypic frequencies (B) in wild, feral and cultivated papayas. Wild populations are abbreviated R1 (Caribbean), R2 (Nothwest Pacific), R3 (Nicoya Peninsula), R4 (Central Pacific), and R5 (Southern Pacific). In (A), the frequency of the *Y* allele is indicated by gray bars, and the frequency of the *y* allele is indicated by black bars. In (B), the frequency of *YY* individuals is indicated by gray bars, *Yy* by striped bars, and *yy* by black bars.

probable. However, the *y* allele was found at low frequency in regional populations, from 0 to 20%, and red-fleshed papayas are uncommon in wild populations. Given these frequencies, we would expect to find one to four naturally occurring homozygous recessive (*yy*) red-fleshed papaya out of every 100 females surveyed. Thirty-two females with ripe fruit were identified in previous morphological analyses of Costa Rican papaya, and all had yellow or yellow to orange flesh; however, the absence of red-fruited papaya seems statistically reasonable given the relatively small sample size. In addition, if the *y* allele is of recent origin in the wild, as suggested by the diversity analysis, we might expect to find it at low frequency in wild populations. It may also be that the *y* allele is found in higher frequencies in Mesoamerican populations outside of Costa Rica, and broader sampling of Mesoamerican wild papaya populations would address this possibility. However, if there is no strong selective pressure for red-fleshed wild papaya, for example, from birds or mammal consumers, the *y* allele may be persisting at low frequencies solely due to drift. Importantly, while the frequency of the *y* allele in the wild is low, it does not rule out the possibility that the *y* allele originated in the wild and was introgressed into cultivars.

In contrast to the low frequency of the *y* allele in wild papaya populations, there was a high frequency of the *y* allele in feral papaya (~80%), and every feral papaya had the *y* allele, either in a homozygous state (~60%) or as heterozygotes (~35%). The high frequency of the *y* allele in ferals makes sense if most ferals were derived from red-fleshed cultivars, where the *y* allele was fixed. That ~35% of ferals were heterozygous suggests gene flow between ferals and wild papaya occurs, accounting for the presence of *Y* alleles in feral individuals. The converse is also likely true, that cultivar alleles may make their way to wild individuals through feral intermediates. Movement of cultivar alleles to wild individuals was supported by the haplotype network analysis of the 100-kb *CYC-b* region, as haplotype 42 was found in both feral and wild individuals and was derived from the red-fleshed cultivar haplotype 43 (Fig. 4).

Ultimately, the story of papaya domestication may be more gradual than abrupt, as seen in other perennial Mesoamerican fruit tree crops, such as *Spondius purpurea* (Miller and Schaal, 2005, 2006; Miller and Gross, 2011), *Leucaena* (Hughes et al., 2007), and avocado (Galindo-Tovar et al., 2008; Clegg et al., 2009; Miller and Gross, 2011). Diversity around the *Y* allele was higher in both cultivated and wild papaya, and estimates of genetic diversity in wild and cultivated papaya based on neutral genetic markers were also roughly equivalent (Brown et al., 2012). Diversity at some loci in the 100-kb region surrounding the *Y* allele was actually lower in wild papaya, though not significantly so (Wilcoxon paired test, Appendix S7). This observation may reflect the geographic limitations of our sampling, which was more toward the southern end of papaya's native distribution. Nonetheless, comparable levels of wild and cultivated diversity suggested the lack of a strong domestication bottleneck in papaya and a pattern of continuous domestication through the occasional introduction of wild papaya germplasm into cultivated stock (Miller and Schaal, 2005; Hughes et al., 2007; Galindo-Tovar et al., 2008; Miller and Gross, 2011).

Furthermore, the phylogenetic analysis of the *CYC-b* locus and the 100-kb *CYC-b* region supported ongoing experimentation by breeding wild germplasm and cultivars. For example, haplotypes for so-called unimproved cultivars were highly similar to wild haplotypes for both the *CYC-b* regional neutral loci haplotypes. Unimproved cultivars might be directly derived from wild- or feral-growing individuals, such as lines UH928 and UH918, which were collected from the Costa Rican landscape (Kim et al., 2002). Population structure analysis identified UH928 as having the genetic signature of wild papaya, whereas UH918 was most similar to exotic cultivars and thus likely represents a feral individual (Brown

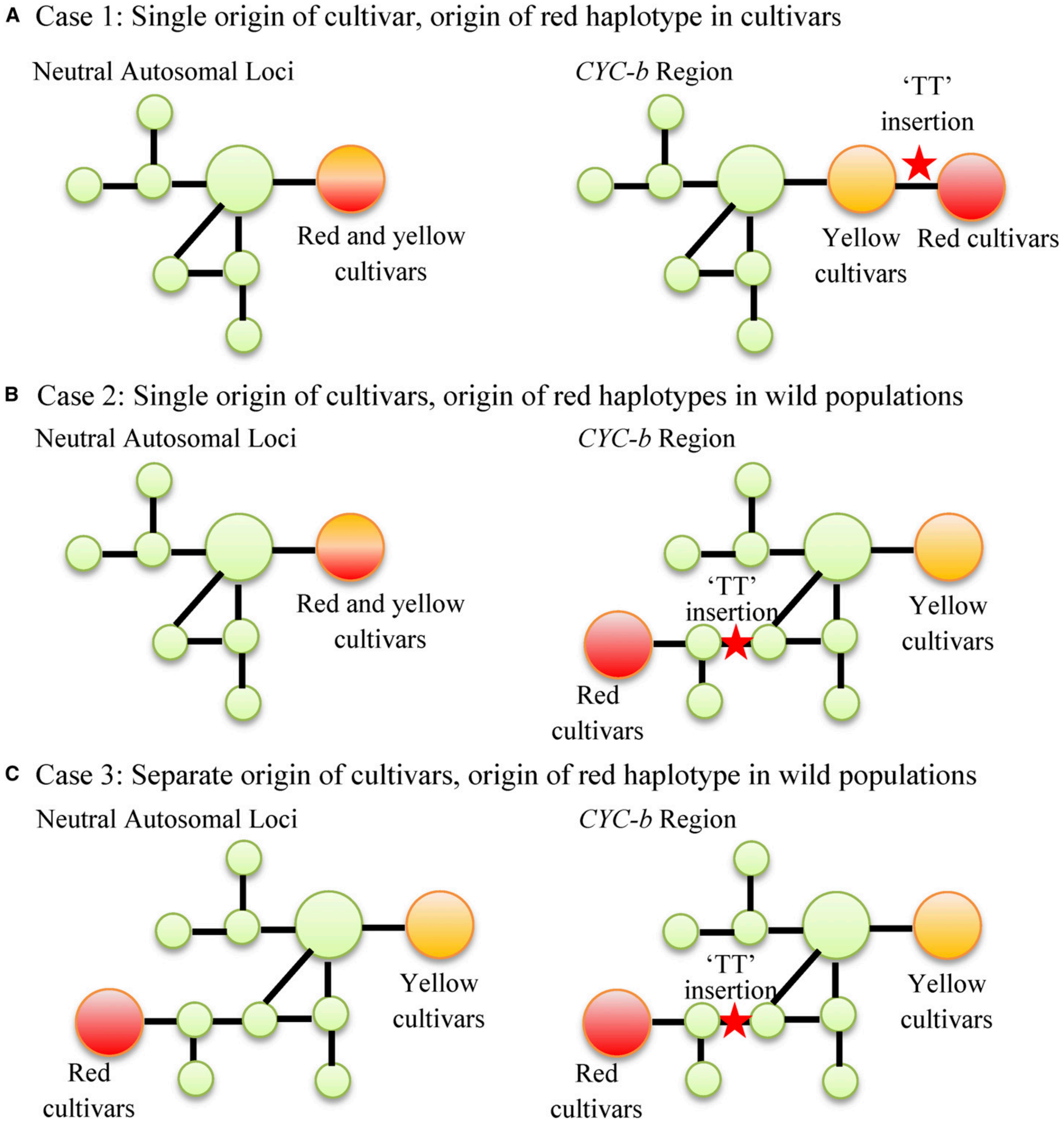

FIGURE 6 Haplotype network predictions for three possible scenarios explaining the origin of the haplotype containing the *y* allele. Each case has specific relationships predicted among haplotypes for the neutral autosomal loci and the *CYC-b* region depending on the number of origins of cultivars (single vs. separate origins of red and yellow cultivars) and the origin of the *y* cultivar haplotype (cultivar vs. wild origin). (A) In case 1, there is a single origin of cultivar haplotypes at the neutral autosomal loci, and the *y* haplotype is derived from a *Y* cultivar haplotype. This scenario would arise if the *y* allele arose in cultivation from yellow cultivars. (B) In case 2, there is also a single origin of cultivars, but the *y* mutation arose in the wild, and the *y* cultivar haplotype is derived from wild haplotypes. This scenario would arise if the *y* allele was introgressed from the wild into a cultivar genetic background. (C) In case 3, the neutral autosomal haplotype network indicates two independent origins of red and yellow cultivars, and the *y* cultivars also arose in the wild. This scenario would arise if red cultivars were domesticated independently from wild red individuals.

et al., 2012). The collection and cultivation of unimproved cultivars represents breeders' attempts to tap the genetic reservoir of naturally occurring papaya, as these populations might harbor variation for traits such as disease resistance or increased yield (Kim et al., 2002; Vázquez Calderón et al., 2014). And, as appears to be the case of the *CYC-b* locus, wild alleles may code for crop diversification traits, such as fruit color, that, while unnecessary for assisting cultivation, confer a desirable aesthetic trait favored by breeders and consumers.

This history of ongoing gene flow between cultivated and wild papaya is likely a two-way street, however, as evidenced by feral papaya in the Costa Rican landscape. Gene flow from cultivated crops to wild populations or related species is an important ecological and environmental issue, especially for transgenic crops (Fuchs and Gonsalves, 2007). Transgenic papaya has been developed to resist PRSV (papaya ringspot virus; Gonsalves, 1998), a major pathogen of the crop. Transgene movement in virus-resistant transgenic plants including squash and sugar beets has been reported in experimental field settings (Bartsch et al., 1996; Fuchs et al., 2004). If wild relatives acquire antivirus transgenes through gene flow from those virus-resistant crops, they will exhibit increased fitness when there is selective advantage for those resisting the corresponding virus (Snow and Palma, 1997; Stewart et al., 2003; Ellstrand et al., 2013). In our study, feralization of the *y* alleles showed the case of gene flow from cultivars to wild populations. If transgenic, cultivated papayas were cultivated in Mesoamerica, gene flow into wild populations would be an ecological risk that would demand consideration.

ACKNOWLEDGEMENTS

We thank Allison Miller (St. Louis University), Ken Olsen (Washington University), and members of the Olsen laboratory for their helpful comments in preparing this manuscript. We are especially grateful for the thorough and thoughtful critiques supplied by two anonymous reviewers. This work was supported by the National Science Foundation (DBI-0922545) to R.C.M. and a National Science Foundation Research Experiences for Undergraduates Site grant to Miami University (DBI-1156703). We are greatly indebted to Oscar Rocha at Kent State University for his help during the specimen collecting process in Costa Rica. We also 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.

LITERATURE CITED

Allaby, R. G., D. Q. Fuller, and T. A. Brown. 2008. The genetic expectations of a protracted model for the origins of domesticated crops. *Proceedings of the National Academy of Sciences, USA* 105: 13982–13986.

Bartsch, D., M. Schmidt, M. Pohlorf, C. Haag, and I. Schuphan. 1996. Competitiveness of transgenic sugar beet resistant to beet necrotic yellow vein virus and potential impact on wild beet populations. *Molecular Ecology* 5: 199–205.

Blas, A. L., R. Ming, Z. Y. Liu, O. J. Veatch, R. E. Paull, P. H. Moore, and Q. Y. Yu. 2010. Cloning of the papaya chromoplast-specific lycopene beta-cyclase, *CpCYC-b*, controlling fruit flesh color reveals conserved microsynteny and a recombination hot spot. *Plant Physiology* 152: 2013–2022.

Brown, J. E., J. M. Bauman, J. F. Lawrie, O. J. Rocha, and R. C. Moore. 2012. The structure of morphological and genetic diversity in natural populations of *Carica papaya* (Caricaceae) in Costa Rica. *Biotropica* 44: 179–188.

Clegg, M. T., H. F. Chen, P. L. Morrell, V. E. T. M. Ashworth, and M. de la Cruz. 2009. Tracing the geographic origins of major avocado cultivars. *Journal of Heredity* 100: 56–65.

Clement, M., D. Posada, and K. A. Crandall. 2000. TCS: A computer program to estimate gene genealogies. *Molecular Ecology* 9: 1657–1659.

Cornille, A., P. Gladieux, M. J. M. Smulders, I. Roldan-Ruiz, F. Laurens, B. Le Cam, A. Nersesyan, et al. 2012. New insight into the history of domesticated apple: Secondary contribution of the European wild apple to the genome of cultivated varieties. *PLOS Genetics* 8: e1002703.

Devitt, L. C., K. Fanning, R. G. Dietzgen, and T. A. Holton. 2010. Isolation and functional characterization of a lycopene beta-cyclase gene that controls fruit colour of papaya (*Carica papaya* L.). *Journal of Experimental Botany* 61: 33–39.

Ellstrand, N. C., P. Meirmans, J. Rong, D. Bartsch, A. Ghosh, T. J. de Jong, P. Haccou, et al. 2013. Introgression of crop alleles into wild or weedy populations. *Annual Review of Ecology, Evolution, and Systematics* 44: 325–345.

Foulongne, M., T. Pascal, F. Pfeiffer, and J. Kervella. 2003. QTLs for powdery mildew resistance in peach × *Prunus davidiana* crosses: Consistency across generations and environments. *Molecular Breeding* 12: 33–50.

Fuchs, M., E. M. Chirco, and D. Gonsalves. 2004. Movement of coat protein genes from a commercial virus-resistant transgenic squash into a wild relative. *Environmental Biosafety Research* 3: 5–16.

Fuchs, M., and D. Gonsalves. 2007. Safety of virus-resistant transgenic plants two decades after their introduction: Lessons from realistic field risk assessment studies. *Annual Review of Phytopathology* 45: 173–202.

Fumagalli, M. 2013. Assessing the effect of sequencing depth and sample size in population genetics inferences. *PLoS One* 8: e79667.

Galindo-Tovar, M. E., N. Ogata-Aguilar, and A. M. Arzate-Fernández. 2008. Some aspects of avocado (*Persea americana* Mill.) diversity and domestication in Mesoamerica. *Genetic Resources and Crop Evolution* 55: 441–450.

Giuliano, G. 2014. Plant carotenoids: Genomics meets multi-gene engineering. *Current Opinion in Plant Biology* 19: 111–117.

Gonsalves, D. 1998. Control of papaya ringspot virus in papaya: A case study. *Annual Review of Phytopathology* 36: 415–437.

Gorbachev, V. V. 2012. Effect of random sample size on the accuracy of nucleotide diversity estimation. *Genetika* 48: 880–884.

Gross, B. L., and K. M. Olsen. 2010. Genetic perspectives on crop domestication. *Trends in Plant Science* 15: 529–537.

Hall, T. A. 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program from Windows 95/98/NT. *Nucleic Acid Symposium Series* 41: 95–98.

Hufford, M. B., P. Lubinksy, T. Pyhajarvi, M. T. Devengenzo, N. C. Ellstrand, and J. Ross-Ibarra. 2013. The genomic signature of crop-wild introgression in maize. *PLOS Genetics* 9: e1003477.

Hughes, C. E., R. Govindarajulu, A. Robertson, D. L. Filer, S. A. Harris, and C. D. Bailey. 2007. Serendipitous backyard hybridization and the origin of crops. *Proceedings of the National Academy of Sciences, USA* 104: 14389–14394.

Kim, M. S., P. H. Moore, F. Zee, M. M. M. Fitch, D. L. Steiger, R. M. Manshardt, R. E. Paull, et al. 2002. Genetic diversity of *Carica papaya* as revealed by AFLP markers. *Genome* 45: 503–512.

Kovach, M. J., M. N. Calingacion, M. A. Fitzgerald, and S. R. McCouch. 2009. The origin and evolution of fragrance in rice (*Oryza sativa* L.). *Proceedings of the National Academy of Sciences, USA* 106: 14444–14449.

Kovach, M. J., and S. R. McCouch. 2008. Leveraging natural diversity: Back through the bottleneck. *Current Opinion in Plant Biology* 11: 193–200.

Librado, P., and J. Rozas. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25: 1451–1452.

Luby, J., P. Forsline, H. Aldwinckle, V. Bus, and M. Geibel. 2001. Silk road apples—Collection, evaluation, and utilization of *Malus sieversii* from Central Asia. *HortScience* 36: 225–231.

Manshardt, R. M., and F. T. P. Zee. 1994. Papaya germplasm and breeding in Hawaii. *Fruit Varieties Journal* 48: 146–152.

Miller, A. J., and B. L. Gross. 2011. From forest to field: Perennial fruit crop domestication. *American Journal of Botany* 98: 1389–1414.

Miller, A., and B. Schaal. 2005. Domestication of a Mesoamerican cultivated fruit tree, *Spondias purpurea*. *Proceedings of the National Academy of Sciences, USA* 102: 12801–12806.

Miller, A. J., and B. A. Schaal. 2006. Domestication and the distribution of genetic variation in wild and cultivated populations of the Mesoamerican fruit tree *Spondias purpurea* L. (Anacardiaceae). *Molecular Ecology* 15: 1467–1480.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York, New York, USA.

Olsen, K. M., and M. D. Purugganan. 2002. Molecular evidence on the origin and evolution of glutinous rice. *Genetics* 162: 941–950.

Quilot, B., B. H. Wu, J. Kervella, M. Genard, M. Foulongne, and K. Moreau. 2004. QTL analysis of quality traits in an advanced backcross between *Prunus persica* cultivars and the wild relative species *P. davidiana. Theoretical and Applied Genetics* 109: 884–897.

Rozen, S., and H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods in Molecular Biology* 132: 365–386.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. P. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419: 832–837.

Sinclair, E. A., and R. J. Hobbs. 2009. Sample size effects on estimates of population genetic structure: Implications for ecological restoration. *Restoration Ecology* 17: 837–844.

Snow, A. A., and P. M. Palma. 1997. Commercialization of transgenic plants: Potential ecological risks. *Bioscience* 47: 86–96.

Stewart, C. N., M. D. Halfhill, and S. I. Warwick. 2003. Transgene introgression from genetically modified crops to their wild relatives. *Nature Reviews. Genetics* 4: 806–817.

Storey, W. B. 1976. Papaya *Carica papaya* (Caricaceae). *In* N. W. Simmonds [ed.], Evolution of crop plants, 21–24. Longman Group, London, UK.

Subramanian, S. 2016. The effects of sample size on population genomic analyses– implications for the tests of neutrality. *BMC Genomics* 17: 123.

VanBuren, R., F. C. Zeng, C. X. Chen, J. S. Zhang, C. M. Wai, J. Han, R. Aryal, et al. 2015. Origin and domestication of papaya Y^h chromosome. *Genome Research* 25: 524–533.

Vázquez Calderón, M., M. J. Zavala León, F. A. Contreras Martín, F. Espadas y Gil, A. Navarrete Yabur, L. F. Sánchez Teyer, and J. M. Santamaría. 2014. New cultivars derived from crosses between commercial cultivar and a wild population of papaya rescued at its center of origin. *Le Journal de Botanique* 2014: 1–10.

Watterson, G. A. 1975. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology* 7: 256–276.

Weingartner, L. A., and R. C. Moore. 2012. Contrasting patterns of X/Y polymorphism distinguish *Carica papaya* from other sex chromosome systems. *Molecular Biology and Evolution* 29: 3909–3920.

Wu, M., and R. C. Moore. 2015. The evolutionary tempo of sex chromosome degradation in *Carica papaya. Journal of Molecular Evolution* 80: 265–277.

Yang, Z. 1996. Statistical properties of a DNA sample under the finite-sites model. *Genetics* 144: 1941–1950.

Zohary, D., and P. Spiegel-Roy. 1975. Beginnings of fruit growing in the Old World. *Science* 187: 319–327.