Domestication of a Mesoamerican cultivated fruit tree, *Spondias purpurea*

Allison Miller[†‡§] and Barbara Schaal[†]

[†]Department of Biology, Campus Box 1137, Washington University, 1 Brookings Drive, St. Louis, MO 63130; and [‡]Missouri Botanical Gardens, 4500 Shaw Boulevard, St. Louis, MO 63110

Contributed by Barbara Schaal, June 28, 2005

Contemporary patterns of genetic variation in crops reflect historical processes associated with domestication, such as the geographic origin(s) of cultivated populations. Although significant progress has been made in identifying several global centers of domestication, few studies have addressed the issue of multiple origins of cultivated plant populations from different geographic regions within a domestication center. This study investigates the domestication history of jocote (*Spondias purpurea*), a Mesoamerican cultivated fruit tree. Sequences of the chloroplast spacer *trn*G–*trn*S were obtained for cultivated and wild *S. purpurea* trees, two sympatric taxa (*Spondias mombin* var. *mombin* and *Spondias radlkoferi*), and two outgroups (*S. mombin* var. *globosa* and *Spondias testudinus*). A phylogeographic approach was used and statistically significant associations of clades and geographical location were tested with a nested clade analysis. The sequences confirm that wild populations of *S. purpurea* are the likely progenitors of cultivated jocote trees. This study provides phylogeographic evidence of multiple domestications of this Mesoamerican cultivated fruit tree. Haplotypes detected in *S. purpurea* trees form two clusters, each of which includes alleles recovered in both cultivated and wild populations from distinct geographic regions. Cultivated *S. purpurea* populations have fewer unique *trn*G–*trn*S alleles than wild populations; however, five haplotypes were absent in the wild. The presence of unique alleles in cultivation may reflect contemporary extinction of the tropical dry forests of Mesoamerica. These data indicate that some agricultural habitats may be functioning as reservoirs of genetic variation in *S. purpurea*.

crop origins | Mesoamerica | phylogeography | genetic variation

The geographic origins of most cultivated plants can be traced to several global centers of domestication (e.g., Fertile Crescent Region of the Middle East, Mesoamerica, the Andes, eastern Asia, sub-Saharan Africa; refs. 1–5). In each center, humans selected and cultivated a suite of native plants. Over time, the cultivated populations became genetically distinct from their wild progenitors as the evolutionary process of domestication proceeded (6). Recently, much attention has focused on the centers of domestication and their associated crops and putative wild ancestors as concerns mount about the lack of genetic diversity in cultivated plants (7). It has been estimated that the diversity found in cultivated populations has declined by as much as 80% over the past 100 years (7–9). How this loss occurs, either all at once or gradually over generations, is not clear. Scientific investigations focusing on the genetic resources contained in cultivated plants and their wild ancestors have documented the historical processes associated with domestication, providing new perspectives on contemporary patterns of genetic variation in cultivated populations and their wild ancestors (e.g., ref. 10).

One of the most elusive questions regarding the evolution of cultivated plants is the number of times a species was taken into cultivation within a domestication center (5, 11). In the Near East center of domestication (the "Fertile Crescent"), the wild ancestors of the crops upon which agriculture was founded are known (e.g., wheats, barley, pea, lentil, and chickpea) (12). The geographic distributions of these wild ancestors, together with biochemical and genetic data, have been used to suggest that emmer wheat, einkorn wheat, peas, chickpeas, and lentils were domesticated from wild progenitors just once or a few times in a single geographic region (12–18)

In contrast to the Near East center, crops domesticated in the Mediterranean region and other parts of the world have been derived more than once from their wild progenitors [e.g., olives (19–21), rice (22, 23), and breadfruit (24)]. Within the Mesoamerican center of domestication (Central Mexico to northwestern Costa Rica), at least 80 native species have been cultivated historically (2, 3 25–29). Some native crop species have complex evolutionary histories, and may have been domesticated multiple times within Mesoamerica [e.g., avocados (30) and one of the cultivated chili pepper species, *Capsicum frutescens* (25)]. Today, many of the native crop species of Mesoamerica are cultivated in traditional agricultural habitats, such as backyard gardens and living fences (1–3, 12, 31–33). They are grown and sold on a regional scale and have not yet undergone the intensive selection and large-scale cultivation characteristic of modern agriculture. Consequently, some Mesoamerican crop populations often resemble their wild relatives, with transitional, morphologically intermediate forms existing between cultivated populations and their progenitors (3). The native crop species of Mesoamerica provide a unique opportunity to document the domestication process in its incipient stages. In this paper, we report on the origins of one of the Mesoamerican cultivated fruit trees, *Spondias purpurea* L. (Anacardiaceae).

Study System

Spondias purpurea L. (known locally as jocote, ciruela mexicana, or hog plum) is a small (3–10 m) tree native to the tropical dry forests of Mexico and Central America (34–39). Jocotes are cultivated throughout the tropics and subtropics for their fruits, which are eaten fresh, sold in local markets, and made into jams and beverages (34, 35). The majority of trees are found in backyard gardens and living fences, although some formal cultivation exists in orchards (39). Jocotes are propagated vegetatively from large branch-size cuttings. The existence of at least 180 common names for the species suggests that jocotes have been used by many cultures, and that there is considerable variation within the species (40). Cultivated, mature fruits vary widely in color (green, yellow, orange, red, violet), size (3–5.5 cm long), texture (chalky, juicy), and taste (sweet, acidic) (39). At the time of the European colonization, jocotes were grown widely from Mexico to the northern regions of South America, as described by the first chroniclers of the region (39).

Today, wild jocote populations are found in the fragmented tropical dry forests of Mexico and Central America (36). The

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ163952–DQ164096).

[§]To whom correspondence should be sent at the present address: University of Colorado Museum, UCB 265, University of Colorado, Boulder, CO 80309. E-mail: allison.j.miller@colorado.edu.

native habitat of wild jocotes is severely restricted; it is estimated that <2% of the Mesoamerican tropical dry forests remain (41, 42). Fruits of wild jocotes are usually bright red (A.M., personal observation). They are smaller and more acidic than the cultivated fruits, with considerably less flesh surrounding the seed. Unlike cultivated populations, wild jocotes reproduce from seed and native populations are age-structured with a variety of juvenile and mature individuals present. The contrast in morphology and method of reproduction between wild and cultivated jocotes indicates that *S. purpurea* is a species that has been altered genetically during the course of domestication. During this process, humans have selected for trees that bear large, fleshy, sweet fruits, and that can be reproduced easily from cuttings.

The genus *Spondias* L. comprises ≈17 species including seven taxa in the Neotropics and ≈10 species in the Asian tropics (refs. 43–45 and D. Daly and J. D. Mitchell, personal communication). Nearly all *Spondias* species have a fibrous endocarp and leaflets with an intramarginal vein (45). Neither the relationship of *Spondias* to closely allied genera nor the relationships among the species of *Spondias* are well understood (46).

Spondias purpurea is one of three *Spondias* species native to Mexico and Central America. *Spondias mombin* L. var. *mombin* (jobo) is a widespread taxon occuring from Mexico to Paraguay (47). Native to the tropical wet forests, *S. mombin* var. *mombin* is a large tree that grows to a height of 30 m. *Spondias mombin* var. *mombin* differs from *S. purpurea* in flower and fruit color, inflorescence structure, leaflet size and number, and bark characteristics (37). In Mesoamerica, *S. mombin* var. *mombin* is cultivated occasionally for its fruits and in living fences, although it is not nearly as common as the cultivated jocotes. The third native *Spondias* species in Mesoamerica is *Spondias radlkoferi* Donn. Sm., is morphologically similar to *S. mombin* var. *mombin*, although it is distinct in fruit shape and flowering time (48). It is rarely cultivated.

In this study, we apply a phylogeographic approach based on chloroplast sequence data to investigate the geographic origins of domesticated jocotes in Mesoamerica and to examine the changes in genetic diversity associated with domestication. Our objectives are to (*i*) identify the ancestors of the cultivated jocotes, (*ii*) document the impact of domestication on the diversity of *trn*G–*trn*S sequences of *S. purpurea*, and (*iii*) to discern one (or more) geographic regions within Mesoamerica where cultivated jocotes originated.

Materials and Methods

Plant Collection. Field studies were conducted during the summers of 2000–2002 in Costa Rica, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, and Panama (Fig. 1). Ninety-six individuals of *S. purpurea* were sampled, representing 11 geographic regions (Table 1). In each region, individuals were sampled from both wild and various cultivated habitats (backyards, living fences, orchards). Multiple accessions of sympatric Mesoamerican *Spondias* species were collected [*S. mombin* var. *mombin* ($n = 25$) and *S. radlkoferi* ($n = 9$)]. Three Brazilian taxa that are morphologically similar to the Mesoamerican species were included as South American outgroups [*S. mombin* var. *mombin* ($n = 4$), *S. mombin* var. *globosa* ($n = 4$), *Spondias testudinus* ($n = 6$)] (A. Costello, Ross School, New York; J. D. Mitchell, personal communication). Leaves for DNA extraction were preserved in silica gel. Herbarium specimens (collected from each population) were deposited at the Missouri Botanical Garden and in herbaria in the country of origin (CR, USJ, ITIC, UVAL, EAP, TEFH, GUADA, CICY, YUC, MEXU, ENAG, SCZ, STRI, PMA). The precise geographic location of each population was determined by using a Garmin Etrex GPS.

Fig. 1. The Mesoamerican Center of Domestication. *Spondias* samples were collected in Mexico, Guatemala, El Salvador, Honduras, Nicaragua, Costa Rica, and Panama.

DNA Extraction, Amplification, and Analysis. Dried leaves were frozen with liquid nitrogen, mixed with powdered glass, and pulverized by using a mortar and pestle. Crushed leaves (≈200 mg) were washed with a Hepes buffer. DNA was extracted from washed leaf material by using a modified (5×) cetyltrimethylammonium bromide (CTAB) extraction (49) and then purified by using the GeneCleanII kit (Qbiogene, Irvine, CA).

Approximately 1,000 bases of the chloroplast spacer *trn*G–*trn*S were amplified by using the forward primer *trn*G 5′-GAA CGA ATC ACA CTT TTA CCA C-3′ and reverse primer *trn*S 5′-GCC GCT TTA GTC CAC TCA GC-3′ (50). The chloroplast genome is assumed to be maternally inherited in *S. purpurea*, as it is maternally inherited in most angiosperms (51, 52). Cycling conditions were 94°C (5 min); then for 30 cycles 94°C (1 min), 55°C (45 s), 72°C (1 min), and a final extension at 72°C (5 min). Two 50-μl reactions were completed for each individual, resulting in 100 μl of PCR product per sample. PCR products were purified by using the Viogene Gel Extraction kit (Viogene, Illkirch, France). Purified templates were sequenced in two directions by using the dideoxy chain termination method. Sequencing reactions were carried out by using the *trn*G forward primer and several *Spondias*-specific internal primers: *trn*G 11.11 5-CGG CAC TGA ACG AAT CAC AC-3′, *trn* GSp 5′-TTT GAC AGA TAT GGC TGG AC-3′, and *trn*S 755 5′-CGG CCT GGC CCT GGC AGT ACC-3′. A string of A's in the middle of the region complicated sequencing efforts and forced the removal of 400 bp in the final alignment. Bases were fluorescently labeled with Big Dye Terminator versions 3.1 and 1.1 sequencing reagents (Applied Biosystems). Sequences were visualized by using the Base Station DNA Fragment Analyzer and CARTOGRAPHER software (MJ Research). Clean sequences were difficult to obtain from some samples; these samples were cloned. Ligation was carried out in a 4.5-μl reaction (2.5 μl of Rapid Ligation Buffer, 1.5 μl of PCR product, 0.5 μl of pGEM T Easy Vector; Promega). Two clones per sample were sequenced in both directions.

Tests of neutrality using Tajima's *D* statistic and Fu and Li's statistic were completed by using DNASP (version 3.53, ref. 53). Gene flow within and among regions (populations) was approximated as N_m, the number of migrants per generation between populations, and was estimated by using the expression $F_{st} = 1/(1 + 2Nm)$, where N is the female effective population size and m is the female migration rate (54). Geographic distances between populations were quantified by using PASSAGE (55) and used to conduct Isolation By Distance analyses with both F_{st} and N_m values. A phylogeographic analysis was used to examine the geographic distribution of alleles (56, 57). Haplotypes (= alleles) were identified and assigned a unique letter code. A haplotype

Table 1. *S. purpurea* populations in Mexico and Central America

Population	Region	Habitat (*n*)*	*n*	Haplotypes (*n*)	Latitude	Longitude
1	Northwestern Costa Rica, Southwestern-Nicaragua	W (5), BY (3), LF (1)	9	AC (5), AA (2), S (1), Z (1)	10°30′ N	85°10′ W
2	Eastern Guatemala	W (7), O (6)	13	AC (1), R (2), V (6), W (1), X (1), Y (1), Z (1)	14°23′ N	90°35′ W
3	Western Guatemala	BY (2), LF (6)	8	AC (7), AE (1)	15°52′ N	89°28′ W
4	El Salvador	W (2), LF (3)	5	AC (2), R (1), Z (2)	13°52′ N	89°29′ W
5	Honduras	W (3), BY (4), LF (3), O (1)	11	AC (7), Q (1), V (2), Z (1)	14°05′ N	86°47′ W
6	Nicaragua	W (3), BY (2), LF (4)	9	AC (4), AD (1), R (1), Z (3)	12°10′ N	86°06′ W
7	Southern Costa Rica, Panama	BY (5), LF (8)	13	AC (11), Z (2)	08°38′ N	81°44′ W
8	Mexico (Yucatan)	BY (5)	5	AC (5)	20°49′ N	89°15′ W
9	Central Western Mexico (Northern Jalisco, Nayarit)	W (9)	9	AC (4), O (3), P (1), T (1)	20°43′ N	105°15′ W
10	Central Western Mexico (Southern Jalisco, Michoacan)	W (5), BY (3)	8	AC (3), AF (1), O (1), P (3)	19°12′ N	103°59′ W
11	Southern Mexico (Chiapas, Oaxaca)	W (5), BY (1)	6	AB (3), AC (2), Z (1)	16°39′ N	93° 58′ W

*W, wild (deciduous/semideciduous tropical dry forests); BY, cultivated in backyards (home gardens); LF, cultivated in a living fence; O, cultivated in an orchard. The number of accessions from each type of habitat is shown in parentheses.

network depicting the mutational relationships among distinct haplotypes was drawn following the principle of parsimony. A nested contingency analysis was conducted (58–60). Haplotypes were grouped into one-, two-, and three-step clades following ref. 59, and statistically significant associations between geographic localities and clades were detected by using the program GEODIS (version 2.0; ref. 61).

Results and Discussion

Chloroplast Sequence Variability. Sequences of the chloroplast *trn*G–*trn*S spacer resulted in 610 aligned bases with 16 substitutions and 13 indels that varied in size from 1 to 219 bp (Table 3, which is published as supporting information on the PNAS web site). Sequence divergence, as measured with Kimura two-parameter algorithm in PAUP* (62), ranged from 0% to 3.37% for all *Spondias* samples, and from 0% from 0.86% within *S. purpurea*. Sequences of the *trn*G–*trn*S spacer in *Spondias* conform to the expectation of neutrality by Tajima's criterion ($D = -1.32250$, $P > 0.10$) and Fu and Li's D^* ($D = -2.41966$, $P > 0.05$) and $F^* = -2.36326$, $P > 0.05$) criteria. Thirty unique haplotypes were identified from the *trn*G–*trn*S sequences. The number of individuals carrying a given haplotype varied from 1 to 60. Pairwise F_{st} values ranged from 0 to 0.66379. Tests for isolation by distance based on N_m and F_{st} were not significant (POPGENE, version 1.31, ref. 63).

Ancestors of Cultivated *S. purpurea* Based on the Distribution of *trn*G–*trn*S Haplotypes. The 30 *trn*G–*trn*S haplotypes recovered in this study fell into two distinct groups (groups 1 and 2, Fig. 2*A*). Haplotypes in the two groups differed by at least 10 mutations. The 13 haplotypes found in group 1 were carried exclusively by *S. mombin*, *S. radlkoferi*, and *S. testudinus* trees from Central and South America; none of the ≥100 *S. purpurea* phenotypes surveyed carried any of the alleles of group 1. All individuals identified as either wild or cultivated *S. purpurea* carried one group 2 haplotype (Fig. 2*A*), reflecting the close relationship of wild and cultivated *S. purpurea* trees. In addition, four of the group 2 haplotypes (haplotypes AA, AC, R, and Z) were recovered in five Mesoamerican trees identified as either *S. mombin* var. *mombin* or *S. radlkoferi*. The *S. mombin* var. *mombin* and *S. radlkoferi* trees carrying group 2 haplotypes occurred in southern Central America (southwestern Nicaragua and northwestern Costa Rica). Shared haplotypes among species can be attributed to secondary gene flow (hybridization) or incomplete lineage sorting, where branching events in gene geneaologies do not correspond to branching in population history (57). Based on the *trn*G–*trn*S sequence data, Mesoamerican *S. purpurea* trees either share a common ancestor with other *Spondias* taxa in southern Central America, are experiencing ongoing gene flow with sympatric congeners in this region, or both.

Haplotype Distribution in Cultivated and Wild Mesoamerican Jocotes. Of the 17 haplotypes detected in *S. purpurea* trees, 12 (71% of the total allele diversity) were recovered in wild populations, and nine (53% of the total allele diversity) were carried by cultivated individuals. During the course of jocote domestication, both the number of *trn*G–*trn*S alleles and the relative abundance of those alleles have changed under the influence of human selection (Fig. 3).

Four alleles (24% of total allele diversity) were found in both wild and cultivated populations. Three of the four shared alleles (R, V, and Z) were recovered sporadically in wild populations, living fences, and backyard gardens in a region spanning from southern Mexico to Panama. Allele AC was found in wild and cultivated populations throughout Mesoamerica. This mosaic-like geographic pattern of shared alleles in cultivated and wild populations is consistent with the idea that genetically distinct individuals from different geographic regions were taken into cultivation and subsequently distributed by humans.

In most cases, the variation detected in cultivated populations is a subset of the total haplotype diversity recorded for a species (10, 64–66). In *S. purpurea*, five haplotypes were found in cultivated populations but were not detected in wild populations. The presence of unique haplotypes in agricultural habitats may be the result of incomplete sampling of wild populations, or it may be the result of new alleles that have arisen in cultivation. Alternatively, it may reflect contemporary extinction of the tropical dry forests of Mexico and Central America and, consequently, the extinction of alleles carried by *S. purpurea* trees in these forests. Four of the five unique alleles were recovered from informal agricultural habitats. These data provide evidence for previous claims that traditional agricultural habitats may be acting as important reservoirs of genetic variation (67–69).

The *trn*G–*trn*S Haplotype Network. The haplotypes recovered in *S. purpurea* trees were organized into haplotype networks based on their mutational differences (Fig. 2*B*). Two most parsimonious networks were identified. Both contain three homoplasious sites and are two steps shorter than the next most parsimonious network. Analyses were conducted with both networks, one of which is shown in Figs. 2*B* and 4. The other most parsimonious network differed in the placement of the clade that included haplotypes AA, AB, S, Q, Y, and Z. In the alternative network,

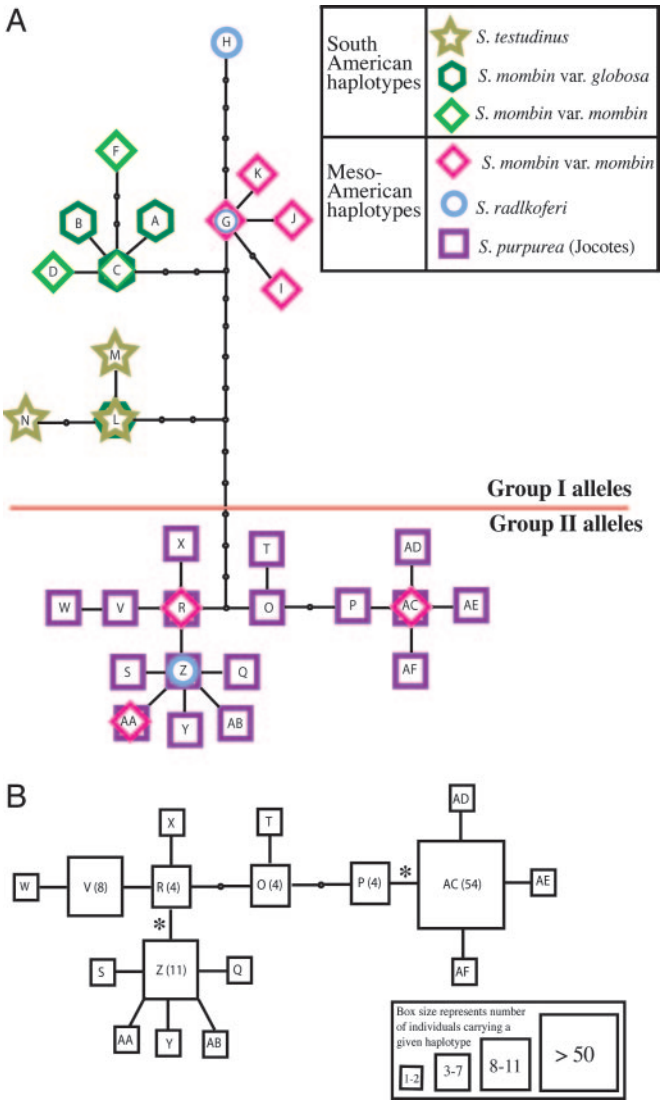

Fig. 2. Relationships among *trn*G–*trn*S alleles recovered in *Spondias* trees. (*A*) Haplotype network depicting mutational relationships among 30 *trn*G–*trn*S alleles found in *S. purpurea* (squares), *S. mombin* var. *mombin* (diamonds), *S. mombin* var. *globosa* (hexagons), *S. radlkoferi* (circles), and *S. testudinus* (stars). The 13 haplotypes found in group 1 were carried exclusively by *S. mombin*, *S. radlkoferi*, and *S. testudinus* trees from Central and South America: none of the ≥100 *S. purpurea* phenotypes surveyed carried any of the alleles of group 1. All individuals identified as either wild or cultivated *S. purpurea* carried one of the 17 haplotypes included in group 2. Group 2 alleles AA, R, and AC were recovered in *S. mombin* var. *mombin* trees in Central America as well. Allele Z was found in an *S. radlkoferi* tree. (*B*) One of two most parsimonious haplotype networks depicting mutational relationships among *trn*G–*trn*S alleles recovered in *S. purpurea* populations (group 2 alleles). The size of the box reflects the number of individuals that carried that haplotype. One gap was mapped on the tree twice, this is indicated with an asterisk.

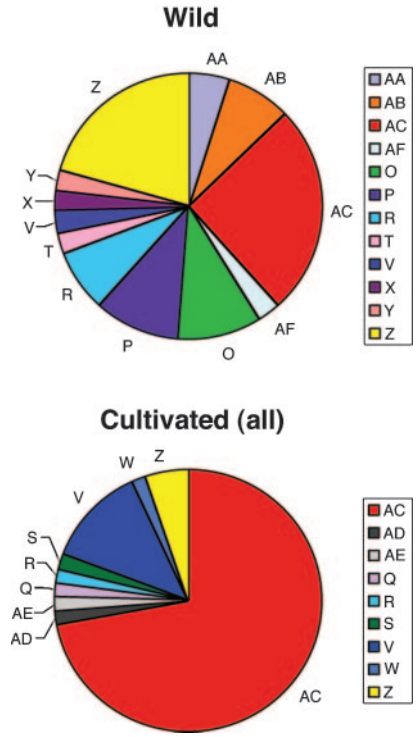

Fig. 3. *trn*G–*trn*S alleles found in wild and cultivated (including backyard trees, living fences, and orchard trees) *S. purpurea*. Alleles were more numerous and more evenly distributed in wild samples as compared with cultivated samples, of which 71% carried haplotype AC. Twelve alleles were found in wild populations, and nine alleles were recovered in cultivated trees. Eight alleles were detected exclusively in wild populations, and five alleles were found only in cultivated populations. Four alleles were found in both cultivated and wild populations (AC, R, V, and Z).

this clade was attached to haplotype O instead of haplotype R. The placement of this clade does not affect overall conclusions of this paper. The distribution of the *trn*G–*trn*S alleles (Fig. 2*B*) conform to the predictions of coalescent theory (70).

The *trn*G–*trn*S spacer had 13 indels, each of which was coded as a single binary character following the six rules described in ref. 71 (see also refs. 72 and 73) (Tables 3 and 4, which are published as supporting information on the PNAS web site). One of the gaps (gap six) was mapped on the haplotype network twice (indicated with an asterisk on network, Fig. 2*B*). Of particular interest is the region from position 169 to 229, a series of three adjacent sequences (19–21 bp in length). The string of gaps created by the lacking nucleotides in this region were labeled 4, 5, and 6 (Table 4). Alleles recovered from wild and cultivated *S. purpurea* individuals contained nucleotides in some or all of the regions corresponding to gaps 4, 5, and 6, whereas those carried by *S. mombin* and *S. radlkoferi* individuals lacked nucleotides in this region. All alleles common in *S. purpurea* contained nucleotides at 169–188 (gap 4), and all but three had an insertion in gap 6 (210–228). The three alleles lacking sequence in gap 6 (O, P, and T) were carried exclusively by *S. purpurea* trees from wild populations in the states of Jalisco, Michoacan, and Nayarit in western central Mexico. In addition to their restricted geographic distribution and their absence from cultivated *S. purpurea* trees, the status of alleles O, P, and T as putatively primitive within *S. purpurea* is further substantiated by their interior status in the *trn*G–*trn*S haplotype network (see below for detailed discussion).

Distribution of Haplotypes in *S. purpurea* Populations. ***Ancestral haplotypes.*** Coalescent theory predicts that older alleles will occupy interior positions in the haplotype network (70). In the *trn*G–*trn*S network, haplotypes R, O, and P are the most interior haplotypes (Fig. 2*B*). These haplotypes were recovered from wild *S. purpurea* populations in western Central Mexico (Jalisco, Michoacan, and Nayarit; alleles O and P) and from wild populations in Guatemala and El Salvador (allele R) (Fig. 4*A*).

***Multiple origins of cultivated* S. purpurea.** The *trn*G–*trn*S haplotype network reveals two groups of *S. purpurea* haplotypes, one centered in western Central Mexico and the other spanning from

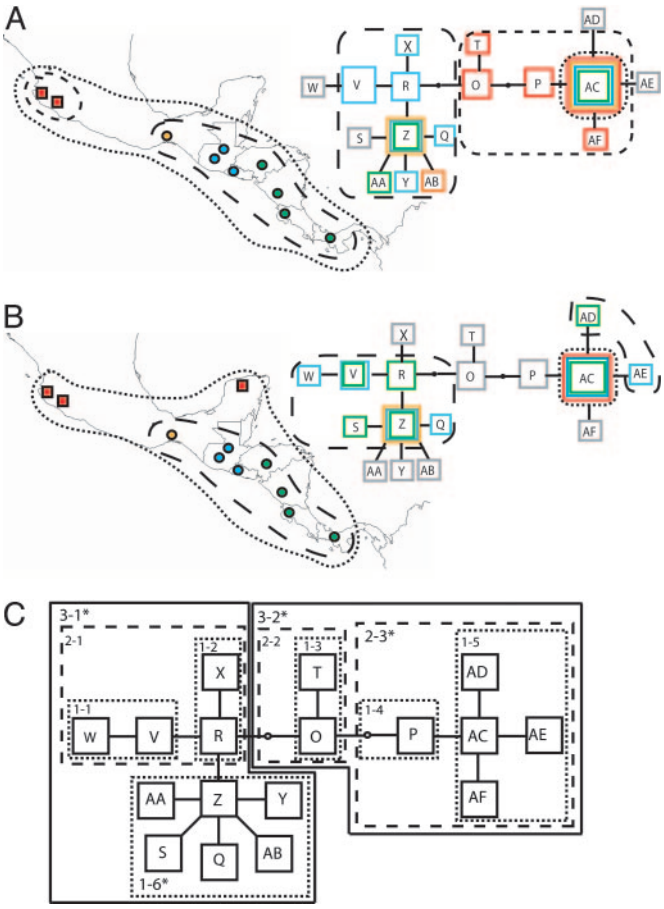

Fig. 4. Geographic distribution of *trn*G–*trn*S alleles in cultivated and wild *S. purpurea* populations. (*A*) Distribution of *trn*G–*trn*S alleles in wild populations. Alleles found in wild populations fall in two groups: southern Mexico and Central America (long dashes) and western central Mexico (short dashes). Alleles boxed with long dashes (AA, AB, R, X, V, Y, and Z) were found in wild populations from southern Mexico to Panama. Alleles boxed with the short dashes (T, O, P, AC, and AF) were found in wild populations in western central Mexico. Allele AC (boxed with small dots on the network) was found in wild populations throughout Mesoamerica as well. Alleles boxed in gray were not recovered in any wild populations. (*B*) Distribution of *trn*G–*trn*S alleles in cultivated populations. Alleles found in wild populations from southern Mexico to Panama (boxed in long dashes) were recovered in cultivated populations in the same region. Allele AC, which was most closely related to alleles found exclusively in wild populations of western central Mexico, was distributed in cultivated populations throughout Mesoamerica. Alleles boxed in gray were not recovered in any cultivated populations. (*C*) Nested clade analysis of *trn*G–*trn*S alleles found in *S. purpurea* trees. Individual alleles (0-step clades) were grouped into 1-step, 2-step, and 3-step clades. A nested contingency analysis rejected the null hypothesis (no association between genealogy and geographic locality) for four clades: 1-6, 2-3, 3-1, and 3-2.

southern Mexico through Central America (Fig. 4 *A* and *B*). The first group comprises alleles recovered from wild and cultivated populations in southern Mexico and Central America (alleles Q, R, S, V, W, X, Y, Z, AA, and AB). The second group includes seven alleles, four of which were recovered exclusively in the wild populations of western Central Mexico (O, P, T, and AF). AC was the most common haplotype; it was detected in wild populations of western Central Mexico, as well as in cultivated and wild populations throughout Mexico and Central America. Alleles AD and AE were recovered from a backyard tree in Nicaragua and living fence in Guatemala, respectively.

Geographical Structuring. A nested clade analysis (NCA) was used to test for statistically significant associations between clades and geographical locations (refs. 59 and 60; but see ref. 74) (Fig. 4*C*). The NCA rejected the null hypothesis of no association between geographical location and clades for clades 1-6, 2-3, 3-1, and 3-2 (Fig. 4*C* and Table 2). Clade 1-6 included alleles in cultivated populations from El Salvador, Costa Rica, and Panama, and wild populations from southern Mexico (Chiapas) to northern Costa Rica. The null hypothesis was rejected for the next level of nesting as well (clade 3-1), which includes alleles found in cultivated and wild populations in southern Mexico and Central America. In addition, statistically significant results were obtained for clades 2-3 and 3-2, which include alleles found exclusively in wild populations in western Central Mexico (alleles AF, T, O, and P), a widespread allele found in cultivated and wild populations throughout the region (AC), and two singletons from Central America (AD and AE). The NCA provides statistical support for two distinct groups of *S. purpurea* haplotypes, corroborating the inference that *S. purpurea* was domesticated more than once in Mesoamerica.

Table 2. Nested clade analysis for *trn*G–*trn*S alleles recovered in cultivated and wild jocote trees (*S. purpurea*)

Clade*	D_c†	D_n	D_n	I/T‡	I/T D_c	I/TD_n
1-6					361.13	−128.98
	XV	0.0001^S	710.33^L	Tip		
2-3					−608.53^S	762.93^L
	1-4	80.38^S	1480.84^L	Int.		
	1-5	688.91^S	717.92^S	Tip		
3-1					−259.15^S	−216.39^S
	2-1	168.10^S	210.72^S	Int.		
	1-6	427.11^L	427.11^L	Tip		
3-2					−696.05^S	667.46^L
	2-2	68.52^S	1464.13^L	Int.		
	2-3	764.57^S	796.67^S	Tip		
Total cladogram						
	3-1	346.67^S	416.49^S	Tip		
	3-2	844.06^L	825.73^L	Tip		

*Only clades with significant effects and containing both tips and interiors are shown. Clades marked with an asterisk show significant geographical structuring ($P < 0.05$).

†Clade distances and nested clade distances that are significant are marked with an S, (significantly small) or L (significantly large).

‡Interior (Int.) or tip (Tip) status of clade.

Conclusions

This study provides phylogeographic evidence of multiple domestications of a cultivated fruit tree within the Mesoamerican center of domestication. Sequences of the chloroplast spacer *trn*G–*trn*S confirm that wild populations of the Mesoamerican tree *S. purpurea* are the likely progenitors of cultivated jocote trees. Our data reveal that *trn*G–*trn*S haplotypes detected in *S. purpurea* trees form two clusters, each of which includes haplotypes recovered in both cultivated and wild populations. One cluster spans the region from southern Mexico through Central America; the second cluster includes haplotypes from western Central Mexico, plus one widespread allele. A nested clade analysis detected significant associations between the mutational relationships and geographic localities of the alleles. Previous studies have suggested that various Mesoamerican species may have been taken into cultivation more than once within this center of domestication. Our study provides genetic evidence for this phenomenon, highlighting the importance of multiple domestications in the evolutionary history of Mesoamerican crops.

As a human-mediated evolutionary process, domestication impacts contemporary patterns of genetic variation in cultivated populations. At the *trn*G–*trn*S locus, >50% of the total genetic variation sampled was recovered in cultivated trees. The elevated

levels of genetic variation in cultivated populations are consistent with vegetative propagation, multiple domestications, and ongoing distribution of domesticated alleles. Notably, five of the 17 haplotypes detected in cultivated *S. purpurea* trees (≈29% of the total variation at this locus in *S. purpurea*) were not recovered in wild populations. The presence of unique haplotypes in informal agricultural habitats (gardens, fences) provides support for traditional agriculture as an important reservoir of genetic variability in cultivated species when native populations of the wild ancestor are declining.

We thank members of the B.S. laboratory, Toby Kellogg, Gayle Fritz, Jason Knouft, and members of the Ranter laboratory (University of Colorado, Boulder) for helpful comments. Yu-Chung Chiang provided valuable assistance with analysis. Leaf material of Brazilian taxa was generously provided by A. Costello (Ross School, New York). We thank field collaborators R. Aguilar, A. Anzueto, G. Borjas, G. Carnevali, J. Castilla Canales, L. Chavez V., I. Diaz, A.-C. Gomez, A. Herrera, M. Diaz, A. MacVean, C. MacVean, M. Merello, A. Molina R., P. Moreño, A. Muehlenbachs, M. Olson, E. O'Mahoney-Cubbison, A. Paschke, E. Pimienta Barrios, B. Ramirez Hernandez, R. Rueda, R. Ruenes, N. Ventura, and B. Wong. This work was supported by National Science Foundation Grant PD 0105134, the Organization for Tropical Studies, the Botanical Society of America, the Andrew W. Mellon Foundation support of Missouri Botanical Garden graduate students, Washington University Division of Biology and Biomedical Sciences Graduate Fellowship, and a Danforth Foundation Grant to support undergraduates in plant biology research at Washington University.

1. Harlan, J. R. (1975) in *Crops and Man*, ed. Harlan, J. R. (Crop Sci. Soc. Am., Madison, WI), pp. 63–81.
2. MacNeish, R. S. (1992) in *The Origins of Agriculture and Settled Life* (Univ. of Oklahoma Press, Norman), p. 123.
3. Vavilov, N. I. (1987) in *Proiskhozhdenie i Geografia Kul'turnykh Rastenii* (Nauka, Leningrad, U.S.S.R.); trans. Löve, D. (1992) *Origin and Geography of Cultivated Plants* (Cambridge Univ. Press, Cambridge, U.K.), pp. 207–238.
4. Smith, B. D. (1998) *The Emergence of Agriculture* (Freeman, New York).
5. Diamond, J. (2002) *Nature* **418,** 700–707.
6. Ladizinsky, G. (1998) in *Plant Evolution Under Domestication* (Kluwer, Dordrecht, The Netherlands), pp. 4–9.
7. Tanksley, S. D. & McCouch, S. L. (1997) *Science* **277,** 1063–1066.
8. Nabhan, G. (1992) in *Chilies to Chocolate: Foods the Americas Gave the World*, eds. Foster, N. & Cordell, L. (Univ. of Arizona Press, Tucson), pp. 143–161.
9. Lanaud, C. (1999) *Plant Genetic Resources Newslett.* **119,** 26–31.
10. Olsen, K. M. & Schaal, B. A. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 5586–5591.
11. Blumler, M. A. (1992) *Econ. Bot.* **46,** 98–111.
12. Zohary, D. (1996) in *The Origins and Spread of Agriculture and Pastoralism in Eurasia*, ed. Harris, D. R. (University College London Press, London), pp. 142–158.
13. Salamini, F., Ozkan, H., Brandolini, A., Schäfer-Pregl, R. & Martin, W. (2002) *Nat. Rev. Genet.* **3,** 429–441.
14. Zohary, D. & Hopf, M. (2000) *Domestication of Plants in the Old World* (Oxford Univ. Press, Oxford).
15. Lev-Yadun, S., Gopher, A. & Abbo, S. (2000) *Science* **288,** 1602–1603.
16. Özkan, H., Brandolini, A., Schäfer-Pregl, R. & Salamini, F. (2002) *Mol. Biol. Evol.* **19,** 1797–1801.
17. Heun, M., Schäfer-Pregl, R., Klaawan, D., Castagna, R., Accerbi, M., Borhi, B. & Salamini, F. (1997) *Science* **278,** 1312–1314.
18. Abbo, S., Berger, J. & Turner, N. C. (2003) *Funct. Plant Biol.* **30,** 1081–1087.
19. Besnard, G., Baradat, P. & Bervillé, A. (2000) *C. R. Acad. Sci.* **323,** 173–181.
20. Besnard, G., Baradat, P. & Bervillé, A. (2001) *Theor. Appl. Genet.* **102,** 251–258.
21. Besnard, G., Khadari, B., Baradat, P. & Bervillé, A. (2002) *Theor. Appl. Genet.* **104,** 1353–1361.
22. Cheng, C., Motohashi, R., Tscuchimoto, S., Fukuta, Y., Ohtsubo, H. & Ohtsubo, E. (2003) *Mol. Biol. Evol.* **20,** 67–75.
23. Yamanaka, S., Nakamura, I., Nakai, H. & Sato, Y.-I. (2003) *Genet. Res. Crop Evol.* **50**, 529–538.
24. Zerega, N. J. C., Ragone, D. & Motley, T. J. (2004) *Am. J. Bot.* **91,** 760–766.
25. Hernández-Verdugo, S., D'Avila Aranda, P. & Oyama, K. (1999) *Bol. Soc. Bot. Méx.* **64,** 65–84.
26. Kaplan, L. & Lynch, T. (1999) *Econ. Bot.* **53,** 261–272.
27. Mapes, C., Basturo, F., Caballero, J. & Bye, R. (1998) *Bol. Soc. Bot. Mex.* **62,** 91–107.
28. Wang, R.-L., Stec, A., Hey, J., Lukens, L. & Doebley, J. (1999) *Nature* **398,** 236–239.
29. Sanjur, O. I., Piperno, D. R., Andres, T. C. & Wessel-Beaver, L. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 535–540.
30. Knight, R. J., Jr. (2002) in *The Avocado*, eds. Whiley, A. W., Schaffer, B. & Wolstenholme, B. N. (CABI, Oxfordshire, U.K.), pp. 1–2.
31. Casas, A. & Barbera, G. (2002) in *Cacti: Biology and Uses*, ed. Nobel, P. S. (Univ. of California Press, Berkeley), pp. 143–162.
32. Gepts, P. (1998) *Hort. Sci.* **33,** 1124–1130.
33. Tuxill, J. (1999) *Nature's Cornucopia: Our Stake in Plant Diversity* (Worldwatch, Washington, DC), paper 148.
34. Avitia-García, E. (1997) *Hort. Mex.* **5,** 282–288.
35. Baraona Cockrell, M. (2000) *Jocote, anona, cas: Tres Frutas Campesinas de América* (Editorial Universidad Nacional, Heredia, Costa Rica).
36. Mandujano, S., Gallina, S. & Bullock, S. H. (1994) *Rev. Biol. Trop.* **42,** 107–114.
37. Mitchell, J. D. (2000) in *Flora de Nicaragua*, eds. Stevens, W. D., Ulloa Ulloa, C., Pool, A. & Montiel, O.-M. (Missouri Botanical Gardens Press, St. Louis), pp. 7–9.
38. Pimienta-Barrios, E. & Ramírez-Hernández, B. C. (2003) *Econ. Bot.* **57,** 481–490.
39. Cuevas, J. A. (1994) in *Neglected Crops: 1492 from a Different Perspective*, eds. Bermejo, J. E. H. & León, J. (Food Agric. Org. U.N., Rome), pp. 111–115.
40. Morton, J. F. (1987) *Fruits of Warm Climates*, ed. Dowling, C. F., Jr. (Creative Resources Systems, Miami).
41. Mooney, H. A., Bullock, S. H. & Medina, E. (1995) in *Seasonally Tropical Dry Forests*, eds. Bullock, S. H., Mooney, H. A. & Medina, A. (Cambridge Univ. Press, Cambridge, U.K.), pp. 1–22.
42. Janzen, D. H. (1988) in *Biodiversity*, ed. Wilson, E. O. (Natl. Acad. Press, Washington, DC), pp. 130–137.
43. Airy Shaw, H. K. & Forman, L. L. (1967) *Kew. Bull.* **21,** 1–21.
44. Ding Hou (1978) *Flora Malesiana Ser. I* **8,** 395–548.
45. Kostermans, A. J. G. H. (1991) *Kedondong, Ambarella, and Amra: The Spondiadeae (Anacardiaceae) in Asia and the Pacific Area* (Bogor, Indonesia).
46. Pell, S. (2004) Ph.D. dissertation (Louisiana State Univ., Baton Rouge).
47. Brücher, H. (1989) in *Useful Plants of Neotropical Origin and their Wild Relatives* (Springer, Berlin), pp. 217–218.
48. Croat, T. (1974) *Ann. Mo. Bot. Gard.* **61,** 483–490.
49. Doyle, J. J. & Doyle, J. L. (1987) *Phytochem. Bull.* **19,** 11–15.
50. Hamilton, M. B. (1999) *Mol. Ecol.* **8,** 521–523.
51. Reboud, X. & Zeyl, X. (1994) *Heredity* **72,** 132–140.
52. Röhr, H., Kües, U. & Stahl, U. (1998) *Prog. Bot.* **60,** 39–87.
53. Rozas, J. & Rozas, R. (1995) *Comput. Appl. Biosci.* **11,** 621–625.
54. Slatkin, M. (1993) *Evolution (Lawrence, Kans.)* **43,** 1349–1368.
55. Rosenberg, M. S. (2001) PASSAGE: *Pattern Analysis, Spatial Statistics and Geographic Exegesis* (Dept. of Biology, Arizona State Univ., Tempe), version 1.0.
56. Schaal, B. A., Hayworth, D. A., Olsen, K. M., Rauscher, J. T. & Smith, W. A. (1998) *Mol. Ecol.* **7,** 465–474.
57. Avise, J. (2000) in *Phylogeography: The History and Formation of Species* (Harvard Univ. Press, Cambridge, MA), p. 11.
58. Templeton, A. R., Boerwinkle, E. & Sing, C. F. (1987) *Genetics* **117,** 343–351.
59. Templeton, A. R. & Sing, C. F. (1993) *Genetics* **134,** 659–669.
60. Templeton, A. R. (1998) *Mol. Ecol.* **7,** 381–397.
61. Posada, D., Crandall, K. A. & Templeton, A. R. (2000) *Mol. Ecol.* **9,** 487–488.
62. Swofford, D. L. (2002) PAUP*: *Phylogenetic Analysis Using Parsimony (*and Other Methods)* (Sinauer, Sunderland, MA).
63. Yeh, F. C., Yang, R.-C. & Boyle, T. (1999) POPGENE (Molecular Biology and Biotechnology Centre, University of Alberta, Edmonton), version 1.31.
64. Eyre-Walker, A., Gaut, R. L., Hilton, H., Feldman, D. L. & Gaut, B. S. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 4441–4446.
65. Saunders, J. A., Pedroni, M. J., Penrose, L. D. J. & Fish, A. J. (2001) *Crop Sci.* **41,** 1596–1601.
66. Anthony, F., Combes, M. C., Astorga, C., Bertrand, B., Graziosi, G. & Lashermes, P. (2002) *Theor. Appl. Genet.* **104,** 894–900.
67. Brown, A. H. D. (2000) in *Genes in the Field*, ed. Brush, S. B. (Lewis, Boca Raton, FL), pp. 29–48.
68. Altieri, M., Anderson, M. K. & Merrick, L. C. (1987) *Conserv. Biol.* **1,** 49–60.
69. Kumar, B. M. & Nair, P. K. R. (2004) *Agrofor. Syst.* **61,** 135–152.
70. Posada, D. & Crandall, K. A. (2001) *Trends. Ecol. Evol.* **16,** 37–45.
71. Simmons, M. P. & Ochoterena, H. (2000) *Syst. Biol.* **49,** 369–381.
72. Kelchner, S. A. & Clark, L. G. (1997) *Mol. Phylogenet. Evol.* **8,** 385–397.
73. Graham, S., Reeves, P. A., Burns, A. C. E. & Olmstead, R. G. (2000) *Int. J. Plant Sci.* **161,** S83–S96.
74. Knowles, L. L. & Maddison, W. P. (2002) *Mol. Ecol.* **11,** 2623–2635.