Evidence on the origin of cassava: Phylogeography of *Manihot esculenta*

KENNETH M. OLSEN[†] AND BARBARA A. SCHAAL

Department of Biology, Washington University, St. Louis, MO 63130

ABSTRACT Cassava (*Manihot esculenta* subsp. *esculenta*) is a staple crop with great economic importance worldwide, yet its evolutionary and geographical origins have remained unresolved and controversial. We have investigated this crop's domestication in a phylogeographic study based on the single-copy nuclear gene glyceraldehyde 3-phosphate dehydrogenase (*G3pdh*). The *G3pdh* locus provides high levels of noncoding sequence variation in cassava and its wild relatives, with 28 haplotypes identified among 212 individuals (424 alleles) examined. These data represent one of the first uses of a single-copy nuclear gene in a plant phylogeographic study and yield several important insights into cassava's evolutionary origin: (*i*) cassava was likely domesticated from wild *M. esculenta* populations along the southern border of the Amazon basin; (*ii*) the crop does not seem to be derived from several progenitor species, as previously proposed; and (*iii*) cassava does not share haplotypes with *Manihot pruinosa*, a closely related, potentially hybridizing species. These findings provide the clearest picture to date on cassava's origin. When considered in a genealogical context, relationships among the *G3pdh* haplotypes are incongruent with taxonomic boundaries, both within *M. esculenta* and at the interspecific level; this incongruence is probably a result of lineage sorting among these recently diverged taxa. Although phylogeographic studies in animals have provided many new evolutionary insights, application of phylogeography in plants has been hampered by difficulty in obtaining phylogenetically informative intraspecific variation. This study demonstrates that single-copy nuclear genes can provide a useful source of informative variation in plants.

Understanding the process of population divergence is fundamental to the study of evolutionary diversification. This process is inherently phylogenetic, such that the present population structure of a species reflects not only current patterns of genetic exchange but also the history of gene flow and isolation among population lineages. In the past decade, workers have begun to employ phylogenetically informative data (most often allele genealogies derived from DNA sequences) for investigating population divergence (1–3). This approach was first described by Avise *et al.* (1), who termed it phylogeography, and it has since been hailed widely as the conceptual bridge linking population-level processes to macroevolutionary phylogenetic relationships (2–4).

Despite a recent explosion in phylogeography studies involving animal species (3), analogous studies in plants remain scarce, primarily because of difficulties in detecting phylogenetically informative intraspecific genetic variation (5). Most attempts to detect such variation in plants have relied on the chloroplast genome, with success varying widely among taxa (e.g., refs. 5–7). The plant mitochondrial genome, with few exceptions (e.g., ref. 8), has not yielded useful amounts population-level variation. An alternative source of variation, the noncoding regions of "single-copy" (low copy number) nuclear genes, has not been extensively explored in plants (5, 9–11), although this genome can potentially provide multiple, unlinked allele genealogies at the intraspecific level (12–14).

This study reports on the use of a single-copy nuclear gene to examine the phylogeography of a plant species, *Manihot esculenta* Crantz (Euphorbiaceae), which includes the important tropical subsistence crop cassava and its wild relatives. A portion of the gene encoding glyceraldehyde 3-phosphate dehydrogenase (*G3pdh*) provides high levels of sequence variation in *Manihot*. The *G3pdh* data are used here to investigate the evolutionary origins of cassava, specifically the crop's geographical origin within the range of its wild relatives, and the potential role that interspecific hybridization may have played in the crop's domestication.

Study System. Cassava (*M. esculenta* subsp. *esculenta*) is a staple root crop for over 500 million people living throughout the tropics (15). It is the primary source of carbohydrates in sub-Saharan Africa (16) and ranks sixth among crops in global production (17). Despite its immense importance in the developing world, cassava has historically received less attention by researchers than have temperate crops, earning it the status of an "orphan crop."

One of the most fundamental questions about cassava that remains unresolved concerns its evolutionary origin. The genus *Manihot* (comprising 98 species; ref. 18) is distributed across much of the Neotropics, and the identity of cassava's closest wild relatives within the genus has been a source of widespread speculation (18–21). Most traditional domestication hypotheses have envisioned the crop to be a "compilospecies" derived from one or more species complexes, either in Mexico and Central America (18, 22) or throughout the Neotropics (21, 23, 24). More recently, wild populations of *M. esculenta* that are likely to be the crop's direct progenitors have been identified in South America (19, 25). However, the evolutionary relationship between cassava and its conspecific wild relatives is only beginning to be examined (19, 26, 27), and there is continued speculation that the crop's origins may extend beyond *M. esculenta* to hybridizing *Manihot* species (26, 27).

Wild populations of *M. esculenta* occur primarily in west central Brazil and eastern Peru (19, 28). All wild populations of this species are referred to here as *M. esculenta* subsp. *flabellifolia* (Pohl) Ciferri (see also ref. 27). This wild subspecies is found in forest patches in the transition zone between the *cerrado* (savanna scrub) vegetation of the Brazilian shield plateau and the lowland rainforest of the Amazon basin, where it grows as a clambering understory shrub or treelet. In Brazil, populations occur along the southern and eastern borders of the Amazon basin, in the states of Tocantins, Goiás, Mato Grosso, Rondônia, and Acre. Although cassava is interfertile with subspecies *flabellifolia* (27) and is cultivated within its range, the population structure of the wild subspecies is not thought to reflect introgression from the crop

---

Evolution: Olsen and Schaal

Proc. Natl. Acad. Sci. USA 96 (1999)   5587

after domestication. Subspecies *flabellifolia* occurs in forested areas where cassava does not grow. In addition, cassava does not survive well in abandoned fields or as an escape from cultivation (refs. 19 and 22; K.M.O., unpublished observation). Finally, cassava is propagated almost exclusively by stem cuttings, minimizing unintentional spread of the crop by humans.

To examine the possibility that cassava's origins extend beyond subspecies *flabellifolia*, we have included in our study *Manihot pruinosa* Pohl, a very closely related, potentially hybridizing species. *M. esculenta* and *M. pruinosa* are very similar, both morphologically (28) and with respect to the internal transcribed spacer region of nuclear ribosomal DNA (B.A.S., unpublished data). Based on this close relationship, *M. pruinosa* has been proposed to fall within cassava's "secondary gene pool" of potentially interfertile species (28); it is the only such species to occur in sympatry with *M. esculenta*. *M. pruinosa* grows as a shrub in the *cerrado* southeast of the Amazon basin, in the Brazilian states of Tocantins, Goiás, and Mato Grosso. Although *M. pruinosa* and *M. esculenta* occur in different habitats, the patchy nature of the *cerrado*–forest transition zone permits the two species to grow in close proximity, providing opportunity for interbreeding.

The goal of the present study is to use DNA sequence variation from the *G3pdh* locus to investigate cassava's evolutionary and geographical origins. (*i*) Can cassava's geographical origin of domestication be localized within the range of its conspecific wild relative? (*ii*) Do patterns of genetic variation in the crop vs. the wild *M. esculenta* subspecies indicate that the crop is derived solely from this subspecies? (*iii*) Is there any evidence that cassava's origin extends beyond *M. esculenta* to the closely related, potentially hybridizing species *M. pruinosa*?

MATERIALS AND METHODS

Populations of wild taxa were sampled in November–December of 1996 and 1997. Two transects spanning the range of subspecies *flabellifolia* were made, one along the eastern border of the Amazon basin and one along the southern border (Fig. 1). These transects include an area of sympatry with *M. pruinosa*. Undisturbed populations of these species usually consist of fewer than 15 individuals, and up to 10 per location were sampled. Young leaf tissue from each sampled plant was dried in silica gel for subsequent DNA extraction. A total of 157 individuals of subspecies *flabellifolia*, representing 27 populations, and 35 individuals of *M. pruinosa*, representing 6 populations, were used in analyses. Voucher herbarium specimens from each population are housed at the Missouri Botanical Garden and at the Centro Nacional de Pesquisa de Recursos Genéticos e Biotecnologia in Brasília, Brazil. To represent the diversity of cassava, 20 cultivars were sampled from the cassava "world core collection" maintained by the Centro Internacional de Agricultura Tropical in Cali, Colombia.

DNA was extracted from dried leaves by using a cetyltrimethylammonium bromide protocol (29). PCR amplification of the *G3pdh* region was performed with primers designed by Strand *et al.* (ref. 9; GPDX7F and GPDX9R, forward and reverse, respectively). These primers were designed from conserved regions identified in published *G3pdh* sequences of *Arabidopsis thaliana* and *Ranunculus acris* (A. Strand, personal communication). There were three 50-μl reactions carried out per individual, and each reaction contained 10 mM Tris·HCl (pH 9.0), 50 mM KCl, 2.5 mM MgCl$_2$, 0.1% Triton X-100, 2 units *Taq* Polymerase (Promega), 200 μM each dNTP, 0.2 μM each primer, and ≈10–20 ng genomic DNA. Cycling conditions were 95°C (2 min); then 30 cycles of 95°C (1 min), 60°C (1 min), 72°C (2 min); and finally 72°C (5 min). The three PCRs were pooled and purified by using a Geneclean II DNA purification kit (Bio 101).

DNA sequencing was performed by using fmol (Promega) direct cycle sequencing with a [^{35}S]α-dATP label, followed by electrophoresis on a 6% Long Ranger acrylamide gel (FMC), blotting and drying onto paper, and exposure to x-ray film. For most individuals, the two haplotypes (alleles) of the locus were sequenced together; in some cases the alleles were separated by size before purification and sequencing (see below). In addition to the external primers, two internal reverse primers were used in sequencing: GPD9R2, CTT GAT TTC CTC ATA TGT TGC C (reverse complement of bases 610–631), and GPD9R4, TCC CTT AAG CTT ACC CTC AG (reverse complement of bases 752–771; Fig. 2). DNA sequences were aligned by eye with the MACCLADE program (30). Sequence variation was tested for deviations from neutrality by using Tajima's (31) *D* statistic (treating all individuals in the study system as a single population), and by and Fu and Li's (32) *D** and *F** statistics for an unrooted tree.

Species of *Manihot* show disomic inheritance (33), such that individuals can be either homozygous or heterozygous at the *G3pdh* locus. For heterozygotes, the sequence of both haplotypes is shown on a single sequencing gel, resulting in double bands at sites that differ between the haplotypes. Scoring of most heterozygotes was accomplished by reading double-banded variable sites directly from the autoradiogram. The identity of the two haplotypes within the heterozygote was then inferred by using "haplotype subtraction" (34), whereby the heterozygote sequence is compared with pairs of homozygotes until the observed combination of double-banded sites can be accounted for. Alternatively, if the two haplotypes differed in length by 20 bp or more, they were separated on a 1.5% agarose gel and sequenced individually.
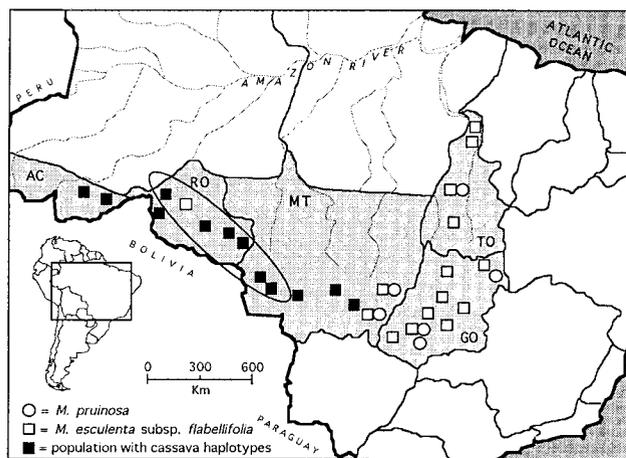

FIG. 1. Sampling locations for populations of *M. esculenta* subsp. *flabellifolia* (squares) and *M. pruinosa* (circles) in Brazilian states surrounding the Amazon basin. Black squares indicate populations of subspecies *flabellifolia* found to share one or more haplotypes with cassava (*M. esculenta* subsp. *esculenta*) accessions. Populations within the oval together contain all of the shared cassava haplotypes (see *Results and Discussion*). TO, Tocantins; GO, Goiás; MT, Mato Grosso; RO, Rondônia; AC, Acre.

FIG. 2. Structure of the *G3pdh* region in *Manihot*. Numbers 1–4 indicate the positions of primers: GPDX7F (1), GPD9R2 (2), GPD9R4 (3), and GPDX9R (4). Uppercase and lowercase letters designate exons and introns, respectively, as follows: partial exon A (52 bp), intron a (84 bp), exon B (98 bp), intron b (265 bp), exon C (143 bp), intron c (101 bp), exon D (84 bp), intron d (101 bp), and partial exon E (34 bp). The cross-hatched region within intron b indicates a variable minisatellite.

Haplotypes were ordered into a maximum parsimony gene tree, which could be constructed by hand because of the low level of homoplasy in the data (see below). The topology of this gene tree was confirmed by performing a "branch and bound" search on the data set with PAUP 3.1 (35).

RESULTS AND DISCUSSION

Structure and Variation of the *G3pdh* Region. The structure of the *G3pdh* region in *Manihot* is shown in Fig. 2. This region spans 4 introns and 3 exons, plus parts of two flanking exons, and has a maximum length of 962 bp. All four of the introns begin with bases GT and end with AG, a motif typical of nuclear introns (36). Exon portions encode 137 amino acids, as is true for the two other dicot species from which end primers were designed (see above). The *Manihot* amino acid sequence differs by less than 5.5% from these other species, which, by comparison, differ from each other by 5.8%.

The *G3pdh* region contains a total of 64 polymorphic sites within the study system (Fig. 3; Table 1). Only one polymorphism encodes an amino acid substitution (base 621, glutamate/

Table 1. Variation in the *G3pdh* locus

Region	Mutation	No.
Exon		
	Nonsynonymous	1
	Synonymous	9
Intron		
	Substitutions	47
	Indels	7
	1 bp	(2)
	3 bp	(3)
	19 bp	(1)
	25 bp*	(1)

*Minisatellite region with 1–5 repeats.

glycine). Of the variable sites within introns, seven are indels, one of which is a minisatellite region with up to five repeats of a 25-bp motif (Table 1). This minisatellite variation was not used in defining haplotypes, as the repeat number was homoplasious relative to the rest of the sequence variation (Fig. 3). Overall,
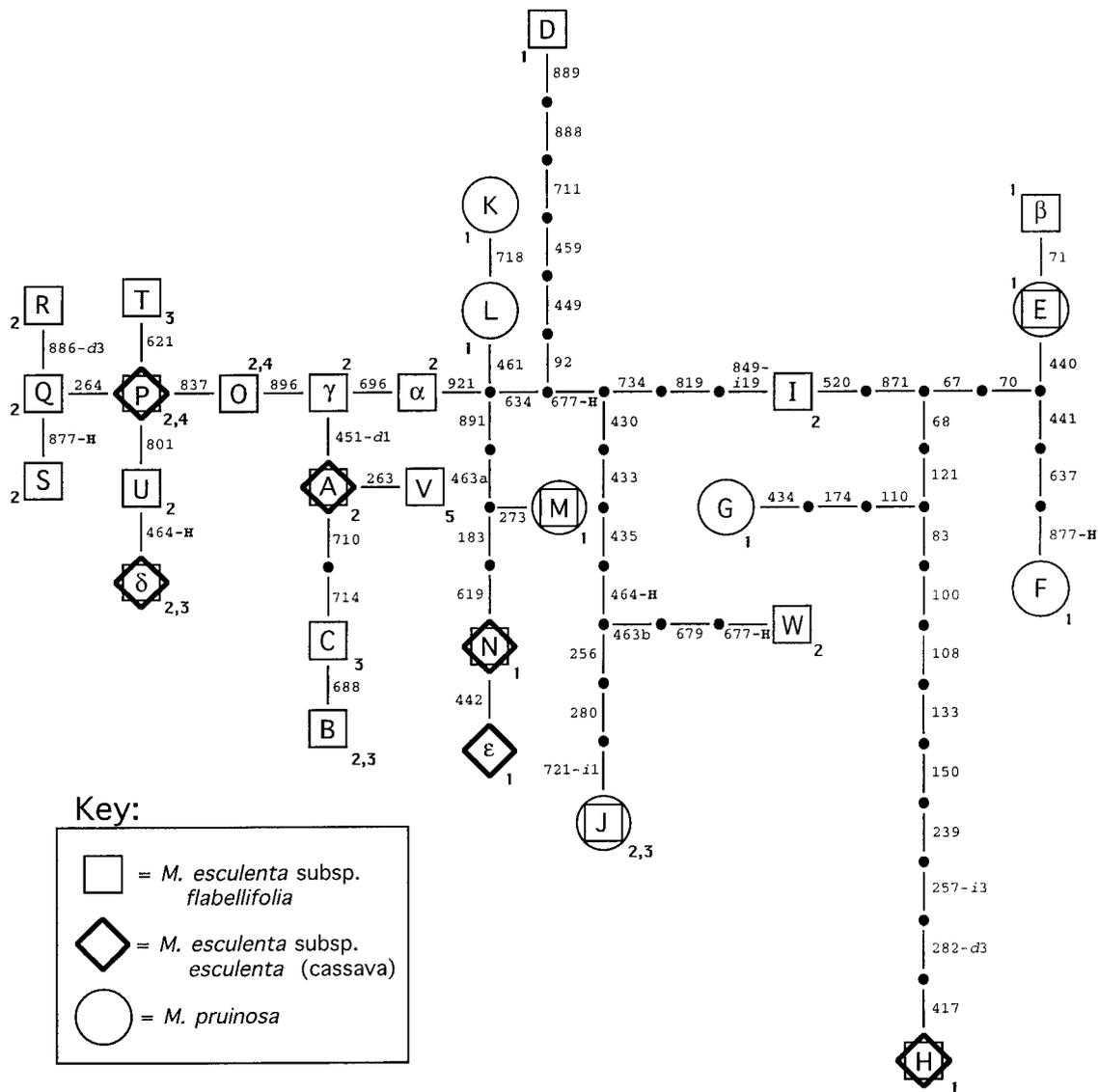
FIG. 3.   The *G3pdh* gene tree. Haplotype letters correspond to those in Table 2. Shapes around letters indicate the taxa in which haplotypes were found. Each line between haplotypes represents a mutational step, with numbers on lines indicating the variable base pair position. For branches of more than one mutational step without intermediate haplotypes (or side branches), the relative placement of the mutations along the branch is arbitrary. Insertions (*i*) and deletions (*d*) relative to the center of the tree are indicated after the base pair number, followed by the number of bases involved. Two different substitutions at position 463 are designated as a and b. Homoplasious (**H**) mutations are indicated after the base pair number. Numbers in bold print indicate the number of 25-bp minisatellite motifs associated with each haplotype.

variation in the *G3pdh* region conforms to expectations of neutrality, both by Tajima's criterion ($D = 0.0899$, $P > 0.5$ for β-distribution approximation) and by Fu and Li's criteria ($D^* = -1.712, P > 0.1; F^* = -1.950, P > 0.1$). These data represent one of the largest samples published to date for a plant phylogeography study.

Variation at polymorphic sites (minisatellite excluded) could be ordered into a maximum parsimony gene tree, with 28 haplotypes defined among the 212 individuals (424 alleles) in the study (Fig. 3; Table 2). Three polymorphic sites were homoplasious, meaning that the same mutational step had to be mapped onto the tree more than once (Fig. 3); all of these sites involve single-base substitutions. For one of the homoplasious sites (base 677), there are two equally parsimonious arrangements for the two mutational steps (relative to the branches leading to haplotypes J and W); this ambiguity affects the relative positions of haplotypes on the tree only minimally. The low level of homoplasy observed for this data set indicates that interallelic recombination has not occurred in the *G3pdh* region.

Cassava's Geographical Origin. Of the 28 haplotypes observed in the study system, 24 were detected in *M. esculenta*. Of these 24, 6 occur in cultivated cassava, 1 exclusively so (Table 2). When

Table 2.   Distribution of *G3pdh* haplotypes in the study system

Population (n)	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	α	β	γ	δ	ε
Frequency:	0.0590	0.0236	0.1533	0.0189	0.1745	0.0283	0.0425	0.0330	0.0189	0.0307	0.0047	0.0425	0.0542	0.0330	0.0165	0.0896	0.0118	0.0142	0.0142	0.0094	0.0495	0.0259	0.0142	0.0071	0.0094	0.0047	0.0118	0.0047
M. esculenta subsp. *flabellifolia*																												
Axixá, TO (6)					E																							
Luzinópolis, TO (6)					E																			α	β			
Miranorte, TO (8)				D	E																		W	α	β	γ		
Dueré, TO (6)																						V				γ		
Campos Belos, GO (6)			C																									
Campinorte, GO (6)			C		E																							
Rialma, GO (6)			C		E																							
Corumbá, GO (6)			C		E																							
Nerópolis, GO (6)			C		E																							
Goiás Velho, GO (6)					E					J																		
Iporá, GO (6)			C		E																							
Caiapônia, GO (6)			C		E																							
Nova Xavantina, MT (6)			C																									
Serra Petrovina, MT (5)									I	J																		
Santa Elvira, MT (2)										J				**N**														
São Vincente, MT (4)																**P**		R										
Lambari d'Oeste, MT (6)	**A**	B								J																		
Pontes e Lacerda, MT-A (6)	**A**	B	C					**H**							O													
Pontes e Lacerda, MT-B (6)										J					O	**P**	Q											
Vilhena, RO (6)				D										**N**														
Pimenta Bueno, RO (6)														**N**					S									
Jarú, RO (6)													M	**N**		**P**												
Ariquemes, RO (6)			C																		U							
Teotônio, RO (6)								**H**							O	**P**				T							**δ**	
Taquaras, RO (6)								**H**							O	**P**												
Rio Branco, AC (6)																**P**				T	U							
Sena Madureira, AC (6)	**A**		C													**P**												
M. esculenta subsp. *esculenta* (cassava) (20)	**A**							**H**						**N**		**P**											**δ**	**ε**
AA:MBRA12, MCOL1684, MPAR110, MVEN25																												
PP:MBOL3, MCOL2215, MNGA2																												
AP:MARG11, MCR32, MCUB74, MECU82, MIND33, MMAL2, MPTR19, MTAI1																												
PN: MCOL1505																												
Aδ: MBRA931, MPAN51																												
Aε:HMC1																												
Hε: MMEX59																												
M. pruinosa																												
Miranorte, TO (6)							G					L	M															
Divinópolis, GO (6)							G					L																
East Iporá, GO (5)							G			J		L	M															
Iporá, GO (6)					E		G					L	M															
Nova Xavantina, MT (6)							G				K		M															
Serra Petrovina, MT (6)						F																						

Populations of wild taxa are listed in clockwise order around the eastern and southern border of the Amazon basin (see Fig. 1 for definition of abbreviations of Brazilian states). *n*, number of individuals sequenced per population. Haplotypes in bold print were detected in cassava. Greek letters indicate haplotypes detected in heterozygous individuals only. Cassava accessions are listed by genotype.

considered in a geographical context, the haplotypes shared between cassava and the wild *M. esculenta* subspecies occur along the southern border of the Amazon basin (the states of Mato Grosso, Rondônia, and Acre) but not in the eastern border region (Goiás and Tocantins; Fig. 1; Table 2). This distribution strongly points to the southern Amazon border region as the crop's geographical origin of domestication. Within this region, all of the shared haplotypes can be accounted for in a limited area of six sampled populations, between western Mato Grosso and western Rondônia (Fig. 1; Table 2). Thus, the crop potentially could be derived from a relatively restricted area in the southern Amazon border region. Alternatively, the origin of domestication may extend further west and include Peruvian populations of subspecies *flabellifolia*, at the westernmost limit of the species range.

The *G3pdh* data provide the most conclusive and most specific evidence to date on the geographical origin of cassava. Previous hypotheses on the crop's origins have been largely speculative and have included such disparate regions as the arid scrub (*caatinga*) of northeastern Brazil (37, 38), the savannas of Colombia and Venezuela (39), the prehistoric savannas of western Peru (24), and the lowlands of Mexico and Central America (22, 23). This earlier conjecture was confounded by two factors: uncertainty about the position of *M. esculenta* within *Manihot* and equivocal anthropological data on early cassava cultivation. Until subspecies *flabellifolia* came to be recognized as cassava's closest wild relative (19, 25), no single wild *Manihot* species could be identified as the crop's obvious progenitor. The numerous candidate species proposed (18, 20) together encompass a vast geographical range in Mexico, Central America, and South America. Many of these species have since been shown to be related to cassava only distantly (refs. 26, 27, and 40; B.A.S., unpublished data). Anthropological data, often useful in tracing a crop's origins, proved similarly inconclusive in pinpointing a region of domestication. Cassava was already widely grown throughout the Neotropics by 3,000 years ago, making its earliest cultivation very difficult to trace (20, 21, 41). Moreover, as a crop of humid lowlands, cassava is preserved particularly poorly in the archaeobotanical record (21), and most recovered remains are from arid sites that do not reflect its earliest use (41).

Haplotype Distributions in Cassava and Wild Taxa. *Within* M. esculenta. One-fourth of the haplotypes observed in *M. esculenta* overall occur in cassava, indicating that genetic variation within the crop is a subset of that found in the wild subspecies (Table 2). This observation concurs with a recent amplified fragment length polymorphism-based study (27), which found that a collection of 38 cassava cultivars contain about 30% of the genetic variation detected in a sample of 14 wild *M. esculenta* accessions. This pattern of reduced genetic diversity with domestication seems to be the rule for crop-relative systems (42–45) and presumably reflects artificial selection and founder-event-induced genetic drift over the course of domestication (10, 42, 46). This pattern also confirms the progenitor-derivative relationship between subspecies *flabellifolia* and cassava.

Cassava differs markedly from subspecies *flabellifolia* with respect to haplotype frequencies and levels of heterozygosity. Among the 20 cassava accessions, 2 haplotypes (A and P) represent 85% of the haplotype variation (i.e., 34 of 40 haplotypes; Table 2), whereas these 2 haplotypes account for only 9% of the haplotypes detected in the wild *M. esculenta* subspecies. In terms of heterozygosity, 13 of the 20 cassava cultivars (65%) are heterozygotes (Table 2), in contrast to 58 of the 157 wild *M. esculenta* individuals (37%). The significance of these differences between the crop and its relatives is difficult to assess, as they reflect a number of interacting evolutionary factors (42). For the wild taxa of the study system, data on haplotype frequencies and heterozygosity have yet to be analyzed by phylogeographic analysis (47), which would focus on relationships among the natural populations.

Interspecific level. At the interspecific level, 4 of the 28 *G3pdh* haplotypes occur exclusively in *M. pruinosa*; however, an addi-

tional 3 haplotypes (E, J, and M) are shared between this species and *M. esculenta* (Fig. 3; Table 2). All three of the transspecific haplotypes are found primarily within one species and are observed in a single individual or single population of the other (Table 2). These shared haplotypes may reflect interspecific introgression or, alternatively, may be ancestral polymorphisms that predate the divergence of the two species (see below). Interspecific introgression is expected only where the two species are sympatric. Thus, two of the shared haplotypes could potentially be explained by introgression (E and J; near Iporá, GO), whereas the third cannot (M; Jarú, RO; Fig. 1; Table 2). Significantly, none of the transspecific haplotypes occur in cassava, suggesting that if there has been interspecific introgression, *M. pruinosa* has not contributed directly to the crop germplasm. This conclusion is supported further by the observation that all cassava haplotypes occur west of the range of *M. pruinosa* (Fig. 1), indicating little contact between *M. pruinosa* and cassava's closest wild progenitors.

Genealogical Relationships Among Haplotypes. The *G3pdh* tree (Fig. 3) indicates no clear association between the genealogical relationships of haplotypes and the taxonomic categories in which they are found. Cassava haplotypes do not cluster into a monophyletic clade within the other *M. esculenta* haplotypes but rather are scattered across the gene tree. Haplotypes of the two different species are similarly intermingled. This visual assessment is confirmed by an analysis of molecular variance (48), which indicates that 83% of the genetic variation among haplotypes occurs within taxa (data not shown). As with the presence of transspecific haplotypes, this lack of congruence between the gene tree and taxonomic categories could arise potentially either by gene flow among the taxa (following the divergence of the separate lineages) or by the persistence and sorting of ancestral polymorphisms that predate the divergence of the taxonomic lineages (lineage sorting).

For the two subspecies within *M. esculenta*, haplotype distributions almost certainly reflect lineage sorting. Cassava was domesticated no more than 10,000 years ago (the origin of human agriculture), which makes the divergence of the crop lineage a very recent evolutionary event. As such, haplotypes observed in cassava most likely predate the crop's origin, and the presence of these haplotypes in subspecies *flabellifolia* therefore reflects their persistence in both lineages after domestication. The haplotype observed only in cassava (ε) may be an exception, possibly having arisen since domestication. This haplotype is positioned as a tip on the gene tree and occurs at a very low frequency (Fig. 3, Table 2), characteristics consistent with a recent evolutionary origin (49, 50). Alternatively, this haplotype may be as old as the other cassava haplotypes and may have either gone extinct in the wild subspecies or simply been missed during population sampling.

Lineage sorting probably has also played an important role in the distribution of haplotypes at the interspecific level. The genus *Manihot* is believed to have arisen and diversified recently (18), an argument supported by a lack of variability in chromosome number, by low levels of divergence in floral morphology (18) and DNA sequence data (40), and by interfertility between morphologically divergent species in artificial crosses (26, 27). Thus, *M. esculenta* and *M. pruinosa* are likely to share very recent common ancestry, indicating a high probability of lineage sorting (51, 52). In addition, the probability of encountering ancestral polymorphisms is greater with the *G3pdh* locus than it would be for chloroplast or mitochondrial genes, all else being equal; the diploid nature of the nuclear genome doubles its effective population size for a monoecious species, thereby doubling the expected time to coalescence (53). Thus, the *G3pdh* haplotypes may reflect ancestral polymorphisms on a time scale where haplotypes from the other genomes would not. Finally, if interspecific introgression were to have played a major role in structuring the taxonomic distribution (shown in Fig. 3), one would expect some additional evidence for hybridization in the study system. Instead, there is very little evidence that the species

Evolution: Olsen and Schaal

Proc. Natl. Acad. Sci. USA 96 (1999)   5591

hybridize. *M. pruinosa* has been included in cassava's secondary gene pool (28) on the basis of morphological similarity, rather than on documented evidence of interfertility, and no putative hybrids were identified during field collections for the present study. Nonetheless, we do not reject the possibility of hybridization between these species.

CONCLUSIONS

This study demonstrates the use of a single-copy nuclear gene to examine the phylogeography of a plant species, in this case a crop-relative system consisting of cassava and its wild progenitors. The level of phylogenetically informative variation in the *G3pdh* region far exceeds levels typically observed at the intraspecific level in the organellar genomes of plants; moreover, the *G3pdh* data are nearly free of homoplasy. These findings suggest that single-copy nuclear genes may provide a rich source of intraspecific variation for plant phylogeography studies, possibly analogous to the rapidly evolving regions of animal mtDNA.

Although much of the *G3pdh* variation likely predates the divergences of the taxa within the study system, these haplotypes have become structured geographically among the wild populations (Fig. 1; Table 2) and therefore provide insight into cassava's origin of domestication. The crop seems to be derived from populations of subspecies *flabellifolia* along the southern border of the Amazon basin. Moreover, the pattern and degree of variation in the crop versus the wild *M. esculenta* and *M. pruinosa* populations indicate that subspecies *flabellifolia* alone can account for the genetic variation observed in cassava. There is no need to infer the involvement of other progenitor species in the crop's origin, at least given present sampling and data. Overall, the picture of cassava's origin provided by the *G3pdh* data represents a fundamental departure from the traditional view, which was accepted until very recently (e.g., refs. 21 and 54), that envisioned this crop as a compilospecies derived from one or more complexes of interbreeding wild species, from one or more regions in the Neotropics. This insight into the crop's origin will be useful for targeting wild germplasm for crop improvement (45, 55) and for understanding the genetic consequences of domestication in this species (10, 44). Ultimately, the findings of this study will be best understood once they can be placed in the broader context of a well resolved *Manihot* species phylogeny.

1. Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. & Saunders, N. C. (1987) *Annu. Rev. Ecol. Syst.* **18,** 489–522.
2. Bermingham, E. & Moritz, C. (1998) *Mol. Ecol.* **7,** 367–369.
3. Avise, J. C. (1998) *Mol. Ecol.* **7,** 371–379.
4. Avise, J. C. (1989) *Evolution* **43,** 1192–1208.
5. Schaal, B. A., Hayworth, D. A., Olsen, K. M., Rauscher, J. T. & Smith, W. A. (1998) *Mol. Ecol.* **7,** 465–474.
6. Soltis, D. E., Gitzendanner, M. A., Strenge, D. D. & Soltis, P. S. (1997) *Plant Syst. Evol.* **206,** 353–373.
7. Levy, F., Antonovics, J., Boyton, J. E. & Gillham, N. W. (1996) *Heredity* **76,** 143–155.
8. Tomaru, N., Takahashi, M., Tsumura, Y., Takahashi, M. & Ohba, K. (1998) *Am. J. Bot.* **85,** 629–636.
9. Strand, A. E., Leebens-Mack, J. & Milligan, B. G. (1997) *Mol. Ecol.* **6,** 113–118.
10. Eyre-Walker, A., Gaut, R. L., Hilton, H., Feldman, D. L. & Gaut, B. S. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 4441–4446.
11. Hilton, H. & Gaut, B. S. (1998) *Genetics* **150,** 863–872.
12. Bernardi, G., Sordino, P. & Powers, D. A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 9271–9274.
13. Hare, M. P. & Avise, J. C. (1998) *Mol. Biol. Evol.* **15,** 119–128.
14. Palumbi, S. R. & Baker, C. S. (1994) *Mol. Biol. Evol.* **11,** 426–435.
15. Best, R. & Henry, G. (1992) in *Report of the First Meeting of the International Network for Cassava Genetic Resources,* eds. Roca, W. M. & Thro, A. M. (Cent. Int. Agric. Trop., Cali, Colombia), pp. 3–11.
16. Cock, J. H. (1985) *Cassava: New Potential for a Neglected Crop* (Westfield, London).
17. Mann, C. (1997) *Science* **277,** 1038–1043.
18. Rogers, D. J. & Appan, S. G. (1973) *Manihot and Manihotoides (Euphorbiaceae): A Computer Assisted Study* (Hafner, New York).
19. Allem, A. C. (1994) *Genet. Res. Crop Evol.* **41,** 133–150.
20. Renvoize, B. S. (1972) *Econ. Bot.* **26,** 352–360.
21. Sauer, J. D. (1993) *Historical Geography of Crop Plants* (CRC, Boca Raton, FL).
22. Rogers, D. J. (1965) *Econ. Bot.* **19,** 369–377.
23. Rogers, D. J. (1963) *Bull. Torrey Bot. Club* **90,** 43–54.
24. Ugent, D., Pozorski, S. & Pozorski, T. (1986) *Econ. Bot.* **40,** 78–102.
25. Allem, A. C. (1987) *Plant Genet. Resour. Newsl.* **71,** 22–24.
26. Fregene, M. A., Vargas, J., Ikea, J., Angel, F., Tohme, J., Asiedu, R. A., Akoroda, M. O. & Roca, W. M. (1994) *Theor. Appl. Genet.* **89,** 719–727.
27. Roa, A. C., Maya, M. M., Duque, M. C., Tohme, J., Allem, A. C. & Bonierbale, M. W. (1997) *Theor. Appl. Genet.* **95,** 741–750.
28. Allem, A. C. (1992) in *Report of the First Meeting of the International Network for Cassava Genetic Resources,* eds. Roca, W. M. & Thro, A. M. (Cent. Int. Agric. Trop., Cali, Colombia), pp. 87–110.
29. Hillis, D. M., Maple, B. K., Larson, A., Davis, S. K. & Zimmer, E. A. (1996) in *Molecular Systematics,* eds. Hillis, D. M., Moritz, C. & Maple, B. K. (Sinauer, Sutherland, MA), pp. 321–381.
30. Maddison, W. P. & Maddison, D. R. (1992) MACCLADE (Sinauer, Sunderland, MA).
31. Tajima, F. (1989) *Genetics* **123,** 585–595.
32. Fu, Y.-X. & Li, W.-H. (1993) *Genetics* **133,** 693–709.
33. Fregene, M., Angel, F., Gomez, R., Rodriguez, F., Chavarriaga, P., Roca, W., Tohme, J. & Bonierbale, M. (1997) *Theor. Appl. Genet.* **95,** 431–441.
34. Clark, A. G. (1990) *Mol. Biol. Evol.* **7,** 111–122.
35. Swofford, D. L. (1993) PAUP, Phylogenetic Analysis Using Parsimony (Smithsonian Institution, Washington, DC), Version 3.1.
36. Griffiths, A. J. F., Miller, J. H., Suzuki, D. T., Lewontin, R. C. & Gelbart, W. M. (1993) *An Introduction to Genetic Analysis* (Freeman, New York), 5th Ed.
37. De Candolle, A. (1883) *L'Origine des plantes cultivées* (Diderot, Paris); 2nd ed. (1886) reprinted, 1967 (Hafner, New York).
38. Vavilov, N. I. (1951) *The Origin, Variation, Immunity, and Breeding of Cultivated Plants: Selected Writings* (Chronica Botanica, Waltham, MA).
39. Sauer, C. O. (1952) *Agricultural Origins and Dispersals* (Am. Geogr. Soc., New York).
40. Schaal, B. A., Olson, P. D., Prinzie, T. P., Carvalho, L. J. C. B., Tonukari, J. & Hayworth, D.A. (1994) *Cassava Biotechnology Network Proceedings* (Cent. Int. Agric. Trop., Cali, Colombia), pp. 62–70.
41. Pearsall, D. M. (1992) in *The Origins of Agriculture: An International Perspective,* eds. Cowan, C. W., Watson, P. J., Benco, N. L. (Smithsonian Institution, Washington, DC), pp. 173–205.
42. Doebley, J. (1989) in *Isozymes in Plant Biology,* eds. Soltis, D. E. & Soltis, P. S. (Dioscorides, Portland, OR), pp. 165–191.
43. Doebley, J. (1992) in *Molecular Systematics of Plants,* eds. Soltis, P. S., Soltis, D. E. & Doyle, J. J. (Chapman & Hall, New York), pp. 202–222.
44. Gepts, P. (1993) in *Evolutionary Biology,* ed. Hecht, M. K. (Plenum, New York), pp. 51–94.
45. Tanksley, S. D. & McCouch, S. R. (1997) *Science* **277,** 1063–1066.
46. Ladizinsky, G. (1985) *Econ. Bot.* **39,** 191–198.
47. Templeton, A. R., Routman, E. & Phillips, C. A. (1995) *Genetics* **140,** 767–782.
48. Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992) *Genetics* **131,** 479–491.
49. Golding, G. B. (1987) *Genet. Res.* **49,** 71–82.
50. Castelloe, J. & Templeton, A. R. (1994) *Mol. Phylogenet. Evol.* **3,** 102–113.
51. Neigel, J. E. & Avise, J. C. (1986) in *Evolutionary Processes and Theory,* eds. Nevo, E. & Karlin, S. (Academic, NY), pp. 515–534.
52. Wu, C.-I. (1991) *Genetics* **127,** 429–435.
53. Birky, C. W., Jr., Fuerst, P. & Maruyama, T. (1989) *Genetics* **121,** 613–627.
54. Jennings, D. L. (1995) in *Evolution of Crop Plants,* eds. Smartt, J. & Simmonds, N. W. (Wiley, New York), pp. 128–132.
55. Hoyt, E. (1992) *Conserving the Wild Relatives of Crops,* (Int. Board Plant Genet. Resour., Rome).