Journal of Heredity 2006:97(4):403–408
doi:10.1093/jhered/esl001
Advance Access publication June 1, 2006

Chloroplast DNA Variation Confirms a Single Origin of Domesticated Sunflower (*Helianthus annuus* L.)

David M. Wills and John M. Burke

From the Department of Biological Sciences, Vanderbilt University, VU Station B 351634, Nashville, TN 37235.

Address correspondence to J. M. Burke at the address above, or e-mail: john.m.burke@vanderbilt.edu. Address as of 1 August 2006 will be Department of Plant Biology, University of Georgia, Athens, GA 30602, or e-mail: jmburke@uga.edu.

Abstract

Although sunflower was long thought to be the product of a single domestication in what is now the east-central United States, recent archaeological and genetic evidence have suggested the possibility of an independent origin of domestication, perhaps in Mexico. We therefore used hypervariable chloroplast simple-sequence repeat markers to search for evidence of a possible Mexican origin of domestication. This work resulted in the identification of 45 chloroplast haplotypes from 26 populations across the range of wild sunflower as well as 3 haplotypes from 15 domesticated lines, representing both primitive and improved cultivars. The 3 domesticated haplotypes were characterized by 1 primary haplotype (found at a frequency of 6.7% in the wild) as well as 2 rare haplotypes, which are most likely the products of mutation or introgression. One of these rare haplotypes was not observed in the wild, bringing the total number of haplotypes identifited to 46. A principal coordinate analysis revealed the presence of 3 major haplotype clusters, one of which contained the primary domesticated haplotype, the 2 rare domesticated variants, as well as haplotypes found across much of the range of wild sunflower. The Mexican haplotypes, on the other hand, fell well outside of this cluster. Although our data do not provide insight into the specific location of sunflower domestication, the relative rarity of the primary domesticated haplotype in the wild, combined with the dissimilarity between this haplotype and those found in the Mexican populations surveyed, provides further evidence that the extant domesticated sunflowers are the product of a single domestication event somewhere outside of Mexico.

The majority of crop plants were domesticated between 4000 and 10 000 years ago (Hancock 2004) and, in most cases, the wild progenitors of these crops have been satisfactorily identified. We are, however, continually gaining insight into the details surrounding the domestication of these plants. For example, it is clear in some cases (such as barley, maize, and potato) that the crop form arose just once (Badr and others 2000; Matsuoka and others 2002; Spooner and others 2005). Thus, the current ranges of cultivation of these crops reflect postdomestication diffusion from their centers of origin. In other cases, such as rice, cotton, and soybean, the crop appears to be the product of multiple origins of domestication (Second 1982; Wendel and others 1995; and Xu and others 2002), sometimes in geographically disparate locales. Here we report the results of an investigation into the origin of domesticated sunflower based on patterns of chloroplast DNA (cpDNA) variation.

Domesticated sunflower (*Helianthus annuus*) is one of the world's most important oilseed crops and is also a major source of confectionery seeds (Putt 1997). Derived from the common sunflower (also *H. annuus*), domesticated sunflower was initially thought to have arisen just once in what is now the east-central United States (Heiser 1954, 1978). In fact, Heiser (1954) first hypothesized that the use of wild sunflowers by Native Americans as a food source resulted in the production of a camp-following weed that eventually spread eastward and that this weed ultimately served as the progenitor of domesticated sunflower. However, Heiser (1985) later discussed the possibility of an additional origin of domestication, perhaps in Mexico. Until recently (see below), the available archaeological evidence (Brewer 1973; Ford 1985; Crites 1993) was most consistent with the single-origin hypothesis, with carbonized achenes (i.e., single-seeded fruits) from the Hayes site in Middle Tennessee providing the earliest record of domesticated sunflower (ca. 4300 years before present [YBP]; Crites 1993).

In terms of genetic data, Rieseberg and Seiler (1990) surveyed a broad collection of wild and domesticated sunflower lines and found that the domesticates exhibited reduced allozyme variability and that they were all characterized by a single cpDNA restriction fragment length polymorphism (RFLP) haplotype. Although this result is consistent with a single origin of domestication, these data are far from conclusive as the domesticated cpDNA haplotype was

geographically widespread and present at relatively high frequency (27%) in the wild. It is thus conceivable that independently derived lines could share the same chloroplast haplotype by chance. In a subsequent survey of allozyme polymorphism, however, Cronn and others (1997) reported that the domesticates form a "genetically coherent group" (p. 532), a result that was once again consistent with the hypothesis of a single origin. The possibility of a second origin of domestication was thus eventually dismissed based on the total weight of the archaeological and genetic evidence available at the time (Seiler and Rieseberg 1997).

The debate over the origin of domesticated sunflower was, however, revived when Lentz and others (2001) reported the discovery of carbonized achenes of domesticated sunflower in southern Mexico, beyond the current range of wild sunflower. These achenes dated to roughly the same time period as those recovered at the Hayes site (ca. 4000 YBP), and no older achenes have been recovered since that time. Shortly after this discovery, Tang and Knapp (2003) used a suite of 122 nuclear simple-sequence repeats (SSRs) to examine patterns of genetic diversity in both wild and domesticated sunflower. Based on their results, these authors concluded that "the single ancestor hypothesis ... seems improbable" (p. 999). Rather, they suggested that the Hopi and Havasupai landraces, which are separated from the balance of the domesticates by a substantial genetic distance, might represent the descendants of the hypothesized "other" origin of domestication. Adding to this is the fact that the Hopi and Havasupai lines are native to the desert southwest, making them the geographically most proximate landraces to the previously hypothesized Mexican origin of domestication. In the most comprehensive molecular analysis to date, however, Harter and others (2004) argued that the 8 extant Native American landraces, from which the modern cultivars are presumably derived, can all be reliably assigned to a single population genetic cluster based on patterns of nuclear SSR diversity. This result led them to conclude that these lines do, in fact, trace back to a single origin of domestication, most likely somewhere in central North America. Under their interpretation, the Hopi and Havasupai landraces represent the most primitive of the extant domesticates.

Here we reconsider the issue of single vs. multiple origins of sunflower domestication based on patterns of cpDNA variation in wild and domesticated sunflower. More specifically, we investigate the question of whether or not the Hopi and Havasupai landraces represent the descendants of an independent origin of sunflower domestication in Mexico. In order to answer this question, we used a suite of highly variable chloroplast SSRs (cpSSRs), which provided us with far greater levels of population genetic resolution than were available at the time of the original RFLP-based survey of cpDNA diversity by Rieseberg and Seiler (1990).

Materials and Methods

Plant Materials and DNA Extractions

Wild and domesticated sunflower accessions were obtained from the North Central Regional Plant Introduction Station (NCRPIS, Ames, IA). Twenty-six wild accessions were selected to represent the range of common sunflower across North America, whereas 15 domesticated lines were selected to represent the Native American landraces as well as improved lines (Table 1). Seeds were sown in flats and allowed to germinate in the greenhouse. After seedling emergence, 200 mg of leaf tissue was collected from each of 4–6 individuals per accession. Total genomic DNA was then extracted from each sample using the Qiagen DNeasy Plant Mini Kit (Qiagen, Valencia, CA).

cpSSR Genotyping and Analysis

Conserved primer pairs that flank cpSSRs have been identified from a number of angiosperm species (Powell and others 1995; Bryan and others 1999; Weising and Gardner 1999). We used 6 such primer pairs, which have previously been shown to reveal polymorphisms within *H. annuus* (ccmp 2, ccmp 7, NTCP 9, NTCP 30, NTCP 40, and NTCP 18; Wills and others 2005), to genotype each individual. We used a modification of the polymerase chain reaction (PCR) methodology presented by Schuelke (2000), wherein we added an arbitrarily selected sequence (the M13 forward [−29] sequencing primer, 5′-CAC GAC GTT GTA AAA CGA C-3′) to the 5′ end of the forward primer. In order to allow for the visualization of multiple loci per lane on an automated DNA sequencer, PCR products were labeled by including a fluorescently tagged M13 forward (−29) primer (carrying either HEX, FAM, or TET) in the reaction mixture. Reactions were performed in 10-μl total volume containing 10 ng of template DNA, 30 mM tricine pH 8.4-KOH, 50 mM KCl, 2 mM $MgCl_2$, 100 μM of each deoxynucleoside triphosphate, 0.02 μM forward primer, 0.1 μM of both the reverse primer and the fluorescently labeled M13 primer, and 2 units of *Taq* polymerase. Cycling conditions were as follows: initial denaturation at 95 °C for 3 min; followed by 10 cycles of 30 s at 94 °C, 30 s at 58 °C (annealing temperature was reduced by one degree per cycle), and 45 s at 72 °C; followed by 30 cycles of 30 s at 94 °C, 30 s at 48 °C, and 45 s at 72 °C; and a final extension time of 20 min at 72 °C.

Amplification products were visualized on an MJ Research BaseStation automated DNA sequencer (South San Francisco, CA) with MapMarker® 1000 ROX size standards (BioVentures Inc., Murfreesboro, TN) included in each lane to allow for accurate determination of fragment size. Alleles were called using the software package CARTOGRAPHER (MJ Research), and the resulting data were analyzed using ARLEQUIN (Excoffier and others 2005) to generate summary statistics and GENALEX 6 (Peakall and Smouse 2006) to perform a principal coordinate (PCO) analysis.

Results and Discussion

Chloroplast Diversity

We determined the chloroplast haplotype of 4–6 individuals from each of the 26 wild populations and 15 domesticated lineages using the 6 cpSSRs described above. Individual

Table 1. Accession numbers and improvement status of wild populations and domesticated lines surveyed. Each haplotype was assigned a number representing its rank from most to least frequent, and the identity of the haplotypes found in each accession is reported. In addition, the PCO cluster to which each haplotype was assigned is included. See text for details

Sample location	Accession ID	Improvement status	Sample size	Haplotype ID	PCO cluster
Arizona, USA	Ames 14400	Wild	6	2, 22	1, 3
Arkansas, USA	PI 613727	Wild	6	5	1
California, USA	PI 613732	Wild	6	12, 29, 30	1, 2
Colorado, USA	PI 586840	Wild	6	3, 15, 25, 31, 32	1, 2
Iowa, USA	PI 597895	Wild	5	10	1
Illinois, USA	PI 547168	Wild	5	1, 23	1
Kansas, USA	PI 413027	Wild	6	13, 24	1
Minnesota, USA	PI 613745	Wild	6	2	1
Missouri, USA	PI 413011	Wild	5	11	1
Montana, USA	PI 531032	Wild	6	3, 21, 34, 35	1, 3
North Dakota, USA	PI 596910	Wild	6	18, 19	1, 3
Nebraska, USA	PI 586865	Wild	6	20, 26, 36	1, 2
Ohio, USA	Ames 23238	Wild	4	14	1
Oklahoma, USA	PI 435619	Wild	6	1	1
Oregon, USA	PI 531015	Wild	6	4	1
South Dakota, USA	Ames 23940	Wild	6	1, 5, 27, 43	1
Tennessee, USA	PI 435552	Wild	6	1	1
Texas, USA	Ames 7442	Wild	6	8	1
Utah, USA	PI 531009	Wild	6	9	1
Washington, USA	PI 531016	Wild	5	4, 15	1
Wyoming, USA	PI 586822	Wild	6	21, 28, 44, 45	1, other
Alberta, Canada	PI 592308	Wild	6	3, 16	1, 3
Manitoba, Canada	PI 592327	Wild	5	17, 25, 33	1
Saskatchewan, Canada	PI 592317	Wild	6	37, 38, 39, 40, 41, 42	1, 2, 3, other
España, Mexico	PI 413067	Wild	6	7	Other
Mayo, Mexico	PI 413123	Wild	6	6	3
Hopi	PI 432504	Domesticated	6	2	1
Havasupai	PI 369358	Domesticated	6	2	1
Seneca	PI 369360	Domesticated	4	2	1
Mandan	PI 600717	Domesticated	4	2	1
Hidatsa	PI 600721	Domesticated	4	46	1
Hidatsa[a]	PI 600720	Domesticated	6	46	1
Arikara	PI 369357	Domesticated	4	2	1
Maíz de Tejas	Ames 6859	Domesticated	4	2, 35	1
Maíz Negro	Ames 19070	Domesticated	4	2	1
Mennonite	Ames 7574	Improved	4	2	1
Jupiter	PI 296289	Improved	4	2	1
Tchernianka Select W-13	PI 343794	Improved	4	2	1
Sunrise	PI 162454	Improved	5	2	1
Mammoth	PI 476853	Improved	5	2	1
Damaya	PI 496263	Improved	5	2	1

[a] Denotes the second Hidatsa accession that was surveyed to confirm the occurrence of a unique haplotype within this landrace.

cpSSR loci harbored an average of 6.0 ± 1.1 alleles per locus (mean ± standard error), resulting in an average gene diversity of 0.59 ± 0.34 (Table 2). Because the chloroplast genome is thought to be a nonrecombining unit, we tested for linkage disequilibrium among loci using Slatkin's (1994) extension of Fisher's exact test. As expected, all loci were found to be in strong disequilibrium with one another (all $P < 0.001$).

In total, we identified 45 unique wild sunflower haplotypes, with the most common haplotype occurring at a frequency of 10.7% (Figure 1). In contrast, 17 wild individuals carried unique haplotypes. With the exception of the Hidatsa and Maíz de Tejas landraces, all domesticated individuals shared a single haplotype (hereafter referred to as the "primary" domesticated haplotype), which was the second most common haplotype found in the wild, occurring at a frequency of 6.7%. All 6 of the Hidatsa individuals that were initially surveyed shared a unique haplotype that differed from the primary domesticated haplotype at 2 of the 6 loci (ccmp 7 and NTCP 18) and was not found in the other domesticated lineages or in any of the wild populations. To confirm this finding, we requested a second Hidatsa accession from the NCRPIS and genotyped 6 additional individuals as described above. All 6 of these individuals contained the same haplotype that was found in the first Hidatsa accession. In the case of Maíz de Tejas, 3 of the 4 individuals surveyed exhibited the primary domesticated haplotype, whereas the fourth contained a haplotype that differed by a single base pair at one locus (NTCP 18).

Table 2. Results of our survey of cpSSR polymorphisms across 26 wild *Helianthus annuus* populations. Allele sizes are reported in base pairs and reflect the inclusion of the 19-bp extension on the 5′ end of the forward primer (see Materials and Methods for additional details)

Locus name	Primer sequences (5′–3′)	Allele size range	Number of alleles	Gene diversity
ccmp 2	F: GATCCCGGACGTAATCCTG R: ATCGTACCGAGGGTTCGAAT	228–230	3	0.452
ccmp 7	F: CAACATATACCACTGTCAAG R: ACATCATTATTGTATACTCTTTC	139–149	9	0.771
NTCP 9	F: CTTCCAAGCTAACGATGC R: CTGTCCTATCCATTAAGACAATG	279–284	6	0.758
NTCP 18	F: CTGTTCTTTCCATGACCCCTC R: CCACCTAGCCAAGCCAGA	207–218	9	0.607
NTCP 30	F: GATGGCTCCGTTGCTTTAT R: TGCCGGAGAGTTCTTAACAATA	176–181	6	0.606
NTCP 40	F: TAATTTGATTCTTCGTCGC R: GATGTAGCCAAGTGGATCA	277–278	3	0.346
Mean		NA	6.0	0.590

PCO analysis was performed using the 46 cpSSR haplotypes that we identified (43 of which were found in wild sunflower only, 2 of which were shared between the wild and domesticated accessions, and 1 of which was unique to the Hidatsa landrace). Using the "distance not standardized" setting in GENALEX 6 (Peakall and Smouse 2006), wherein each of the 3–9 alleles per locus (mean = 6.0) was considered to be a single mutational step from all others, the first 2 coordinates explained 44.8% of the total variance. Inspection of Figure 2 reveals that the haplotypes appear to form 3 main clusters with 3 outlying haplotypes and all 3 of the domesticated haplotypes occurring in the largest cluster. This cluster contains 30 of the 46 haplotypes, including the 5 most common haplotypes across the range of wild sunflower. Note that the 2 wild accessions from Mexico (#6 and #7) fell outside of this cluster, one as an outlier and the other within cluster #3.

Insights into the Origin of Domesticated Sunflower

Our data revealed the presence of 3 haplotypes within the primitive domesticates but only 1 in all other domesticates. Although this finding is superficially consistent with the occurrence of independent origins, the 2 rare domesticated haplotypes seem more likely to be the result of other processes. For example, although the occurrence of a unique haplotype in the Hidatsa lineage could have resulted from a separate origin of domestication, previous researchers have found

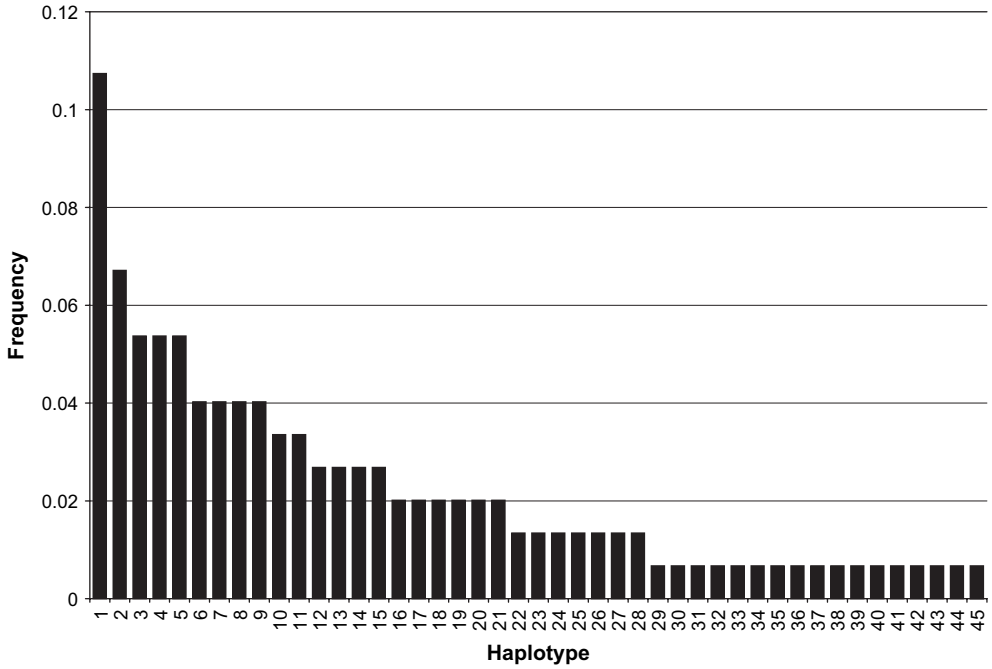

Figure 1. Frequency distribution of the 45 unique cpSSR haplotypes identified in wild *Helianthus annuus*.

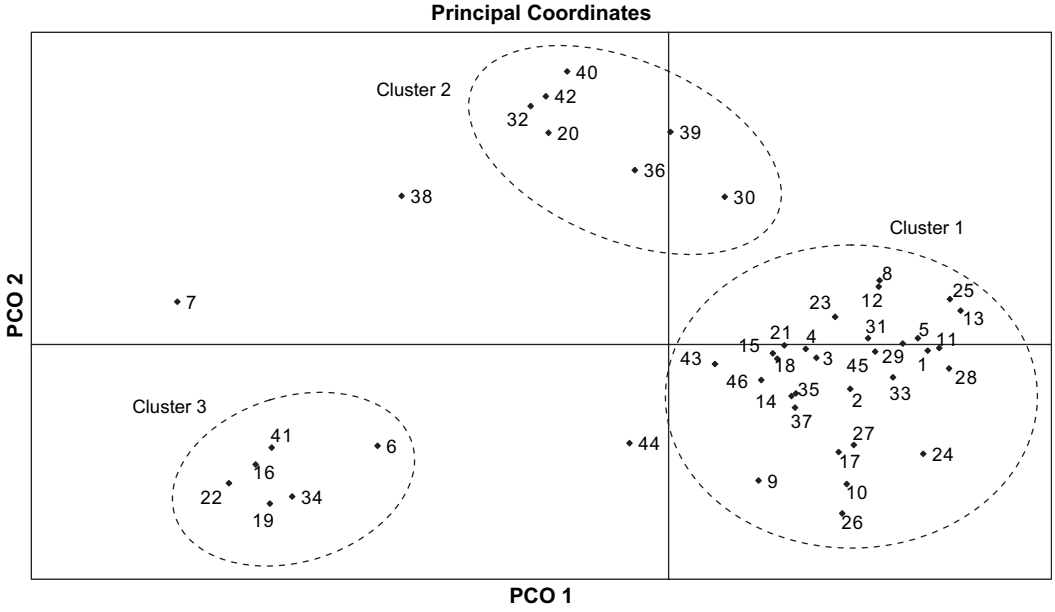

Figure 2. PCO analysis of the cpSSR haplotypes observed in *Helianthus annuus*.

no evidence (based on nuclear markers) to suggest that this landrace arose independently of the others and these plants come from North Dakota, which is geographically distant from the hypothesized other origin of domestication. Thus, it seems most likely that this haplotype is the result of mutation and subsequent fixation within the Hidatsa landrace or possibly introgression/chloroplast capture. Although this haplotype has not been found in the wild, it falls within PCO cluster #1 (Figure 2). In the case of the Maíz de Téjas individual that differs from the primary domesticated haplotype by a single base pair at 1 of the 6 loci, the most likely explanation seems to be that the haplotype carried by this individual arose as a result of a unique mutational event.

With regard to the hypothesis that the Hopi and Havasupai landraces trace to a second origin of domestication in Mexico, the relative rarity of the primary domesticated haplotype in the wild (ca. 6.7%) makes a second origin rather unlikely. Moreover, this haplotype was not found outside of the United States, and both of the Mexican haplotypes that we identified fell well outside of the PCO cluster that contains all the domesticated lines that we surveyed. Thus, although we cannot rule out the possibility of a second origin of domestication in Mexico (or elsewhere), the descendants of which ultimately went extinct, our data point to a single origin of the extant domesticated sunflowers somewhere outside of Mexico.

Acknowledgments

We thank Preston Aldrich, Mark Chapman, Aizhong Liu, Natasha Sherman, Jessica Wenzler, and an anonymous reviewer for comments on an earlier version of the manuscript. Tamara Berthel, Chris Buckner, Jonathan Ertelt, Natasha Sherman, and Jessica Wenzler provided assistance in the greenhouse and/or laboratory. This work was supported by grants to J.M.B. from the National Science Foundation Plant Genome Research Program (DBI-0332411) and the National Research Initiative of the United States Department of Agriculture Cooperative State Research, Education, and Extension Service (#03-35300-13104 and 03-39210-13958).

References

Badr A, Muller K, Schafer-Pregl R, El Rabey H, Effgen S, Ibrahim HH, Pozzi C, Rohde W, Salamini F. 2000. On the origin and domestication history of barley (*Hordeum vulgare*). Mol Biol Evol 17:499–510.

Brewer AJ. 1973. Analysis of floral remains from the Higgs site (40L045). In: McCullough MCR and Faulknor CJ, editors. Excavation of the Higgs and Doughty site. I-75 salvage archaeology. Tennessee Archaeological Society, Miscellaneous Papers No. 12. Knoxville, TN: University of Tennessee. p 141–7.

Bryan GJ, McNicoll J, Ramsay G, Meyer RC, DeJong WS. 1999. Polymorphic simple sequence repeat markers in chloroplast genomes of Solanaceous plants. Theor Appl Genet 99:859–67.

Crites GD. 1993. Domesticated sunflower in 5th millennium BP temporal context—new evidence from Middle Tennessee. Am Antiq 58:146–8.

Cronn R, Brothers M, Klier K, Bretting PK, Wendel JF. 1997. Allozyme variation in domesticated annual sunflower and its wild relatives. Theor Appl Genet 95:532–45.

Excoffier L, Laval G, Schneider S. 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50.

Ford RI. 1985. Some botanical considerations of the early domesticated plants north of Mexico. In: Ford RI, editor. Prehistoric food production in North America, Museum of Anthropology, Anthropological Paper Nr. 75. Ann Arbor, MI: University of Michigan. p 1–18.

Hancock JF. 2004. Plant evolution and the origin of crop species. Cambridge, MA: CABI Publishing.

Harter AV, Gardner KA, Falush D, Lentz DL, Bye RA, Rieseberg LH. 2004. Origin of extant domesticated sunflowers in eastern North America. Nature 430:201–5.

Heiser CB Jr. 1954. Variation and subspeciation in the common sunflower, *Helianthus annuus*. Am Midl Nat 51:287–305.

Heiser CB Jr. 1978. Taxonomy of *Helianthus* and origin of domesticated sunflower. In: Carter JF, editor. Sunflower science and technology. Madison, WI: American Society of Agronomy. p 31–53.

Heiser CB Jr., 1985. Some botanical considerations of the early domesticated plants north of Mexico. In: Ford RI, editor. Prehistoric food production in North America, Museum of Anthropology, Anthropological Paper Nr. 75. Ann Arbor, MI: University of Michigan. p 57–72.

Lentz DL, Pohl MED, Pope KO, Wyatt AR. 2001. Prehistoric sunflower (*Helianthus annuus* L.) domestication in Mexico. Econ Bot 55:370–6.

Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J. 2002. A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–4.

Peakall R, Smouse PE. 2006. Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–95.

Powell W, Morgante M, McDevitt R, Vendramin GG, Rafalski JA. 1995. Polymorphic simple sequence repeat regions in chloroplast genomes: applications to the population genetics of pines. Proc Natl Acad Sci USA 92:7759–63.

Putt ED. 1997. Early history of sunflower. In: Schneiter AA, editor. Sunflower production and technology. Madison, WI: American Society of Agronomy. p 1–19.

Rieseberg LH, Seiler GJ. 1990. Molecular evidence and the origin and development of the domesticated sunflower (*Helianthus annuus*, Asteraceae). Econ Bot 44(Suppl 3):79–91.

Schuelke M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nat Biotechnol 18:233–4.

Second G. 1982. Origin of the genic diversity of cultivated rice (*Oryza* spp.)—study of the polymorphism scored at 40 isoenzyme loci. Jpn J Genet 57: 25–57.

Seiler GJ, Rieseberg LH. 1997. Systematics, origin, and germplasm resources of the wild and domesticated sunflower. In: Schneiter AA, editor. Sunflower production and technology. Madison, WI: American Society of Agronomy. p 21–65.

Slatkin M. 1994. Linkage disequilibrium in growing and stable-populations. Genetics 137:331–6.

Spooner DM, McLean K, Ramsay G, Waugh R, Bryan GJ. 2005. A single domestication for potato based on multilocus amplified fragment length polymorphism genotyping. Proc Natl Acad Sci USA 102:14694–9.

Tang S, Knapp SJ. 2003. Microsatellites uncover extraordinary diversity in Native American land races and wild populations of cultivated sunflower. Theor Appl Genet 106:990–1003.

Weising K, Gardner RC. 1999. A set of conserved PCR primers for the analysis of simple sequence repeat polymorphisms in chloroplast genomes of dicotyledonous angiosperms. Genome 42:9–19.

Wendel JF, Schnabel A, Seelanan T. 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). Proc Natl Acad Sci USA 92:280–4.

Wills DM, Hester ML, Liu A, Burke JM. 2005. Chloroplast SSR polymorphisms in the Compositae and the mode of organellar inheritance in *Helianthus annuus*. Theor Appl Genet 110:941–7.

Xu DH, Abe J, Gai JY, Shimamoto Y. 2002. Diversity of chloroplast DNA SSRs in wild and cultivated soybeans: evidence for multiple origins of cultivated soybean. Theor Appl Genet 105:645–53.

Received January 9, 2006
Accepted April 10, 2006

Corresponding Editor: James Hamrick