

# Sunflower domestication alleles support single domestication center in eastern North America

Benjamin K. Blackman<sup>a,b,1</sup>, Moira Scascitelli<sup>c</sup>, Nolan C. Kane<sup>c</sup>, Harry H. Luton<sup>a</sup>, David A. Rasmussen<sup>a,b</sup>, Robert A. Bye<sup>d</sup>, David L. Lentz<sup>e</sup>, and Loren H. Rieseberg<sup>a,c</sup>

<sup>a</sup>Department of Biology, Indiana University, Bloomington, IN 47405; <sup>b</sup>Department of Biology, Duke University, Durham, NC 27701; <sup>c</sup>Department of Botany, University of British Columbia, Vancouver, BC, Canada V6R 2A5; <sup>d</sup>Jardín Botánico Exterior, Instituto de Biología, Universidad Nacional Autónoma de México, Mexico DF, 04510, Mexico; and <sup>e</sup>Department of Biological Sciences, University of Cincinnati, Cincinnati, OH 45221

Edited by John Doebley, University of Wisconsin, Madison, WI, and approved July 20, 2011 (received for review March 29, 2011)

Phylogenetic analyses of genes with demonstrated involvement in evolutionary transitions can be an important means of resolving conflicting hypotheses about evolutionary history or process. In sunflower, two genes have previously been shown to have experienced selective sweeps during its early domestication. In the present study, we identified a third candidate early domestication gene and conducted haplotype analyses of all three genes to address a recent, controversial hypothesis about the origin of cultivated sunflower. Although the scientific consensus had long been that sunflower was domesticated once in eastern North America, the discovery of pre-Columbian sunflower remains at archaeological sites in Mexico led to the proposal of a second domestication center in southern Mexico. Previous molecular studies with neutral markers were consistent with the former hypothesis. However, only two indigenous Mexican cultivars were included in these studies, and their provenance and genetic purity have been questioned. Therefore, we sequenced regions of the three candidate domestication genes containing SNPs diagnostic for domestication from large, newly collected samples of Mexican sunflower landraces and Mexican wild populations from a broad geographic range. The new germplasm also was genotyped for 12 microsatellite loci. Our evidence from multiple evolutionarily important loci and from neutral markers supports a single domestication event for extant cultivated sunflower in eastern North America.

agriculture | *Helianthus annuus* | phylogeography

Many genetic changes essential for crop evolution have been identified, and these alleles provide useful tools for inferring how, where, and when early societies transformed wild plants into agricultural staples (1, 2). Examining sequence diversity of these alleles in cultivated lineages, wild progenitors, and archaeological specimens can reveal critical information about the rate, timing, geography, and order of the domestication process (3). For instance, sequences of domestication loci obtained from archaeological maize cobs indicated that although domestication alleles for two traits became fixed early, fixation of domestication alleles for a third trait occurred several thousand years later (4). Patterns of sequence diversity around domestication alleles have yielded estimates of the strength of selection during domestication for several loci in rice and maize (5–7). In addition, the geographic distributions of domestication alleles in extant cultivated and wild germplasm have been used to assess how crops and crop alleles have spread from domestication centers and reveal whether convergent traits in independent lineages evolved from the same or unique suites of mutations (5, 8–11).

Here, we examine three genes that experienced selective sweeps during early domestication of sunflower, *Helianthus annuus*, to address a recent controversy about the number and location of sunflower domestication centers (12–17). Molecular, archaeological, and linguistic evidence has long supported the presence of a domestication center in eastern North America (ENA) (16). However, pre-Columbian archaeological finds at several sites in southern Mexico have suggested the possibility of a second independent domestication center and challenged the view that domesticated sun-

flower reached Mexico only after introduction by Spanish traders (14, 15, 17, 18). The archaeological evidence suggests two possible alternative explanations: an independent origin of domesticated sunflower in Mexico or dispersal to Mexico through trade routes established before Spanish colonization. The identification and significance of the archaeological remains, the historic record, and modern linguistic evidence have been debated heavily (16, 17, 19–26). Previous molecular studies using neutral nuclear and chloroplast DNA markers have indicated that extant Mexican landraces are descended from the same lineage as ENA landraces, suggesting that if an independent Mexican lineage evolved, it did not contribute to modern cultivated sunflower germplasm (27, 28). However, the provenance and genetic purity of the two Mexican landrace samples included in those studies have been questioned. In addition, few Mexican wild *H. annuus* populations were included in these analyses.

We have made new, extensive collections of Mexican wild sunflowers and domesticated cultivars grown by indigenous Mexican populations. Sequencing genes demonstrated to have experienced selective sweeps during domestication in ENA provided an opportunity to test whether extant Mexican landraces descended from the same or a unique lineage. In particular, if domesticated sunflower reached Mexico through trade routes from the North, we predicted that ENA and Mexican landraces would share the same haplotypes at these genes. If two independent domestication events occurred in isolation, it is highly unlikely these parallel events would have selected for the same causative mutations arising on the same haplotype backgrounds at multiple loci (8–10). In addition, the presence of the domesticated haplotype in ENA wild populations and its absence from wild Mexican populations would exclude the possibility of a parallel selective sweep from equivalent standing variation.

Two strong candidate early domestication genes in sunflower have been described. A genomic screen found that sequence diversity in the gene *c4973*, a chorismate synthase homolog involved in aromatic amino acid synthesis, is significantly reduced in North American landraces relative to its sequence diversity in neutral loci (29). *HaFT1*, a homolog of the floral inducer *FLOWERING LOCUS T*, shows a similar signature of a selective sweep during early domestication (30). Although this evidence indicates that the genomic regions containing these genes reflect a history of recent selection, we describe these genes as candidate domestication genes because additional work is required to distinguish whether they or other tightly linked loci carry the specific muta-

Author contributions: B.K.B., M.S., R.A.B., D.L.L., and L.H.R. designed research; B.K.B., M.S., N.C.K., H.H.L., D.A.R., R.A.B., and D.L.L. performed research; B.K.B. and M.S. analyzed data; and B.K.B., M.S., and D.L.L. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. JF428180–JF428779).

<sup>1</sup>To whom correspondence should be addressed. E-mail: bkb7@duke.edu.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1104853108/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1104853108/-DCSupplemental).

**Table 1. Domesticated and wild allele frequencies in eastern North American landraces (ENA domesticated), American and Canadian wild populations (ENA wild), Mexican landraces (MX domesticated), and Mexican wild populations (MX wild)**

Gene	ENA domesticated	ENA wild	MX domesticated	MX wild
<i>HaFT1</i>				
Domesticated allele	95% (36)	5% (2)	98% (41)	0% (0)
Wild allele	5% (2)	95% (38)	2% (1)	100% (152)
	(38 total)	(40 total)	(42 total)	(152 total)
<i>HaGA2ox</i>				
Domesticated haplotype 1	95% (36)	7% (3)	100% (48)	0% (0)
Other domesticated haplotypes	5% (2)	2% (1)	0% (0)	0% (0)
Wild haplotypes	0% (0)	91% (42)	0% (0)	100% (152)
	(38 total)	(46 total)	(48 total)	(152 total)
<i>c4973</i>				
Domesticated haplotype 1	100% (44)	29% (4)	76% (47)	7% (8)
Domesticated haplotype 2	0% (0)	7% (1)	18% (11)	4% (5)
Wild haplotype 1	0% (0)	43% (6)	5% (3)	20% (23)
Other wild haplotypes	0% (0)	21% (3)	1% (1)	69% (79)
	(44 total)	(14 total)	(62 total)	(115 total)

The number of haplotype sequences per allele class is shown in parentheses.

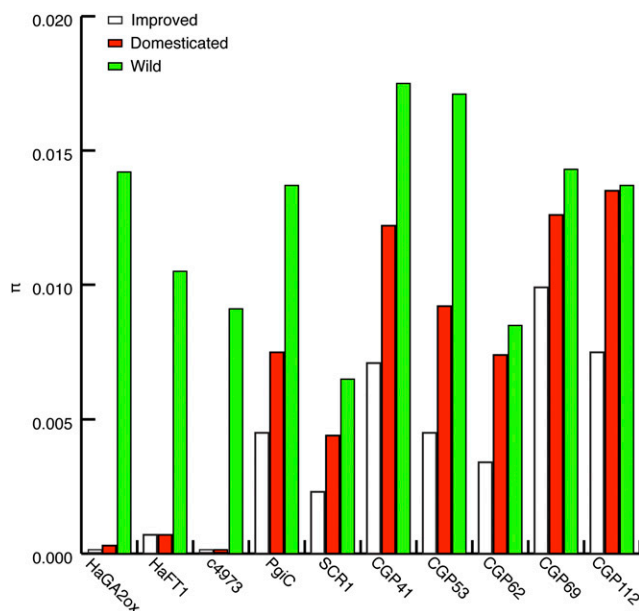
tions responsible for domestication traits and targeted by selection. Notably, however, several lines of genetic and functional evidence have provided strong support for a frameshift mutation in the domesticated allele of *HaFT1* as the likely causative mutation that responded to selection through its effects on flowering time or head diameter (30). Here, we identify a third candidate early domestication gene and show that patterns of diversity in these genes, as well as 12 neutral markers, in Mexican landraces and wild *H. annuus* are consistent with all extant landraces originating from a single domestication center in eastern North America.

## Results

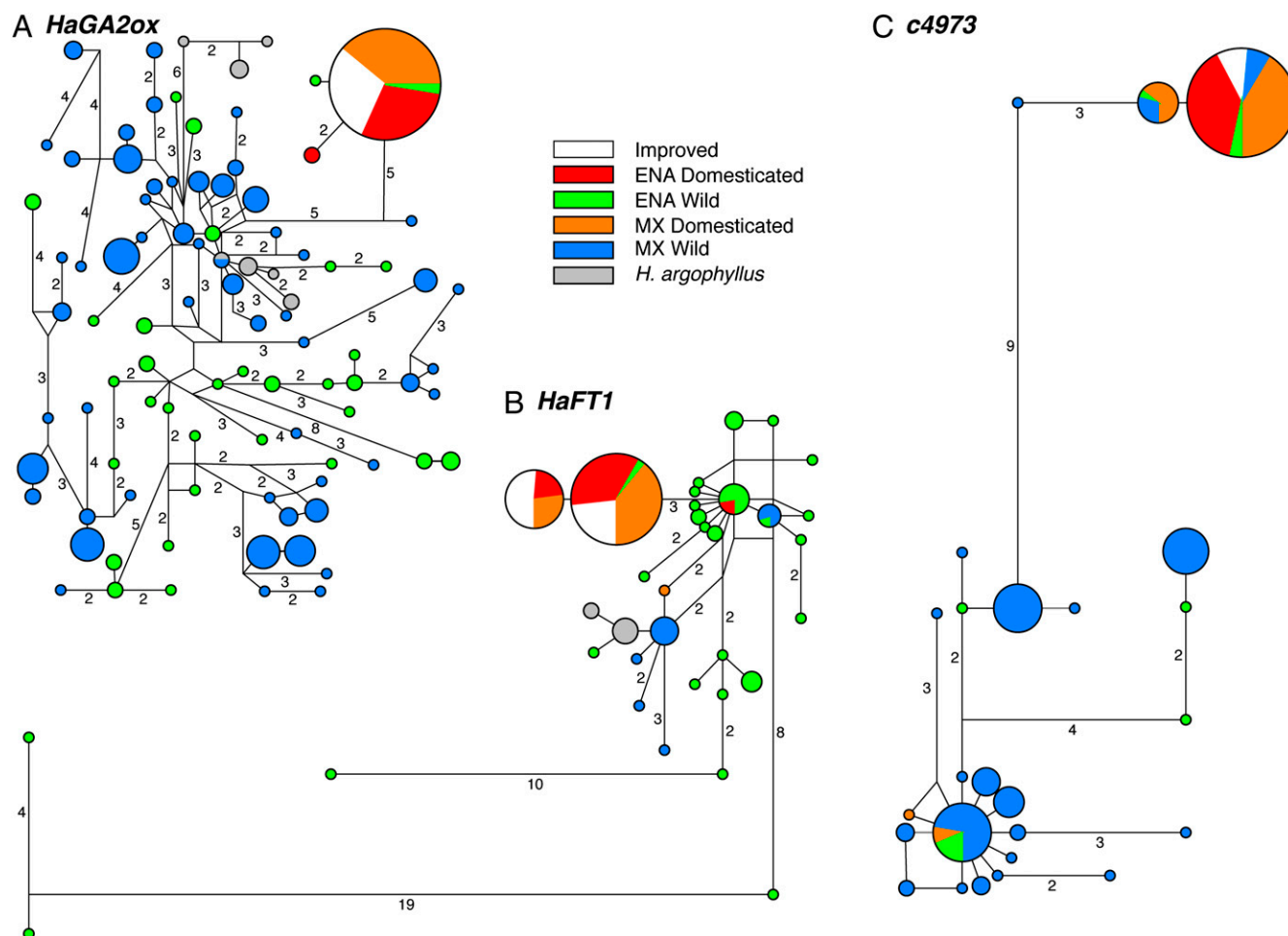
**Selection on Gibberellin 2-Oxidase Homolog *HaGA2ox* During Early Domestication.** An 885-bp portion of the gibberellin 2-oxidase homolog *HaGA2ox* was examined as part of an ongoing study of

a genetic diversity panel from the United States and Canada (31). This panel included 18 individuals from improved modern crop lines, 19 individuals from Native American landraces, and 23 individuals from a geographically diverse sample of wild sunflower populations (Table S1). Six *Helianthus argophyllus* individuals also were sequenced to obtain divergence estimates. Average pairwise nucleotide diversity ( $\pi$ ) in *HaGA2ox* was much reduced in domesticated and improved lines relative to neutral loci, similar to previous observations for *HaFT1* and *c4973* (Fig. 1). Only two haplotypes—distinguished from each other by two SNPs—segregated in Native American landraces (Fig. 2A). Improved lines were fixed for the major landrace haplotype, and this haplotype was present only at minor frequency in wild populations, indicative of a selective sweep (Table 1). All coding substitutions along the branch distinguishing the domesticated and wild haplotypes were synonymous changes. A maximum-likelihood Hudson–Kreitman–Aguadé (HKA) test, which compares the likelihood of a neutral evolutionary model of polymorphism and divergence with the likelihood of a model in which a candidate gene is under selection, was conducted for *HaGA2ox* and seven previously sequenced putative neutral loci (31, 32). Although *HaGA2ox* was found to be evolving neutrally in wild populations ( $P = 0.9748$ ), there were significant signatures of a selective sweep in the landrace ( $P = 0.0032$ ) and improved pools ( $P = 0.0002$ ).

**Description of Collections.** Mexican landraces and wild populations used in this study were collected between 2001 and 2005 (Table S2). Wild specimens were obtained from habitats in all states of Mexico north of Veracruz. Detailed collection procedures and complete morphological descriptions of Mexican landrace achenes can be found in ref. 33. In addition to the genetic affinities of sunflower, we were interested in learning more about how this diverse crop was used by Native Americans in Mexico. Accordingly, we interviewed numerous representatives of indigenous peoples in their homes and fields. Cultivars used in this study were collected from the Tarahumara in Chihuahua, the Nahuia in Guerrero, the Tzeltal in Chiapas, and the Mixe in Oaxaca. One domesticated collection was obtained from Mennonites living in Chihuahua who had emigrated recently from Canada and likely brought their sunflower germplasm with them.



**Fig. 1.** Average pairwise nucleotide diversity of three candidate domestication genes (*HaFT1*, *HaGA2ox*, and *c4973*) is significantly reduced relative to seven neutral loci during early domestication.



**Fig. 2.** Median-joining haplotype networks for three candidate early domestication genes, *HaGA2ox* (A), *HaFT1* (B), and *c4973* (C). ENA domesticated and ENA wild are the ENA landraces and American and Canadian wild populations, respectively. MX domesticated and MX wild are the Mexican landraces and Mexican wild populations, respectively. The number of substitutions along a branch is shown for branches with more than one substitution.

#### Domestication Alleles in Mexican Landraces and Wild Populations.

We next examined sequence diversity of *HaGA2ox*, *HaFT1*, and *c4973* in 21–31 individuals from the 16 Mexican landrace collections and 57–76 individuals from 15 geographically diverse Mexican wild populations (Table S2 and Fig. 3). For *HaGA2ox* and *c4973*, we fully sequenced the same portions previously sequenced from ENA accessions. Although we sequenced the *HaFT1* haplotype from all Mexican landraces, we genotyped the *HaFT1* allele in wild populations because the putative causal frameshift mutation in the domesticated allele of *HaFT1* creates an easily diagnosed restriction site polymorphism.

As would be expected if the extant ENA and Mexican landraces are descended from a single domestication event, the Mexican landraces were fixed for the same major domesticated haplotype for *HaGA2ox* (Fig. 2A and Table 1). Although present at minor frequency in ENA wild populations, this haplotype was not observed in wild Mexican populations.

Nearly the same results were observed for *HaFT1*. All but one haplotype sequenced from Mexican landraces contained the frameshift allele (Fig. 2B and Fig. S1). Although the frameshift-containing domesticated allele does segregate at minor frequency in wild populations from Oklahoma, this allele was not found in any Mexican wild individuals, an observation inconsistent with a parallel sweep from standing variation in Mexico (Table 1). We predicted that the single Mexican wild haplotype present in the Mexican landraces is the result of a recent hybridization event. If

true, then nearby genes also should segregate for one American domesticated-like allele and one Mexican wild-like allele, because adequate time will not have passed to break down linkage disequilibrium between alleles in closely linked genes. We sequenced portions of *HaFT2* and *HaFT3*, which are within 2.7 cM of *HaFT1* (30), from this individual and observed this predicted pattern. For both genes, one allele exactly matched the major al-



**Fig. 3.** Collection locations of sunflower domesticates (diamonds) and wild populations (circles) in Mexico.

lele in improved lines and ENA landraces. The other allele at each locus was similar to alleles in ENA wild populations but contained one or more unique SNPs, likely indicative of Mexican derivation.

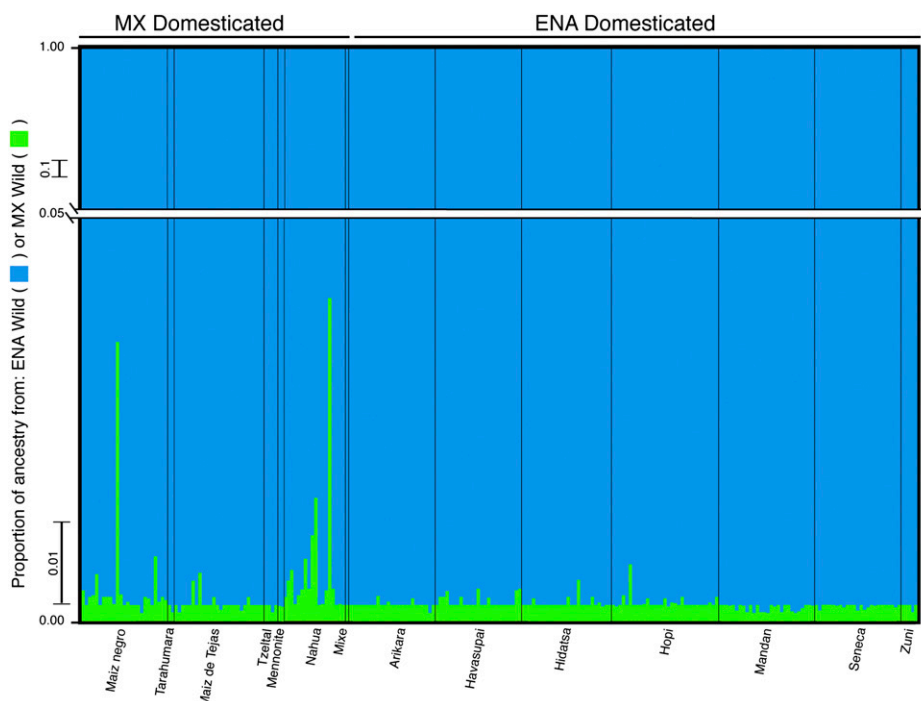
Although the results of *HaFT1* and *HaGA2ox* sequencing were completely consistent with a single origin of extant domesticated sunflower, the results for *c4973* did not distinguish unambiguously between the single- and multiple-origin scenarios. As for the other two genes, the haplotype distribution is disjunct, with most wild sequences forming one cluster and most domesticated sequences forming a second cluster (Fig. 2C). Two haplotypes, which differ from each other by only a single SNP, were at high frequency in Mexican landraces. One of these two haplotypes was fixed in our sample that included previously collected data from a small set of ENA domesticates (29) and additional sequencing we conducted on the landraces in our own diversity panel (Fig. 2C and Table 1). Wild haplotypes also were observed to segregate at minor frequency in the Mexican landraces, possibly indicating recent hybridization as observed for *HaFT1*. Notably, both domesticated haplotypes were found in wild populations from both regions, meaning a parallel selective sweep from standing variation in Mexico cannot be excluded. The domesticated alleles were present at higher frequency in American wild populations, however (Table 1), and these alleles were found in populations from the eastern and northern United States. Thus, ENA is more likely to be the geographic source of these domesticated alleles.

**Ancestry Assignment with Neutral Markers.** Under certain evolutionary scenarios, incongruities may arise between domestication gene trees and lineage histories because a domestication allele that contributes to desirable phenotypes may be bred from one lineage into an independent lineage through hybridization and selective breeding (5, 6, 11). Therefore, to determine ancestry and corroborate our findings with neutral markers, we genotyped our new, more thorough collections from Mexico for 12 microsatellite loci that had been scored previously on a large set of ENA cultivars and wild populations (27). As expected, domes-

ticated lines contained a subset of the diversity present in wild populations, with the exception of two cultivar-specific alleles present in both ENA and Mexican domesticates (Table S3). In addition, although several alleles present in ENA wild populations but not in Mexican wild populations segregated in domesticates from either or both regions, only one allele observed in one Mexican domesticated individual was found solely in wild individuals from Mexico (Table S3). The estimation of the ancestral provenance of ENA and Mexican landrace samples—as performed in the program Structure through analysis of multi-locus genotypes (34)—yielded a clear, unequivocal result. All Mexican and ENA cultivars clustered with the ENA wild samples and not with the Mexican wild samples (Fig. 4), consistent with the results from the three selected loci. For the Mexican landraces, the proportion of genotype with ancestry assigned to the ENA-wild group was >0.96 in all cases.

## Discussion

Taken together, the findings from three candidate domestication loci and additional neutral markers indicate that domesticated sunflowers grown in Mexico today are descended from the same cultivated genetic lineage as ENA domesticated sunflower. All ENA and Mexican landraces shared the same divergent *HaGA2ox* haplotypes that, although observed in ENA wild populations, never were present in Mexican wild populations. Although the signature of a selective sweep on *HaGA2ox* could also be explained by selection acting on a closely linked locus, this enzyme is an attractive candidate for the actual target of selection. Gibberellin 2-oxidases convert gibberellins from active to inactive forms and in doing so act to regulate seed development, delay germination, and promote dwarfism (35–38). Reduced seed dormancy and increased plant biomass have been selected for strongly during domestication of many crops (39), and this pressure may have driven evolutionary changes affecting *HaGA2ox* expression or



**Fig. 4.** Estimated ancestry of domesticated *H. annuus* from Mexico (MX domesticated) and North America (ENA domesticated). Each vertical bar represents an individual's genome and is colored in proportion to the estimated ancestry from MX wild (green) or ENA wild (blue). Vertical black lines delimit individuals into groups by landrace, as reported below each group.



activity. We are planning future genetic studies to look at the functional evolution of this gene during sunflower domestication.

*HaFT1* exhibited the same pattern as *HaGA2ox*, except for a single Mexican cultivar that was heterozygous for a Mexican wild allele because of a recent hybridization event. Since *c4973* domesticated alleles segregate in both ENA and Mexican wild populations, the possibility of a parallel selective sweep on the same allele cannot be excluded. Nevertheless, the higher frequency of domesticated alleles in ENA wild populations suggest that the eastern United States is the more likely geographic source. The low frequency of these alleles in Mexican wild populations and the low frequency of wild alleles in Mexican landraces may have resulted from recent hybridization, as seen with *HaFT1*. An intriguing alternative explanation for the latter pattern may be that the selective sweep at *c4973* was incomplete when cultivated sunflower was dispersed southward to Mexico. If true, then this result would further support an early introduction of cultivated sunflower to Mexico and indicates that, as more domestication alleles and their functions are identified, it may be possible to time selective events as pre- and post-dispersal, allowing trait evolution to be ordered relative to this chronological benchmark. Finally, the ambiguous findings with *c4973* could also have resulted from altered selection pressures on the domesticated haplotype in Mexico or because *c4973* is not an actual domestication gene but is tightly linked to one.

Although these three genes and neutral markers exhibit patterns of variation in concordance with a single-origin scenario for extant cultivated sunflower, the possibility of multiple origins cannot be discounted fully. For instance, despite thorough searching (by R.A.B. and D.L.L.), it is possible that we were unable to locate modern domesticates descended from an independent Mexican lineage. Also, although the data presented herein are more consistent with a single origin and dispersal before the arrival of Spanish explorers (40), our findings do not invalidate the possibility that a separate center of sunflower domestication in Mexico did indeed occur during the pre-Columbian period. An ancient lineage may not be represented in modern germplasm because no descendants of the hypothesized independent Mexican lineage have survived. Although current evidence from archaeological deposits in ENA and Mesoamerica indicate that ancient Mesoamericans cultivated sunflower before the advent of the Spanish in the region, more discoveries will be needed before a full understanding of the trajectory of sunflower crop development in Mesoamerica and ENA emerges.

Lack of modern descendants from a possible independent lineage raises the question of how descendants of that lineage may have gone extinct. It has been suggested that colonial Spanish Christians may have tried to eliminate sunflower because of its important symbolic role in the religious rituals of the solar-worshipping Aztecs and other Native American cultures (17). Alternatively, native domesticates in Mexico may have been replaced much more recently by a substantial influx of seed imports made possible by the North American Free Trade Agreement.

Although the majority of the Mexican archaeological sunflower discoveries were destroyed for radiocarbon dating, archaeological studies at sites throughout Mexico likely will yield additional samples. Furthermore, new sequencing technologies have made ancient DNA studies increasingly successful means of revealing

genealogical and phenotypic characteristics of extinct lineages (4, 8, 41–46). Studies of these and newly identified domestication genes in DNA isolated from archaeological plant remains discovered in ENA and Mexico promise to yield even further insight into the timing and geography of sunflower domestication.

## Materials and Methods

DNA was extracted from leaf tissue using DNeasy 96 Plant or Plant Mini Kits (Qiagen). Portions of *HaFT1*, *HaGA2ox*, and *c4973* were amplified by PCR and sequenced using gene-specific primers (Table S4) from Mexican cultivars and wild Mexican individuals (Table S2). *HaGA2ox* also was sequenced on the entire American and Canadian diversity panel (Table S1). We included previously published *HaFT1* (30) and *c4973* (29) sequences from American and Canadian *H. annuus* domesticated and wild populations and *H. argophyllus* (Table S5). Because *c4973* had been sequenced previously only on a small sample of American landraces, we obtained additional sequences from our diversity panel to sample *c4973* sequence diversity more thoroughly in North American landraces (Tables S1 and S5). Sequence data for seven putative neutral loci were obtained for a previously published study of this diversity panel (Table S5) (31). *HaFT1* PCR products were digested with BstEII to genotype individuals for the frameshift mutation nearly fixed in domesticated haplotypes.

For heterozygous individuals, PCR products were cloned (TOPO TA Cloning; Invitrogen), and multiple clones per individual were sequenced using T7/T3 or M13F/M13R primers. Cloned sequences were compared with each other and with sequencing reads from the initial PCR product to ensure that both haplotypes were captured and to eliminate sequence errors introduced by PCR and cloning. Median-joining haplotype networks were constructed with Network 4.5.16 (Fluxus Technology). For all genes, insertion/deletion polymorphisms were recoded as SNPs, and a complex intronic region of nested insertion/deletion polymorphisms was excluded from analysis of *HaGA2ox*.

Average pairwise nucleotide diversity ( $\pi$ ) was calculated using DnaSP (47). The program MLHKA (32) was used to conduct the maximum-likelihood HKA tests for *HaGA2ox*. A likelihood-ratio test was conducted to compare the likelihood of a neutral model and the likelihood of a model in which *HaGA2ox* was under selection. Separate tests were performed for improved, domesticated American landraces and American wild samples to determine the timing of selection.

Landraces and wild population samples from Mexico and ENA were genotyped with 12 microsatellite loci (Tables S1, S2, and S4) that had been scored previously on another panel of domesticated and wild *H. annuus* (27). PCR amplification, fragment separation by capillary electrophoresis, and fragment length scoring were conducted as previously described (48). Data were combined with genotypes from ref. 27 using one to five overlapping samples to assure that allele assignments in each data set were congruent. The ancestry of *H. annuus* cultivated individuals and the composition of their genotype was evaluated with a Bayesian clustering method, as implemented in Structure v. 2.3.3 (34, 49). Two ancestral source clusters were identified by ref. 27: (i) a group including all Mexican and Arizonan wild populations (MX Wild) and (ii) a group of central United States populations (ENA Wild). Ten independent simulations were performed using the same parameters and prior assumptions as in ref. 27. Results were formatted with HARVESTER v. 0.6.6 (<http://taylor0.biology.ucla.edu/structureHarvester/>) and permuted with CLUMPP v. 1.1.2 (50). The final matrix of estimated ancestry coefficients was visualized with DISTRUCT v. 1.1 (51).

**ACKNOWLEDGMENTS.** We thank A. Harter for generously providing data from her previous work and for providing guidance with the neutral marker analysis. We thank Z. Lai, Q. Yu, A. Raduski, J. Strasburg, M. Kreitzman, L. Washington, and M. Chapman for technical assistance and scientific input. This work was supported by National Science Foundation Grants DEB0608118 and DBI0905958 (to B.K.B.), DBI0421630 and DBI0820451 (to L.H.R.), and BCS0228049 (to R.A.B., D.L.L., and L.H.R.) and by National Geographic Society Grant 7030-01 (to D.L.L. and R.A.B.).

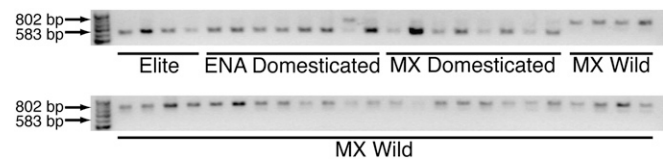
1. Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. *Cell* 127:1309–1321.
2. Purugganan MD, Fuller DQ (2009) The nature of selection during plant domestication. *Nature* 457:843–848.
3. Gross BL, Olsen KM (2010) Genetic perspectives on crop domestication. *Trends Plant Sci* 15:529–537.
4. Jaenicke-Després V, et al. (2003) Early allelic selection in maize as revealed by ancient DNA. *Science* 302:1206–1208.

5. Zhang L-B, et al. (2009) Selection on grain shattering genes and rates of rice domestication. *New Phytol* 184:708–720.
6. Olsen KM, et al. (2006) Selection under domestication: Evidence for a sweep in the rice waxy genomic region. *Genetics* 173:975–983.
7. Wang H, et al. (2005) The origin of the naked grains of maize. *Nature* 436:714–719.
8. Palmer SA, Moore JD, Clapham AJ, Rose P, Allaby RG (2009) Archaeogenetic evidence of ancient nubian barley evolution from six to two-row indicates local adaptation. *PLoS ONE* 4:e6301.

9. Azhaguel P, Komatsuda T (2007) A phylogenetic analysis based on nucleotide sequence of a marker linked to the brittle rachis locus indicates a diphyletic origin of barley. *Ann Bot (Lond)* 100:1009–1015.
10. Komatsuda T, et al. (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc Natl Acad Sci USA* 104:1424–1429.
11. Sweeney MT, et al. (2007) Global dissemination of a single mutation conferring white pericarp in rice. *PLoS Genet* 3:e133.
12. Heiser CB (1951) The sunflower among the North American Indians. *Proc Am Philos Soc* 95:432–448.
13. Heiser CB (1998) The domesticated sunflower in old Mexico? *Genet Resour Crop Evol* 45:447–449.
14. Pope KO, et al. (2001) Origin and environmental setting of ancient agriculture in the lowlands of Mesoamerica. *Science* 292:1370–1373.
15. Lentz DL, Pohl MED, Pope KO, Wyatt AR (2001) Prehistoric sunflower (*Helianthus annuus* L.) domestication in Mexico. *Econ Bot* 55:370–376.
16. Smith BD (2006) Eastern North America as an independent center of plant domestication. *Proc Natl Acad Sci USA* 103:12223–12228.
17. Lentz DL, Pohl MD, Alvarado JL, Tarighat S, Bye R (2008) Sunflower (*Helianthus annuus* L.) as a pre-Columbian domesticate in Mexico. *Proc Natl Acad Sci USA* 105:6232–6237.
18. Bye R, Linares E, Lentz DL (2009) México: Centro de Origen de la Domesticación del Girasol. *TIP Revista Especializada en Ciencias Químico-Biológicas* 12:5–12.
19. Smith BD (2008) Winnowing the archaeological evidence for domesticated sunflower in pre-Columbian Mesoamerica. *Proc Natl Acad Sci USA* 105:E45, author reply E49–E50.
20. Rieseberg L, Burke JM (2008) Molecular evidence and the origin of the domesticated sunflower. *Proc Natl Acad Sci USA* 105:E46; author reply E49–E50.
21. Brown CH (2008) A lack of linguistic evidence for domesticated sunflower in pre-Columbian Mesoamerica. *Proc Natl Acad Sci USA* 105:E47; author reply E49–E50.
22. Heiser CB (2008) How old is the sunflower in Mexico? *Proc Natl Acad Sci USA* 105:E48, author reply E49–E50.
23. Lentz DL, Pohl MD, Bye R (2008) Reply to Rieseberg and Burke, Heiser, Brown, and Smith: Molecular, linguistic, and archaeological evidence for domesticated sunflower in pre-Columbian Mesoamerica. *Proc Natl Acad Sci USA* 105:E49–E50.
24. Heiser CB (2008) The sunflower (*Helianthus annuus*) in Mexico: Further evidence for a North American domestication. *Genet Resour Crop Evol* 55:9–13.
25. Heiser CB (2008) Comment: Sunflowers among Aztecs? *Int J Plant Sci* 169:980.
26. Lentz DL (2008) Reply to Heiser. *Int J Plant Sci* 169:980.
27. Harter AV, et al. (2004) Origin of extant domesticated sunflowers in eastern North America. *Nature* 430:201–205.
28. Wills DM, Burke JM (2006) Chloroplast DNA variation confirms a single origin of domesticated sunflower (*Helianthus annuus* L.). *J Hered* 97:403–408.
29. Chapman MA, et al. (2008) A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*). *Plant Cell* 20:2931–2945.
30. Blackman BK, Strasburg JL, Raduski AR, Michaels SD, Rieseberg LH (2010) The role of recently derived FT paralogs in sunflower domestication. *Curr Biol* 20:629–635.
31. Blackman BK, et al. (2011) Contributions of flowering time genes to sunflower domestication and improvement. *Genetics* 187:271–287.
32. Wright SI, Charlesworth B (2004) The HKA test revisited: A maximum-likelihood-ratio test of the standard neutral model. *Genetics* 168:1071–1076.
33. Lentz DL, Bye R, Sanchez-Cordero V (2008) Ecological niche modeling and distribution of wild sunflower (*Helianthus annuus* L.) in Mexico. *Int J Plant Sci* 169:541–549.
34. Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
35. Oh E, et al. (2006) Light activates the degradation of PIL5 protein to promote seed germination through gibberellin in *Arabidopsis*. *Plant J* 47:124–139.
36. Thomas SG, Phillips AL, Hedden P (1999) Molecular cloning and functional expression of gibberellin 2-oxidases, multifunctional enzymes involved in gibberellin deactivation. *Proc Natl Acad Sci USA* 96:4698–4703.
37. Schomburg FM, Bizzell CM, Lee DJ, Zeevaert JAD, Amasino RM (2003) Overexpression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants. *Plant Cell* 15:151–163.
38. Singh DP, Jermakow AM, Swain SM (2002) Gibberellins are required for seed development and pollen tube growth in *Arabidopsis*. *Plant Cell* 14:3133–3147.
39. Hammer K (1984) Das Domestikationssyndrom. *Genet Resour Crop Evol* 32:11–34.
40. Smith BD, Yarnell RA (2009) Initial formation of an indigenous crop complex in eastern North America at 3800 B.P. *Proc Natl Acad Sci USA* 106:6561–6566.
41. Erickson DL, Smith BD, Clarke AC, Sandweiss DH, Tuross N (2005) An Asian origin for a 10,000-year-old domesticated plant in the Americas. *Proc Natl Acad Sci USA* 102:18315–18320.
42. Lalueza-Fox C, et al. (2007) A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. *Science* 318:1453–1455.
43. Miller W, et al. (2008) Sequencing the nuclear genome of the extinct woolly mammoth. *Nature* 456:387–390.
44. Rasmussen M, et al. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463:757–762.
45. Larson G, et al. (2010) Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proc Natl Acad Sci USA* 107:7686–7691.
46. Green RE, et al. (2010) A draft sequence of the Neandertal genome. *Science* 328:710–722.
47. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
48. Heesacker A, et al. (2008) SSRs and INDELs mined from the sunflower EST database: Abundance, polymorphisms, and cross-taxa utility. *Theor Appl Genet* 117:1021–1029.
49. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164:1567–1587.
50. Jakobsson M, Rosenberg NA (2007) CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23:1801–1806.
51. Rosenberg NA (2004) Distruct: aA program for the graphical display of population structure. *Mol Ecol Notes* 4:137–138.

# Supporting Information

Blackman et al. 10.1073/pnas.1104853108



**Fig. S1.** Sample of restriction digest genotyping of *FLOWERING LOCUS T* (FLW) PCR products. The frameshift mutation in the domesticated allele introduces a BstEII restriction site. No Mexican (MX) wild populations contained the frameshift, and their undigested PCR products ran at 802 bp. All improved, elite crop lines, all MX early domesticates, and all but one of the eastern North American (ENA) early domesticates included in this sample were homozygous for the domesticated allele (longer digestion product at 583 bp). One ENA early domesticate (seventh from left) shown is heterozygous for wild and domesticated alleles. All MX wild individuals were genotyped by restriction digest. All MX landrace genotypes were confirmed additionally by cloning and sequencing of PCR products.

**Table S1. American and Canadian diversity panel**

Species	Type	GRIN ID	Line name or population location
<i>Helianthus argophyllus</i>	Wild	PI 494571	Corpus Christi, TX
		PI 494572	North Padre Island, TX
		PI 494573	Port Aransas, TX
		PI 494580	Rachal, TX
		PI 494576	Skidmore, TX
		PI 494582	Victoria, TX
<i>Helianthus annuus</i>	Improved	PI 552943	RHA280
		PI 599984	HA821
		PI 534655	HA369
		PI 599768	RHA801
		PI 600000	RHA417
		PI 552937	HA292
		PI 578872	HA383
		PI 578873	HA384
		PI 607505	HA414
		PI 633744	HA434
		PI 599773	HA89
		PI 599759	RHA274
		PI 560141	RHA373
		PI 560145	RHA377
	Domesticated	PI 633746	RHA436
		PI 633748	RHA438
		PI 340790	USSR VNIIMK 8931 '66
		PI 650650	Ames7574, Mennonite
		PI 369357	Arikara
		PI 369360	Seneca
		PI 600717	Mandan #1
		PI 600718	Mandan #2
		PI 600719	Mandan #3
		PI 600720	Hidatsa #1
		PI 600721	Hidatsa #2
		PI 432504	Hopi dye
		PI 432505	Hopi
		PI 432507	Hopi dye
		PI 432508	Hopi dye
		PI 432509	Hopi dye
		PI 369358	Havasupai
		PI 369359	Hopi
		PI 432510	Hopi dye, possible hybrid w/Mammoth
		PI 432515	Zuni
		PI 432516	Pueblo

Table S1. Cont.

Species	Type	GRIN ID	Line name or population location
		PI 432521	Anzac Pueblo
		PI 432522	Laguna Pueblo
	Wild	PI 659440	Ann1238, Cedar Point Biological Station, NE
		PI 613750	Dickinson, ND
		PI 592325	Carievale, Saskatchewan
		PI 592316	Keeler, Saskatchewan
		PI 435434	Riviera, TX
		PI 435619	Tulsa, OK
		PI 494567	Skidmore, TX, Ann1811
		PI 468439	Colfax, ND
		PI 586879	Norden, NE
		PI 586872	Axtell, NE
		PI 435616	Topaz, MO
		PI 586869	Silver Creek, NE
		PI 586856	Great Bend, KS
		PI 586849	Colby, KS
		PI 613751	Minot, ND
		PI 613723	Crete, ND
		PI 613722	Onida, SD
		PI 613720	Garden City, KS
		PI 613711	Woonsocket, SD
		PI 592326	Boissevain, Manitoba
		PI 435505	McLoud, OK
		PI 468475	Childress, TX
		PI 597890	Yankton, SD

GRIN ID, US Department of Agriculture Germplasm Resources Information Network (<http://www.ars-grin.gov/npgs/>) identification number.



Type	N	Population ID number	Culture or line	Location	Latitude	Longitude
Landrace	2	33655	Tarahumara	Chihuahua	26.84509	−106.93494
	2	34177	Nahua	Guerrero	17.92778	−99.52629
	2	34111	Nahua	Guerrero	17.92372	−99.51938
	2	34285	Tzeltal	Chiapas	16.90928	−92.09041
	2	34055	Nahua	Guerrero	17.92528	−99.52958
	2	34112	Nahua	Guerrero	17.92372	−99.51938
	2	34282	Tzeltal	Chiapas	16.90928	−92.09041
	1	34324	Mixe	Oaxaca	16.9601	−95.84218
	2	Ames 6859/PI 650646	Maiz de Tejas			
	2	34120	Nahua	Guerrero	17.9248	−99.33792
	2	34110	Nahua	Guerrero	17.92372	−99.51938
	2	34102	Nahua	Guerrero	17.92098	−99.52727
	2	34109	Nahua	Guerrero	17.92372	−99.51938
	2	33724	Mennonite	Chihuahua	28.3095	−106.10909
	2	34107	Nahua	Guerrero	17.9236	−99.53328
	2	Ames 19070/PI 650761	Maiz negro			
Wild	9	28280		Tamaulipas	22.86712	−99.10377
	6	28321		Nuevo Leon	24.11155	−99.82345
	11	28353		Zacatecas	22.78198	−103.5582
	7	28315		San Luis Potosi	22.30028	−100.837
	6	28339		Nuevo Leon	27.27295	−100.18203
	9	28395		Sinaloa	24.75317	−107.49903
	5	28392		Sinaloa	26.30297	−109.03337
	14	28376		Chihuahua	27.14178	−104.92337
	8	28307		Tamaulipas	23.00933	−99.69512
	8	28366		Zacatecas	24.71628	−101.3117
	12	28381		Chihuahua	28.57192	−105.95682
	10	28390		Sonora	27.5891	−109.92963
	7	28385		Sonora	30.20812	−111.09735
	4	PI 413123		Mayo Experimental Station		
	4	PI 413067		North of España		

Numbers with a "PI" (GRIN ID) were obtained from the US Department of Agriculture. Other accessions were collected by R.A.B. and D.L.L. from 2001 to 2005. All wild collections retrieved by R.A.B. and D.L.L. displayed the wild- type morphology with small heads, thin pedicels, and multiple flowering stalks per plant.





Gene	Samples	GenBank accession numbers
<i>HaGA2ox</i>	American/Canadian panel	JF428412–JF428543
	Mexican panel	JF428388–JF428411
		JF428544–JF428719
<i>HaFT1</i>	American/Canadian panel	GQ884463–GQ884584
	Mexican panel	JF428724–JF428779
<i>c4973</i>	Chapman 2008 panel	FJ373761–FJ373782
	American landraces	JF428180–JF428353
	Mexican panel	JF428354–JF428387
PgiC (reference gene; homolog of phosphoglucose isomerase C)	American/Canadian Panel	GQ884717 –GQ884848
SCR1 (reference gene; homolog of SCARECROW)	American/Canadian panel	GQ884849–GQ884980
CGP41 (reference gene; homolog of AVRPPHB SUSCEPTIBLE 1)	American/Canadian panel	GQ884988–GQ885119
CGP53 (reference gene; homolog of ARABIDOPSIS THALIANA BASIC LEUCINE ZIPPER 11)	American/Canadian panel	GQ884199–GQ884330
CGP62 (reference gene; homolog of CYCLIC NUCLEOTIDE-GATED CHANNEL 15)	American/Canadian panel	HQ110110–HQ110241
CGP69 (reference gene; homolog of PURPLE ACID PHOSPHATASE 17)	American/Canadian panel	HQ110242–HQ110373
CGP112 (reference gene; homolog of HUA1, ENHANCER OF AG-4 1)	American/Canadian panel	HQ119174–HQ110505
<i>HaFT2</i>	Guerrero 34111	JF428720J–F428721
<i>HaFT3</i>	Guerrero 34111	JF428722–JF428723