DOI: 10.1534/genetics.105.051110

Patterns of Nucleotide Diversity in Wild and Cultivated Sunflower

Aizhong Liu and John M. Burke[1]

Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235

Manuscript received September 13, 2005
Accepted for publication November 16, 2005

ABSTRACT

Interest in the level and organization of nucleotide diversity in domesticated plant lineages has recently been motivated by the potential for using association-based mapping techniques as a means for identifying the genes underlying complex traits. To date, however, such data have been available only for a relatively small number of well-characterized plant taxa. Here we provide the first detailed description of patterns of nucleotide polymorphism in wild and cultivated sunflower (*Helianthus annuus*), using sequence data from nine nuclear genes. The resuflts of this study indicate that wild sunflower harbors at least as much nucleotide diversity as has been reported in other wild plant taxa, with randomly selected sequence pairs being expected to differ at 1 of every 70 bp. In contrast, cultivated sunflower has retained only 40–50% of the diversity present in the wild. Consistent with this dramatic reduction in polymorphism, a phylogenetic analysis of our data revealed that the cultivars form a monophyletic clade, adding to the growing body of evidence that sunflower is the product of a single domestication. Eight of the nine loci surveyed appeared to be evolving primarily under purifying selection, while the remaining locus may have been the subject of positive selection. Linkage disequilibrium (LD) decayed very rapidly in the self-incompatible wild sunflower, with the expected LD falling to negligible levels within 200 bp. The cultivars, on the other hand, exhibited somewhat higher levels of LD, with nonrandom associations persisting up to ~1100 bp. Taken together, these results suggest that association-based approaches will provide a high degree of resolution for the mapping of functional variation in sunflower.

THE domestication of crop plants is typically accompanied by a genomewide loss of genetic diversity (Tanksley and McCouch 1997). This reduction in diversity is typically due, at least in part, to the population bottleneck that occurs during the founding of a new crop lineage (*e.g.*, Eyre-Walker *et al.* 1998). In addition to this so-called "domestication bottleneck," the transition to self-fertilization that often accompanies domestication can further reduce levels of genetic diversity (Pollack 1987; Nordborg 2000), as can selection on the genes underlying agronomically important traits (although this latter effect occurs in a locus-specific fashion; *e.g.*, Hanson *et al.* 1996; Tenaillon *et al.* 2004). While the effects of domestication on genetic diversity are likely to vary across taxa, comprehensive surveys of nucleotide diversity in crop plants and their wild progenitors have been performed in only a handful of systems. On the basis of data from the major cereal crops, it appears that genomewide reductions in diversity on the order of 30–40% are not uncommon (Buckler *et al.* 2001), with selectively important loci often exhibiting even greater losses (*e.g.*, Whitt *et al.* 2002). In addition to these effects on the overall level of polymorphism, domestication can also have a major impact on the organization of genetic diversity within the genome. Indeed, population bottlenecks can produce transient increases in linkage disequilibrium (LD, the nonrandom association of alleles at different sites) throughout the genome. Similarly, the increase in homozygosity associated with a transition to partial or full self-fertilization reduces the effective recombination rate, once again resulting in elevated LD across the genome (Nordborg 2000). Selection can have a similar, albeit localized, effect on LD in and around the targeted loci (*e.g.*, Clark *et al.* 2004).

Beyond the obvious concern that reduced genetic variability might limit the potential for crop improvement over the long term (Harlan 1984), interest in the level and organization of nucleotide variability in domesticated plant lineages has recently been motivated by the potential for using association-mapping techniques as a means for identifying the genes underlying agronomically important traits (Flint-García *et al.* 2003). In the extreme, association-based approaches can even be used to identify the single-nucleotide polymorphisms (SNPs) that are actually responsible for particular trait differences (*i.e.*, so-called quantitative trait nucleotides, QTNs) (Long and Langley 1999). While association mapping promises to provide a great deal of insight into the genetic basis of complex traits,

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. DQ503586–DQ504161.

[1]*Corresponding author:* Department of Biological Sciences, Vanderbilt University, VU Station B 351634, Nashville, TN 37235.
E-mail: john.m.burke@vanderbilt.edu

this approach requires a detailed understanding of the distribution of genetic variation across the genome, including data on the density of SNPs and the structure of LD.

To date, the vast majority of nucleotide polymorphism data in plants have come from a relatively small (but growing) number of well-characterized study systems, such as Arabidopsis (*e.g.*, Savolainen *et al.* 2000; Aguadé 2001; Nordborg *et al.* 2002; Wright *et al.* 2003; Ramos-Onsins *et al.* 2004), several major crops (*e.g.*, White and Doebley 1999; Tenaillon *et al.* 2001, 2002; Garris *et al.* 2003; Zhu *et al.* 2003; Hamblin *et al.* 2004), and a handful of other taxa (*e.g.*, Lin *et al.* 2001; Tiffin and Gaut 2001; Dvornyk *et al.* 2002; García-Gil *et al.* 2003; Kado *et al.* 2003; Brown *et al.* 2004; Ingvarsson 2005). While some generalities have emerged from these studies (*e.g.*, a tendency toward reduced levels of polymorphism and elevated LD in selfers *vs.* outcrossers), it is clear that the details learned from the study of any one system do not necessarily apply to another, even if they share similar mating systems, demographic histories, etc. With this in mind, we set out to provide the first detailed description of the level of nucleotide diversity and the extent of LD in a broad collection of wild and cultivated sunflower accessions.

Derived from the wild sunflower (*Helianthus annuus*), the cultivated sunflower (also *H. annuus*) is one of the world's most important oilseed crops and is also a major source of confectionery seeds (Putt 1997). Despite being fully interfertile and considered to be members of the same species, wild and cultivated sunflower exhibit a number of striking morphological differences. In short, wild sunflower has a highly branched growth form with numerous small flowering heads and relatively small achenes (*i.e.*, single-seeded fruits) that are dispersed at maturity. In contrast, cultivated sunflower is characterized by an unbranched stem that is topped by a single, large head and relatively large achenes that are retained until harvest. Moreover, wild sunflower is an obligate outcrosser, whereas cultivated sunflower has lost the sporophytic self-incompatibility that is typical of the genus. Despite this potentially major shift in breeding system, however, the extent to which cultivated sunflower actually self-pollinates in the field remains unknown.

Although cultivated sunflower was long thought to be the product of a single origin of domestication >4000 years ago (Heiser 1954, 1955; Rieseberg and Seiler 1990; Crites 1993), this premise was subsequently called into question on the basis of both archaeological and genetic evidence (*e.g.*, Heiser 1985; Lentz *et al.* 2001; Tang and Knapp 2003). In the most comprehensive molecular analysis to date, however, Harter *et al.* (2004) argued convincingly that the eight extant Native American landraces, from which the modern cultivars are presumably derived, can all be reliably assigned to a single population genetic cluster. This result led them to conclude that these lines do, in fact, trace back to a single origin of domestication, most likely somewhere within what is now considered to be the central United States.

While sunflower has recently been the subject of a substantial amount of EST sequencing (see http://compgenomics.ucdavis.edu), detailed analyses of sequence diversity derived from a broad sample of germplasm are lacking. Rather, analyses of genetic diversity in sunflower have thus far relied primarily on techniques such as allozyme (Rieseberg and Seiler 1990; Cronn *et al.* 1997) and SSR (Tang and Knapp 2003; Harter *et al.* 2004; Burke *et al.* 2005) genotyping. Here we seek to rectify this situation by reporting on patterns of nucleotide polymorphism in a widespread sample of wild sunflower individuals, as well as a diverse collection of cultivars.

MATERIALS AND METHODS

Sampling strategy and plant materials: Seeds of 16 wild *H. annuus* populations and 16 cultivated lines were obtained from the North Central Regional Plant Introduction Station (NCRPIS, Ames, IA; Table 1). The wild populations included in this study were selected to provide broad geographic coverage of the species' range in North America. The 16 cultivated lines were composed of 8 Native American landraces, which represent the most primitive sunflower domesticates available, and 8 improved lines that were selected such that, when combined with the landraces, our collection of cultivars contained at least one representative from 9 of the 10 subsets that make up the NCRPIS *H. annuus* core collection. With the exception of cmsHA89, which is an elite inbred oilseed line that has been used in a variety of other studies (*e.g.*, Burke *et al.* 2002, 2004; Tang and Knapp 2003), the improved lines included here represent open-pollinated cultivars. Thus, the 16 cultivars included in this study are largely comparable to the "exotic" lines employed by Burke *et al.* (2005). Upon receipt, seeds from each accession were germinated and the resulting seedlings were reared in the greenhouse. Following emergence, 200 mg of leaf tissue was collected from each seedling, and total genomic DNA was extracted from one individual per accession using the QIAGEN (Valencia, CA) DNeasy plant mini kit.

Loci studied: The nine genes that were selected for inclusion in this study are briefly outlined below (see also Table 2). Calmodulin (*CAM*) plays a central role in calcium-mediated signaling in plants. Chalcone synthase (*CHS*; EC 2.3.1.74) plays an essential role in the biosynthesis of plant phenylpropanoids. Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*; EC 1.2.1.12) is a tetrameric NAD+ binding protein that is involved in glycolysis and gluconeogenesis. Cytosolic phosphoglucose isomerase (*PGIC*; EC 5.3.1.9) catalyzes the reversible isomerization of 6-phosphoglucose and 6-phosphofructose, an essential reaction that precedes sucrose biosynthesis. *GIA/RGA* is a putative gibberellin response modulator. Glutathione peroxidase (*GPX*; EC 1.11.1.9) and glutathione *S*-transferase (*GST*; EC 2.5.1.18) are antioxidants that are thought to play an important role in protecting against oxidative damage. Finally, *SCR*-1 and *SCR*-2 show homology to SCARECROW (*SCR*) or SCARECROW-like gene regulators. *SCR* is known to be involved in asymmetric cell division in plants (*e.g.*, Kamiya *et al.* 2003). The genetic map positions of all nine of these genes are currently unknown.

TABLE 1

List of populations/lines surveyed in this study, along with accession IDs and an indication of improvement status

Population/line name (location)	Accession ID	Improvement status
AZ (Arizona)	Ames 14400	Wild
CA (California)	PI 613732	Wild
CO (Colorado)	PI 586840	Wild
IA (Iowa)	PI 597895	Wild
KS (Kansas)	PI 413027	Wild
MEX (north of España, Mexico)	PI 413067	Wild
MO (Missouri)	PI 531032	Wild
MT (Montana)	PI 531032	Wild
NE (Nebraska)	PI 586865	Wild
OK (Oklahoma)	PI 435619	Wild
SAS (Saskatchewan, Canada)	PI 592317	Wild
SD (South Dakota)	Ames 23940	Wild
TN (Tennessee)	PI 435552	Wild
TX (Texas)	Ames 7442	Wild
UT (Utah)	PI 531009	Wild
WY (Wyoming)	PI 586822	Wild
Arikara	PI 369357	Primitive
Havasupai	PI 369358	Primitive
Hidatsa	PI 600721	Primitive
Hopi	PI 432504	Primitive
Maíz de Tejas	Ames 6859	Primitive
Maíz Negro	Ames 19070	Primitive
Mandan	PI 600717	Primitive
Seneca	PI 369360	Primitive
cmsHA89	Ames 3963	Improved
Jupiter	PI 296289	Improved
Mennonite	Ames 7574	Improved
Peredovik	PI 372173	Improved
Sundak	Ames 4114	Improved
VIR 847	PI 386230	Improved
VK-47	Ames 3361	Improved
VNIIMK 8931	PI 340790	Improved

PCR amplification and sequencing: Primers for all loci except *PGIC* were designed solely on the basis of sunflower EST sequences contained within the Compositae Genome Project Database (http://cgpdb.ucdavis.edu; see Table 2 for contig IDs). In contrast, exons 16–21 of *PGIC* were initially amplified using universal primers (yamV and AA16F) developed by L. D. Gottlieb (unpublished data). Once this region was sequenced (see below for details), an internal primer (16R) was designed from our sequences and used along with an EST-derived primer from exon 12 (12F) to amplify exons 12–16. Thus, we were able to sequence exons 12–21 of this gene. Similarly, *CHS* was amplified on the basis of primer pairs designed from two contigs that overlapped, but that had not previously been assembled into a single unigene. The internal primers were designed from the region of overlap, such that the sequences could subsequently be assembled end-to-end.

Wherever possible, PCR products were purified using the QIAGEN PCR purification kit and directly sequenced. Heterozygotes were dealt with in several ways. First, if the two alleles within an individual differed sufficiently in length, they were separated via agarose gel electrophoresis, isolated, and sequenced. In cases where alleles could not be separated, but direct sequence could be generated, the forward and reverse sequences were assembled, heterozygous sites were identified using Sequencher (Gene Codes, Ann Arbor, MI), and haplotypes were inferred via "haplotype substraction" (Clark 1990; Olsen and Schaal 1999). In cases where PCR products could not be directly sequenced, or where haplotypes could not be readily inferred, PCR products were cloned using the QIAGEN PCR cloning kit prior to sequencing. In such cases, multiple clones were sequenced from each individual to distinguish between alleles within an individual and to control for *Taq* polymerase errors. Thus, each individual was represented by two alleles at each locus. All genes were sequenced in both directions using DYEnamic ET cycle sequencing kits (Amersham Biosciences, Piscataway, NJ) following the manufacturer's protocol on an MJ BaseStation automated DNA sequencer (MJ Research, South San Francisco).

Sequence analyses: Multiple sequence alignments were made using Se-Al version 2.0a11 (Rambaut 1996; http://evolve.zoo.ox.ac.uk). The coding and noncoding regions of each gene were then identified by aligning our sequences against the original EST sequences and via BLAST searches. Estimates of nucleotide polymorphism (π and θ, calculated on a per site basis), population subdivision (*i.e.*, F_{ST} between wild and cultivated sunflower), and Tajima's (1989) *D* were obtained using the software package DnaSP 4.00.5 (Rozas and Rozas 1999). DnaSP was also used to estimate the minimum number of recombination events (RM) in the history of the wild and cultivated subsamples, using the four-gamete test (Hudson and Kaplan 1985) as well as the strength of linkage disequilibrium between pairs of polymorphic sites (computed as the squared allele frequency correlation, r^2; Weir 1990). The population recombination parameter ($\rho = 4N_e r$, where N_e is the effective population size and r is the recombination rate) was estimated using the composite-likelihood estimator of Hudson (2001) as implemented in the software package LDhat (available from http://www.stats.ox.ac.uk/~mcvean/LDhat/), and Wall's (1999) *B* was estimated using COMPUTE (Thornton 2003). Contiguous indels were treated as single polymorphisms, and singletons were excluded from all analyses of linkage disequilibrium.

The decay of linkage disequilibrium over physical distance was investigated following the methods of Remington *et al.* (2001). Briefly, the expected value of r^2 at drift-recombination equilibrium is $E(r^2) = 1/(1 + \rho)$ (Hill and Weir 1988). Allowing for a low level of mutation and correcting for finite sample size, this relationship becomes

$$E(r^2) = \left[\frac{10 + \rho}{(2 + \rho)(11 + \rho)}\right]\left[1 + \frac{(3 + \rho)(12 + 12\rho + \rho^2)}{n(2 + \rho)(11 + \rho)}\right], \quad (1)$$

where n is the number of sequences sampled. The nonlinear equation based on this relationship contains a single coefficient (b_1), which corresponds to the least-squares estimate of ρ per base pair. We pooled our data across genes and fit this model separately for the wild and cultivated samples using PROC NLIN in SAS Ver. 6.12 (SAS Institute, Cary, NC). Although factors such as nonindependence among linked sites and nonequilibrium populations can reduce the precision of and/or bias such analyses, possibly resulting in unreliable estimates of ρ (Weir and Hill 1986), such analyses are still useful for investigating the overall rate of decay of linkage disequilibrium (*e.g.*, Remington *et al.* 2001; Ingvarsson 2005). Following the methods of Macdonald *et al.* (2005), we also summarized the observed r^2-values using the ksmooth function in the statistical programming language R (http://www.R-project.org/).

Phylogenetic analyses: To further investigate the origin of cultivated sunflower, we constructed a phylogeny of the 16 wild

TABLE 2

Summary of genes surveyed and primer sequences employed

Gene name	Contig ID[a]	Functional association via BLAST	Primer sequences
CAM	Contig1037	Calmodulin	5′-GAAACTCTGGGAAGTCGATTG
			5′-TCAGCGCAAATTACCCAAAT
CHS	Contig1392	Chalcone synthase	5′-GTGTGCTCCAAACCACATATC
			5′-TTTGAGCAGCAGAAATGATCT
	Contig4492		5′-TACCGGGACTTATCTCGAAAC
			5′-CAAACAAGTGTCCCAAAGTT
GAPDH	Contig1499	Glyceraldehyde-3-phosphate dehydrogenase	5′-TTTGGTGATTGAAGGCACAG
			5′-TCGTTCCAAATACTTCAAACTCT
GIA/RGA	Contig3138	GIA/RGA-like gibberellin response modulator	5′-CGGTGATTTCGGATACAATCT
			5′-CTGAAACTCACCACCAATTCA
GPX	Contig268	Glutathione peroxidase	5′-CCCAACCTCCAAGTTGACAC
			5′-TGCATGCATAGAAAGTTTGTTATT
GST	Contig5423	Glutathione *S*-transferase	5′-GGTGCCATCTCTTGAACACA
			5′-ATCCAGGCTGCTAATTTTGG
PGIC	Contig1461	Cytosolic phosphoglucose isomerase	12F: 5′-TATCTCTCCATACGGGTTTTCC
			16R: 5′-GATTTCACCAGCTTCAAAAGGA
	NA		AA16F: 5′-ATGGARAGYAAYGGNAARGG
			yamv: 5′-TCIACICCCCAITGRTCAAAIGARTTIAT
SCR-1	Contig1850	Putative Scarecrow gene regulator	5′-GGAATCCTGTCTGCTGATAAGT
			5′-TTCACTTGCGAAACAAGCTC
SCR-2	Contig1874	Putative Scarecrow gene regulator	5′-TTGGAAACGGACTAAACAGTTG
			5′-CGCAACCGAACAACTAAACC

[a] Refers to contig IDs from the Compositae Genome Project Database (http://cgpdb.ucdavis.edu/).

sunflower accessions, as well as the 8 Native American landraces (*i.e.*, the "primitive" lines), which, unlike the "improved" lines that made up the balance of our sequencing panel, are free from the confounding effects of human-mediated introgression during the postdomestication era. We used the neighbor-joining algorithm of PAUP Ver. 4.0b10 (Swofford 2002) to construct a phylogeny on the basis of the combined sequence data. Indels were recoded as numerical characters prior to analysis, and branch support was estimated on the basis of 1000 bootstrap replicates of the data.

RESULTS

Sequence diversity: All nine gene regions were sequenced in each of the 32 sampled individuals. Including indels, sequence lengths varied from 504 to 1642 bp (Table 3), and sequences from all genes but *SCR*-1 and *SCR*-2 included both coding and noncoding (*i.e.*, intron and/or UTR) regions. Thus, we were able to analyze 8207 bp of aligned sequence per individual, with nearly two-thirds (5328 bp) coming from coding regions. Across samples, the number of indel polymorphisms per gene varied from 0 to 12, with a total of 31 indel polymorphisms in the data set. Of these, all but 1 (a 3-bp indel in the coding region of *GIA/RGA*) occurred in noncoding regions. Indel size was highly variable, ranging from a single nucleotide in some cases (including three single-base indels embedded within mononucleotide repeat motifs in *PGIC*) to >100 bp in others. More specifically, two wild individuals harbored *CAM* indels spanning ≥100 bp, and the largest indel observed (250 bp) was found within one of the *PGIC* introns (flanked by exons 12 and 13) in two wild individuals. All indels were excluded from subsequent analyses of nucleotide polymorphism.

Single-nucleotide polymorphisms were considerably more frequent than indels, with a total of 444 polymorphic sites being identified across all individuals and all genes, resulting in an average of 1 SNP for every 16.8 bp of sequence (excluding indels). When considered separately, the wild sunflowers harbored 392 polymorphic sites (1 SNP/19.1 bp), whereas the cultivars

TABLE 3

Lengths of gene regions analyzed in base pairs, excluding indels

Gene	Total length	No. of indels	Length of noncoding region	Length of coding region
CAM	822	9	720	102
CHS	1336	0	268	1068
GAPDH	920	2	248	672
GIA/RGA	752	4	67	685
GPX	830	4	236	594
GST	660	3	309	351
PGIC	1642	9	1031	611
SCR-1	741	0	—	741
SCR-2	504	0	—	504
Total	8207	31	2879	5328

TABLE 4

Summary of measures of nucleotide variability and Tajima's *D*

Gene	Sample	S[a]	θ_W	π_T	π_{sil}	π_{syn}	π_{nonsyn}	π_{nonsyn}/π_{syn}	Tajima's *D*
CAM	Total	65 (49)	0.0266	0.0176	0.0207	0.0500	0.0012	0.02	−1.25
	Wild	59 (46)	0.0270	0.0218	0.0238	0.0610	0.0015	0.03	−0.81
	Cultivated	39 (22)	0.0180	0.0137	0.0160	0.0359	0.0008	0.02	−0.88
CHS	Total	19 (17)	0.0030	0.0016	0.0035	0.0027	0.0004	0.15	−1.45
	Wild	17 (14)	0.0032	0.0023	0.0046	0.0048	0.0008	0.17	−0.97
	Cultivated	4 (4)	0.0007	0.0006	0.0017	0.0005	0	0	−0.37
GAPDH	Total	18 (16)	0.0042	0.0018	0.0033	0.0017	0.0006	0.35	−1.72
	Wild	17 (15)	0.0046	0.0026	0.0044	0.0033	0.0011	0.33	−1.49
	Cultivated	2 (2)	0.0005	0.0009	0.0020	0	0	—	1.35
GIA/RGA	Total	27 (27)	0.0078	0.0055	0.0173	0.0227	0.0007	0.03	−1.07
	Wild	27 (27)	0.0091	0.0089	0.0271	0.0354	0.0014	0.04	−0.32
	Cultivated	4 (3)	0.0013	0.0008	0.0029	0.0038	0	0	−0.95
GPX	Total	39 (37)	0.0101	0.0041	0.0055	0.0027	0.0031	1.15	−2.02*
	Wild	29 (29)	0.0090	0.0058	0.0087	0.0036	0.0036	1.01	−1.27
	Cultivated	14 (12)	0.0042	0.0025	0.0032	0.0018	0.0020	1.11	−1.48
GST	Total	84 (13)	0.0317	0.0328	0.0527	0.0837	0.0117	0.14	−0.28
	Wild	76 (10)	0.0337	0.0356	0.0585	0.0898	0.0116	0.13	−0.12
	Cultivated	39 (20)	0.0173	0.0122	0.0177	0.0325	0.0064	0.20	−1.20
PGIC	Total	132 (110)	0.0205	0.0156	0.0218	0.0209	0.0030	0.14	−1.13
	Wild	115 (94)	0.0208	0.0187	0.0250	0.0269	0.0044	0.16	−0.76
	Cultivated	66 (58)	0.0121	0.0091	0.0126	0.0102	0.0015	0.15	−1.07
SCR-1	Total	29 (27)	0.0083	0.0067	0.0179	0.0179	0.0019	0.11	−0.71
	Wild	24 (22)	0.0081	0.0072	0.0188	0.0188	0.0022	0.12	−0.52
	Cultivated	15 (14)	0.0051	0.0060	0.0157	0.0157	0.0016	0.10	0.59
SCR-2	Total	31 (22)	0.0130	0.0093	0.0302	0.0302	0.0028	0.09	−1.07
	Wild	28 (20)	0.0138	0.0122	0.0398	0.0398	0.0037	0.09	−0.61
	Cultivated	11 (10)	0.0054	0.0042	0.0148	0.0148	0.0010	0.07	−0.69
Average	Total	49.3 (35.3)	0.0139	0.0106	0.0192	0.0258	0.0028	0.24	
	Wild	43.6 (30.8)	0.0144	0.0128	0.0234	0.0315	0.0034	0.23	
	Cultivated	21.6 (16.1)	0.0072	0.0056	0.0096	0.0128	0.0015	0.21	

* $P < 0.05$.

[a] The number of segregating sites is indicated. The number of nonsingletons is included in parentheses.

harbored 194 polymorphic sites (1 SNP/38.8 bp). Inspection of Table 4 confirms that levels of nucleotide polymorphism are generally quite high. More specifically, estimates of total nucleotide diversity (π_T) for the data set as a whole ranged from 0.0016 to 0.0328 (mean = 0.0106), and Watterson's θ (θ_W) ranged from 0.0030 to 0.0317 (mean = 0.0139). Not surprisingly, a comparison of the wild and cultivated subsamples revealed that π_T and θ_W are both significantly higher in wild sunflower as compared to the cultivars (0.0128 *vs.* 0.0056; paired $t = 3.14$, d.f. = 8, $P = 0.007$ and 0.0144 *vs.* 0.0072; paired $t = 5.03$, d.f. = 8, $P = 0.0005$, respectively). Similarly, silent-site diversity (π_{sil}) as well as synonymous (π_{syn}) and nonsynonymous (π_{nonsyn}) nucleotide diversity was significantly higher in the wild subsample than in the cultivars (all $P \leq 0.008$). In terms of the extent of divergence between subsamples, F_{ST} values averaged 0.1837 ± 0.038 (mean ± SE), indicating that the wild and cultivated sunflower gene pools are moderately differentiated.

Tests for nonneutral evolution: For all loci except *GPX*, π_{nonsyn} was markedly lower than π_{syn}, with the π_{nonsyn}/π_{syn} ratio ranging from 0.024 to 0.353 in the full data set (the corresponding values for the wild and cultivated subsamples were 0.025–0.333 and 0–0.197, respectively), suggesting that diversity at these eight loci is largely governed by purifying selection. For *GPX*, on the other hand, the π_{nonsyn}/π_{syn} ratio for the full data set was 1.148, whereas the corresponding values for the wild and cultivated subsamples were 1.005 and 1.110, respectively. This result suggests either that *GPX* has experienced a relaxation of the purifying selection that has presumably shaped diversity at the other eight genes or that some portion of the *GPX* coding region has been under positive selection. In terms of allele frequency distributions, Tajima's *D* was significantly negative at *GPX* in the full data set (Table 4), indicating an excess of rare alleles. While superficially consistent with the hypothesis that *GPX* was the target of recent positive selection, it must be kept in mind that: (1) the corresponding estimates from both the wild and cultivated subsamples were not significantly different from zero when they were considered separately, and (2) estimates of Tajima's *D* were generally negative (albeit nonsignificantly so)

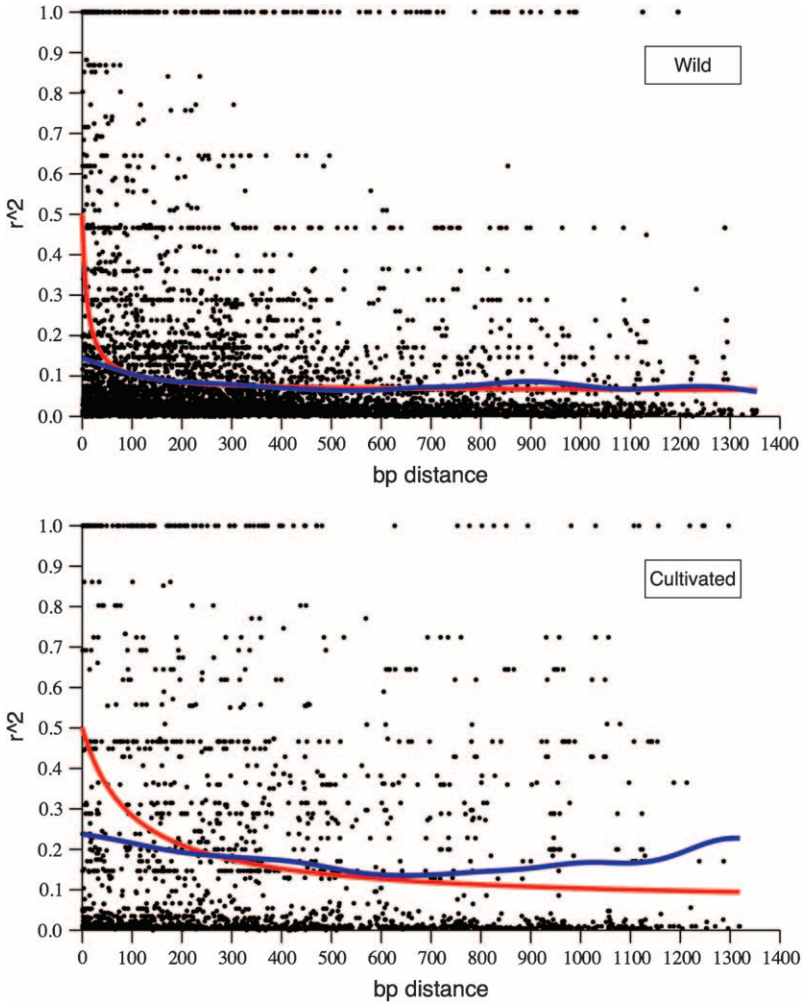

FIGURE 1.—Plots of the squared allele frequency correlations (r^2) as a function of physical distance between sites in wild and cultivated sunflower. The red line on each graph depicts the expected decline in linkage disequilibrium based on a nonlinear regression of r^2 against distance, using a mutation–recombination–drift model. The blue line represents a smoothed line through the raw data and thus provides a summary of the observed decline in linkage disequilibrium. See text for details.

across all other loci. Thus, factors other than selection may be responsible for the observed excess of rare alleles.

Linkage disequilibrium: Results of the LD analyses are summarized in Figure 1 and Table 5. Data from all nine genes were pooled for the wild and cultivated subsamples and Equation 1 was used to model the decay of LD across physical distance. In general terms, the expected value of r^2 declines very rapidly in wild sunflower, falling to negligible levels (*i.e.*, ≤0.10) within 200 bp, whereas somewhat higher levels of LD are maintained across greater distances in cultivated sunflower. Observed levels of LD are actually somewhat lower than the expected values at short distances, although the wild data largely follow the expectation. For the cultivar data, observed LD declines out to ~650 bp, at which point it begins to drift upward. This pattern is likely due, at least in part, to increased sampling variation in the cultivars resulting from lower overall levels of polymorphism. While we were unable to estimate certain recombination parameters for all nine genes in the cultivated subsample due to a lack of sufficient polymorphism, estimates of the population recombination parameter using HUDSON's (2001) composite-likelihood estimator ranged from 0.0012 to 0.1483 (0.0528 ± 0.016) in wild sunflower and from 0.0036 to 0.0298 (0.0155 ± 0.007) in cultivated sunflower. Similarly, the minimum number of recombination events ranged from 2 to 11 (7.7 ± 1.8) in wild sunflower and from 0 to 9 (2.7 ± 1.1) in cultivated sunflower, and Wall's B ranged from 0 to 0.4286 (0.1180 ± 0.051) in wild sunflower and from 0.0526 to 0.3611 (0.1760 ± 0.065) in cultivated sunflower. In all three cases, the differences were significant (paired t-test, both $P < 0.05$) with wild sunflower exhibiting higher recombination (and thus lower LD) than cultivated sunflower. Estimates of interlocus disequilibrium were negligible for all nine genes within both the wild and cultivated subsamples (data not shown).

Another method for investigating patterns of linkage disequilibrium, particularly when it comes to making comparisons among populations, is to scale estimates of ρ against θ. The rationale for doing so is that, under the assumptions of the standard neutral model, both values are proportional to the effective population size ($\rho = 4N_e r$ and $\theta = 4N_e\mu$, respectively; HUDSON 1987), such that the ratio ρ/θ becomes the recombination rate divided by the mutation rate (*i.e.*, r/μ). This ratio ranged from 0.98 to 11.87 (4.10 ± 2.6) in wild sunflower and from 0.21 to 5.84 (1.94 ± 1.3) in the cultivars at the

TABLE 5

Summary of the observed number of unique haplotypes within the wild and cultivated sunflower subsamples, as well as estimates of the minimum number of recombination events (RM), Hudson's (2001) estimate of the population recombination parameter (ρ), and Wall's *B*

Gene	Sample	No. haplotypes	RM	ρ[a]	Wall's *B*[a]
CAM	Wild	21	11	0.0645	0
	Cultivated	15	3	0.0242	0.0526
CHS	Wild	14	4	0.0142	0
	Cultivated	5	1	—	—
GAPDH	Wild	12	2	0.0164	0.3125
	Cultivated	3	0	—	—
GIA/RGA	Wild	25	7	0.1483	0
	Cultivated	4	0	—	—
GPX	Wild	14	3	0.0012	0.4286
	Cultivated	13	0	—	—
GST	Wild	27	14	0.0268	0.0870
	Cultivated	13	4	0.0036	0.3611
PGIC	Wild	27	18	0.0241	0.0680
	Cultivated	15	9	0.0044	0.1475
SCR-1	Wild	20	5	0.0962	0.0455
	Cultivated	14	6	0.0298	0.1429
SCR-2	Wild	24	5	0.0833	0.1200
	Cultivated	5	1	—	—
Average	Wild	20.4	7.9	0.0528	0.1180
	Cultivated	9.7	2.7	0.0155	0.1760

[a] Hudson's ρ and Wall's *B* could be estimated only for four of the nine genes in the cultivated sunflower subsample due to a lack of sufficient polymorphism in the remaining five loci.

four loci for which we were able to estimate ρ (using Hudson's 2001 estimator) from both the wild and cultivated subsamples (see Tables 4 and 5). While this result is consistent with greater recombination in wild as compared to cultivated sunflower, the difference was not significant (paired *t*-test $P = 0.10$).

Phylogenetic insights: Inspection of the neighbor-joining tree in Figure 2 reveals that the primitive domesticates form a relatively well-supported, monophyletic clade. Note that the inclusion of the "improved" accessions resulted in a similar overall pattern (data not shown). The primary difference following the inclusion

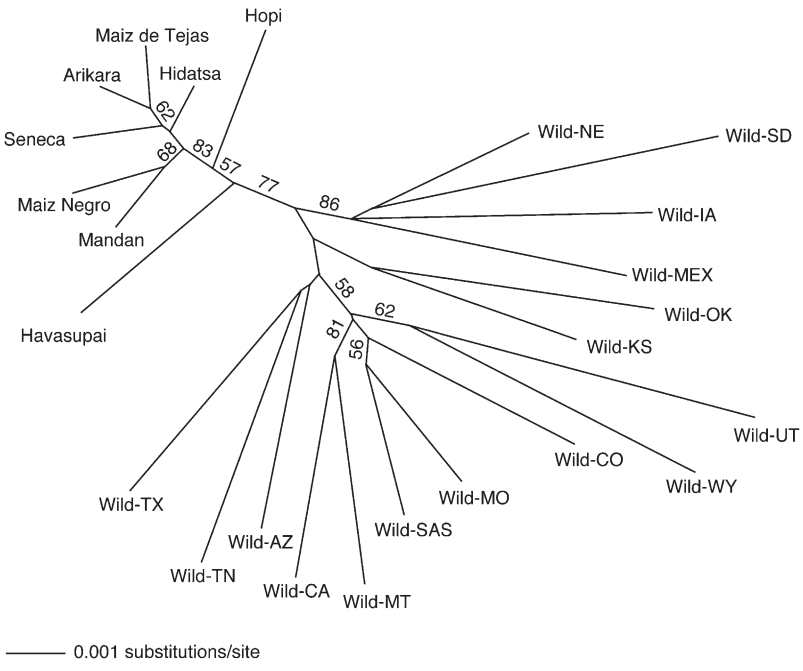

Figure 2.—Neighbor-joining tree of 16 wild sunflower accessions and the eight extant Native American landraces. Numbers refer to bootstrap values for branches with >50% support.

of the improved lines was a decrease in bootstrap support, perhaps due to wild × cultivar introgression during sunflower improvement. The cultivars, however, still formed a monophyletic clade. Note that the apparent monophyly of the cultivars is apparent only when the data are combined across loci. While these results are fully consistent with the conclusions of Harter *et al.* (2004) regarding a single origin of domesticated sunflower, our data do not provide sufficient resolution to corroborate their placement of the domestication event in the central United States.

DISCUSSION

Sequence diversity: This study represents the most comprehensive analysis of DNA sequence variation in wild and cultivated sunflower to date. Although there was considerable locus-to-locus variation, with estimates of nucleotide diversity varying >10-fold across loci, it is clear that wild sunflower contains substantial levels of nucleotide diversity (Table 4). Indeed, wild sunflower appears to harbor at least as much silent-site diversity ($\pi_{sil} = 0.0234 \pm 0.006$) as do a number of other wild taxa that have been studied to date. For example, species-wide silent-site diversity (π_{sil}) in the selfing *Arabidopsis thaliana* is 0.011 (Aguadé 2001), whereas in the outcrossing species *A. lyrata* and *A. halleri* the corresponding values are 0.023 and 0.015 (Wright *et al.* 2003; Ramos-Onsins *et al.* 2004). Similarly, three wild relatives of maize, *Zea diploperennis*, *Z. perennis*, and *Z. parviglumis* have $\pi_{sil} = 0.012$, 0.013, and 0.023, respectively (White and Doebley 1999; Tiffin and Gaut 2001), and the highly outcrossed tree species *Populus tremula* has $\pi_{sil} = 0.016$ (Ingvarsson 2005).

In contrast to wild sunflower, cultivated sunflower contains markedly less nucleotide variation, with the cultivars included in our survey exhibiting only 40–50% as much diversity (depending on the measure) as was found in the wild. In terms of the overall density of polymorphisms across the regions that we analyzed, we found an average of 1 SNP/19.1 and 38.8 bp across our samples of wild and cultivated sunflower, respectively. Because θ_W is roughly proportional to heterozygosity, we can further conclude that a randomly selected pair of wild (or cultivated) sunflower sequences would be expected to differ at an average of 1 of every ~70 (or ~140) nucleotides (*i.e.*, $1/0.0144 \approx 70$ and $1/0.0072 \approx 140$). For the sake of comparison, randomly selected pairs of maize sequences are expected to differ at 1 of every ~105 nucleotides (Tenaillon *et al.* 2001), whereas pairs of soybean sequences are expected to differ at 1 of every ~1030 nucleotides (Zhu *et al.* 2003).

While the pattern documented here is qualitatively similar to what has been found in previous surveys of genetic variation in sunflower, the loss of diversity is somewhat greater with sequence data than with either allozymes or SSRs. For example, Rieseberg and Seiler (1990) and Cronn *et al.* (1997) found that cultivated sunflower contains ~50–60% of the allozyme diversity present in wild sunflower. Both of these studies, however, reported only the mean level of within-population heterozygosity, as opposed to a true specieswide estimate of diversity, and thus are not strictly comparable to our data. With regard to SSR variation, the exotic cultivated sunflower gene pool has previously been shown to contain ~65–80% of the diversity present across the range of wild sunflower (Tang and Knapp 2003; Harter *et al.* 2004; Burke *et al.* 2005). The fact that this portion of the cultivated sunflower gene pool appears to have lost comparatively little SSR diversity is most likely a result of the relatively high mutation rates that are typical of SSRs (*e.g.*, Diwan and Cregan 1997; Vigouroux *et al.* 2002). Given an initial loss of variation, SSR diversity would be expected to rebound much more rapidly than would nucleotide diversity.

The observed loss of diversity from wild to cultivated sunflower is likely due, at least in part, to a population bottleneck during the domestication of sunflower. It is, however, also possible that the loss of self-incompatibility in cultivated sunflower played a role in producing this pattern. Indeed, inbreeding is known to result in both a reduction of effective population size (Pollack 1987) and an amplification of the effects of background selection (Charlesworth *et al.* 1993), both of which would act to reduce genetic variation across the genome (see also Nordborg 2000). It is worth noting here that the primitive and improved accessions that composed the cultivar portion of our sequencing panel (Table 1) contained similar levels of nucleotide diversity when compared to each other (data not shown), indicating that much of the diversity that made it through the initial stages of domestication can be found in the open-pollinated cultivars.

Evidence of selection: Eight of the nine gene regions that we analyzed exhibited low π_{nonsyn}/π_{syn} ratios and thus appear to be evolving primarily under purifying selection. The one exception to this pattern (*GPX*) exhibited a somewhat elevated nonsynonymous substitution rate (Table 4). It is important to note here that the elevated nonsynonymous substitution rate is evident not only across the full sample of 32 sequences, but also within the wild and cultivated subsamples. Thus, it seems unlikely that this pattern arose as a result of selection during domestication. Rather, the most likely explanation is that *GPX*, which is an antioxidant that is thought to play an important role in the defense against oxidative damage in the face of a variety of environmental stresses (Rodriguez Milla *et al.* 2003), has been under divergent selective pressures across both wild and cultivated sunflower accessions.

Linkage disequilibrium: As might be expected of an obligate outcrosser, LD decays extremely rapidly in wild sunflower. More specifically, expected levels of LD decline to negligible levels ($r^2 < 0.10$) within 200 bp

(Figure 1). In contrast, nonrandom associations appear to be maintained over somewhat longer distances in the self-compatible cultivated sunflower, with a predicted decline to $r^2 < 0.10$ within ~1100 bp. While the extent of LD differs somewhat across genes, the overall pattern of higher LD (and lower recombination) in cultivated *vs.* wild sunflower holds across loci (Table 5). This increase in the extent of LD in cultivated sunflower is likely due to a decrease in effective population size owing to the presumptive domestication bottleneck, as well as to a possible increase in the occurrence of inbreeding. Even with the transition to self-compatibility, however, the expected level of LD appears to decay relatively rapidly in cultivated sunflower as compared to predominantly autogamous crops. For example, Zhu *et al.* (2003) concluded that there is little decline in LD over distances as great as 50 kbp in soybean, whereas Garris *et al.* (2003) found that LD in rice approaches $r^2 = 0.10$ only after ~100 kbp (but see Morrell *et al.* 2005 for an example of rapid decline of LD in a predominantly selfing taxon). In contrast, r^2 declines to <0.10 within ~1 kb in maize, which is highly outcrossed (Remington *et al.* 2001; Tenaillon *et al.* 2001). Thus, the patterns documented here appear to be typical of a taxon with a history of relatively frequent outcrossing. It should be noted, however, that our sampling strategy was primarily designed to investigate domestication-related changes in nucleotide diversity and LD. Thus, our data do not provide any insight into patterns of polymorphism in the elite inbred lines, where nonrandom associations might be expected to extend over longer distances.

Conclusions: Taken together, our results indicate that wild sunflower harbors at least as much nucleotide polymorphism as has been reported in other wild plant taxa and that the cultivated sunflower gene pool has retained only 40–50% of this diversity. Our results also add to the growing body of evidence that cultivated sunflower is the product of a single domestication event. As noted above, the issue of diversity loss during domestication has been most thoroughly investigated in the major cereal crops, where losses of 30–40% have been documented (Buckler *et al.* 2001). The fact that cultivated sunflower has experienced a greater domestication-related loss of diversity than is typical of the cereals suggests that sunflower may have experienced an even smaller and/or lengthier domestication bottleneck than did the various cereal crops. The results of our work also suggest that association-based approaches may provide a high degree of resolution for the identification of genes underlying trait variation in sunflower. Indeed, even the self-compatible cultivated sunflower lines included in this study exhibited relatively low levels of LD as compared to predominantly autogamous crops. Given this pattern, most SNPs that are significantly associated with a trait would be expected to reside in relatively close proximity to the causative genetic variant. In the case of the cultivars surveyed here, this should allow functional variation to be mapped to the level of the gene, whereas even finer-scale localization may be possible in wild sunflower.

We thank Mark Chapman, Peter Morrell, Catherine Pashley, Natasha Sherman, Jessica Wenzler, David Wills, and two anonymous reviewers for comments on an earlier version of this manuscript. Stuart Macdonald and David Remington provided assistance with the linkage disequilibrium analyses. This work was supported by grants to J.M.B. from the National Science Foundation (DBI-0332411) and the United States Department of Agriculture (03-35300-13104 and 03-39210-13958). EST sequence data were obtained from the Compositae Genome Project website, which was funded by the USDA IFAFS program.

LITERATURE CITED

Aguadé, M., 2001 Nucleotide sequence variation at two genes of the phenylpropanoid pathway, the *FAH1* and *F3H* genes in *Arabidopsis thaliana*. Mol. Biol. Evol. **18:** 1–9.

Brown, G. R., G. P. Gill, R. J. Kuntz, C. H. Langley and D. B. Neale, 2004 Nucleotide diversity and linkage disequilibrium in loblolly pine. Proc. Natl. Acad. Sci. USA **101:** 15255–15260.

Buckler, E. S., IV, J. F. Thornsberry and S. Kresovich, 2001 Molecular diversity, structure and domestication of grasses. Genet. Res. **77:** 213–218.

Burke, J. M., S. Tang, S. J. Knapp and L. H. Rieseberg, 2002 Genetic analysis of sunflower domestication. Genetics **161:** 1257–1267.

Burke, J. M., Z. Lai, M. Salmaso, T. Nakazato, S. Tang *et al.*, 2004 Comparative mapping and rapid karyotypic evolution in the genus Helianthus. Genetics **167:** 449–457.

Burke, J. M., S. J. Knapp and L. H. Rieseberg, 2005 Genetic consequences of selection during the evolution of cultivated sunflower. Genetics **171:** 1933–1940.

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

Clark, A. G., 1990 Inference of haplotypes from PCR-amplified samples of diploid populations. Mol. Biol. Evol. **7:** 111–122.

Clark, R. M., E. Linton, J. Messing and J. F. Doebley, 2004 Pattern of diversity in the genomic region near the maize domestication gene *tb1*. Proc. Natl. Acad. Sci. USA **101:** 700–707.

Crites, G. D., 1997 Domesticated sunflower in fifth millennium B.P. temporal context: new evidence from Middle Tennessee. Am. Antiq. **58:** 146–148.

Cronn, R., M. Brothers, K. Klier and P. K. Bretting, 1997 Allozyme variation in domesticated annual sunflower and its wild relatives. Theor. Appl. Genet. **95:** 532–545.

Diwan, N. and P. B. Cregan, 1997 Automated sizing of fluorescent-labeled simple sequence repeat (SSR) markers to assay genetic variation in soybean. Theor. Appl. Genet. **95:** 723–733.

Dvornyk, V., A. Sirviö, M. Mikkonen and O. Savolainen, 2002 Low nucleotide diversity at the *pall* locus in the widely distributed *Pinus sylvestris*. Mol. Biol. Evol. **19:** 179–188.

Eyre-Walker, A., R. L. Gaut, H. Hilton, D. L. Feldman and B. S. Gaut, 1998 Investigation of the bottleneck leading to the domestication of maize. Proc. Natl. Acad. Sci. USA **95:** 4441–4446.

Flint-García, S. A., J. M. Thornsberry and E. S. Buckler, IV, 2003 Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. **54:** 357–374.

García-Gil, M. L., M. Mikkonen and O. Savolainen, 2003 Nucleotide diversity at two phytochrome loci along a latitudinal cline in *Pinus sylvestris*. Mol. Ecol. **12:** 1195–1206.

Garris, A. J., S. R. McCouch and S. Kresovich, 2003 Population structure and its effects on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.) Genetics **165:** 759–769.

Hamblin, M. T., S. E. Mitchell, G. M. White, J. Gallego, R. Kukatla *et al.*, 2004 Comparative population genetics of the panicoid grasses: sequence polymorphism, linkage disequilibrium and selection in a diverse sample of *Sorghum bicolor*. Genetics **167:** 471–483.

Hanson, M. A., B. S. Gaut, A. O. Stec, S. I. Fuerstenberg, M. M. Goodman *et al.*, 1996 Evolution of anthocyanin biosynthesis in maize kernels: the role of regulatory and enzymatic loci. Genetics **143:** 1395–1407.

Harlan, J. R., 1984 Gene centers and gene utilization in American agriculture, pp. 111–129 in *Plant Genetic Resources: A Conservation Imperative*, edited by C. W. Yeatman, D. Kafton and G. Wilkes. AAAS Selected Symposium 87, Westview Press, Boulder, CO.

Harter, A. V., K. A. Gardner, D. Falush, D. L. Lentz, R. A. Bye *et al.*, 2004 Origin of extant domesticated sunflowers in eastern North America. Nature **430:** 201–205.

Heiser, C. B., 1954 Variation and subspeciation in the common sunflower, *Helianthus annuus*. Am. Midl. Nat. **51:** 387–405.

Heiser, C. B., 1955 The origin and development of cultivated sunflower. Am. Biol. Teach. **17:** 161–167.

Heiser, C. B., 1985 Some botanical considerations of the early domesticated plants north of Mexico, pp. 57–72 in *Prehistoric Food Production in North America*, edited by R. Ford. Anthropological Paper 75, Museum of Anthropology, University of Michigan, Ann Arbor, MI.

Hill, W. G., and B. S. Weir, 1988 Variances and covariances of squared linkage disequilibria in finite populations. Theor. Popul. Biol. **33:** 54–78.

Hudson, R. R., 2001 Two-locus sampling distributions and their application. Genetics **159:** 1805–1817.

Hudson, R. R., and N. L. Kaplan, 1985 Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics **111:** 147–164.

Ingvarsson, P. K., 2005 Nucleotide polymorphism and linkage disequilibirum within and among natural populations of European aspen (*Populus termula* L., Salicaceae). Genetics **169:** 945–953.

Kado, T., H. Yoshimaru, Y. Tsumura and H. Tachida, 2003 DNA variation in a conifer, *Cryptomeria japonica* (Cupressaceae *sensu lato*). Genetics **164:** 1547–1599.

Kamiya, N., J. I. Itoh, A. Morikama, Y. Nagato and M. Matsuoka, 2003 The SCARECROW gene's role in asymmetric cell divisions in rice plants. Plant J. **36:** 45–54.

Lentz, D. L., M. E. D. Pohl, K. O. Pope and A. R. Wyatt, 2001 Prehistoric sunflower (*Helianthus annuus* L.) domestication in Mexico. Econ. Bot. **55:** 370–376.

Lin, J.-Z., A. H. D. Brown and M. T. Clegg, 2001 Heterogeneous geographic patterns of nucleotide sequence diversity between two alcohol dehydrogenase genes in wild barley (*Hordeum vulgare* subspecies *spontaneum*). Proc. Natl. Acad. Sci. USA **98:** 531–536.

Long, A. D., and C. H. Langley, 1999 The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res. **9:** 720–731.

Macdonald, S. J., T. Pastinen and A. D. Long, 2005 The effect of polymorphisms in the *Enhancer of split* gene complex on bristle number variation in a large wild-caught cohort of *Drosophila melanogaster*. Genetics **171:** 1741–1756.

Morrell, P. L., D. M. Toleno, K. E. Lundy and M. T. Clegg, 2005 Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *spontaneum*) despite high rates of self-fertilization. Proc. Natl. Acad. Sci. USA **102:** 2442–2447.

Nordborg, M., 2000 Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. Genetics **154:** 923–929.

Nordborg, M., J. O. Borevitz, J. Bergelson, C. C. Berry, J. Chory *et al.*, 2002 The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat. Genet. **30:** 190–193.

Olsen, K. M., and B. A. Schaal, 1999 Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. Proc. Natl. Acad. Sci. USA **96:** 5586–5591.

Pollack, E., 1987 On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics **117:** 353–360.

Putt, E. D., 1997 Early history of sunflower, pp. 1–19 in *Sunflower Production and Technology*, edited by A. A. Schneiter. American Society of Agronomy, Madison, WI.

Rambaut, A. 1996 *Se-Al: Sequence Alignment Editor* (http://evolve.zoo.ox.ac.uk/).

Ramos-Onsins, S. E., B. E. Stranger, T. Mitchell-Olds and M. Aguadé, 2004 Multilocus analysis of variation and specialization in the closely relates species *Arabidopsis halleri* and *A. lyrata*. Genetics **166:** 373–388.

Remington, D. L., J. M. Thornsberry, Y. Matsuoka, L. M. Wilson, S. R. Whitt *et al.*, 2001 Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. USA **98:** 11479–11484.

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

Rodriguez Milla, M., A. Maurer, A. Rodriguez Huete and J. P. Gustafson, 2003 Glutathione peroxidase genes in *Arabidopsis* are ubiquitious and regulated by abiotic stresses through diverse signaling pathways. Plant J. **36:** 602–615.

Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:** 174–175.

Savolainen, O., C. H. Langley, B. P. Lazzaro and H. Fréville, 2000 Contrasting patterns of nucleotide polymorphism at the alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the selfing *Arabidopsis thaliana*. Mol. Biol. Evol. **17:** 645–655.

Swofford, D. L., 2002 *PAUP* 4.0 b10: Phylogenetic Analysis Using Parsimony (* and Other Methods)*. Sinauer, Sunderland, MA.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

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

Tanksley, S. D., and S. R. McCouch, 1997 Seed banks and molecular maps: unlocking genetic potential from the wild. Science **277:** 1063–1066.

Tenaillon, M. I., M. C. Sawkins, A. D. Long, R. L. Gaut, J. F. Doebley *et al.*, 2001 Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc. Natl. Acad. Sci. USA **98:** 9161–9166.

Tenaillon, M. I., M. C. Sawkins, L. K. Anderson, S. M. Stack, J. F. Doebley *et al.*, 2002 Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Genetics **162:** 1401–1413.

Tenaillon, M. I., J. U'Ren, O. Tenaillon and B. Gaut, 2004 Selection versus demography: a multilocus investigation of the domestication process in maize. Mol. Biol. Evol. **21:** 1214–1225.

Thornton, K., 2003 libsequence: a C++ class library for evolutionary genetic analyses. Bioinformatics **19:** 2325–2327.

Tiffin, P., and B. S. Gaut, 2001 Sequence diversity in the tetraploid *Zea perennis* and the closely related diploid *Z. diploperennis*: insights from four nuclear loci. Genetics **158:** 401–412.

Vigouroux, Y., J. S. Jaqueth, Y. Matsuoka, O. S. Smith, W. D. Beavis *et al.*, 2002 Rate and pattern of mutation at microsatellite loci in maize. Mol. Biol. Evol. **19:** 1251–1260.

Wall, J. D., 1999 Recombination and the power of statistical tests of neutrality. Genet. Res. **74:** 65–79.

Weir, B. S., 1990 *Genetic Data Analysis*. Sinauer, Sunderland, MA.

Weir, B. S., and W. G. Hill, 1986 Nonuniform recombination within the human beta-globin gene cluster. Am. J. Hum. Genet. **38:** 776–781.

White, S. E., and J. F. Doebley, 1999 The molecular evolution of *terminal ear1*, a regulatory gene in the genus *Zea*. Genetics **153:** 1455–1462.

Whitt, S. R., L. M. Wilson, M. I. Tenaillon, B. Gaut and E. S. Buckler, IV, 2002 Genetic diversity and selection in the maize starch pathway. Proc. Natl. Acad. Sci. USA **99:** 12959–12962.

Wright, S., B. Lauga and D. Charlesworth, 2003 Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. Mol. Ecol. **12:** 1247–1263.

Zhu, Y. L., Q. J. Song, D. L. Hyten, C. P. Van Tassell, L. K. Matukumalli *et al.*, 2003 Single-nucleotide polymorphisms in soybean. Genetics **163:** 1123–1134.

Communicating editor: A. H. D. Brown