The Plant Cell, Vol. 20: 2931–2945, November 2008, www.plantcell.org © 2008 American Society of Plant Biologists

RESEARCH ARTICLES

A Genomic Scan for Selection Reveals Candidates for Genes Involved in the Evolution of Cultivated Sunflower (*Helianthus annuus*) [W]

Mark A. Chapman,[a,b] Catherine H. Pashley,[b,1] Jessica Wenzler,[b] John Hvala,[a] Shunxue Tang,[c] Steven J. Knapp,[c] and John M. Burke[a,b,2]

[a] Department of Plant Biology, University of Georgia, Athens, Georgia 30602
[b] Department of Biological Sciences, Vanderbilt University, Nashville, Tennessee 37235
[c] Center for Applied Genomic Technologies, University of Georgia, Athens, Georgia 30602

Genomic scans for selection are a useful tool for identifying genes underlying phenotypic transitions. In this article, we describe the results of a genome scan designed to identify candidates for genes targeted by selection during the evolution of cultivated sunflower. This work involved screening 492 loci derived from ESTs on a large panel of wild, primitive (i.e., landrace), and improved sunflower (*Helianthus annuus*) lines. This sampling strategy allowed us to identify candidates for selectively important genes and investigate the likely timing of selection. Thirty-six genes showed evidence of selection during either domestication or improvement based on multiple criteria, and a sequence-based test of selection on a subset of these loci confirmed this result. In view of what is known about the structure of linkage disequilibrium across the sunflower genome, these genes are themselves likely to have been targeted by selection, rather than being merely linked to the actual targets. While the selection candidates showed a broad range of putative functions, they were enriched for genes involved in amino acid synthesis and protein catabolism. Given that a similar pattern has been detected in maize (*Zea mays*), this finding suggests that selection on amino acid composition may be a general feature of the evolution of crop plants. In terms of genomic locations, the selection candidates were significantly clustered near quantitative trait loci (QTL) that contribute to phenotypic differences between wild and cultivated sunflower, and specific instances of QTL colocalization provide some clues as to the roles that these genes may have played during sunflower evolution.

INTRODUCTION

The search for genes underlying phenotypic variation can be performed using either top-down or bottom-up genetic approaches (Wright and Gaut, 2004; Ross-Ibarra et al., 2007). In top-down investigations, researchers start with a phenotype of interest and drill down to the underlying genetic basis. This approach can involve positional cloning of quantitative trait loci (QTL) or association analyses targeting particular candidate genes identified based on homology to genes that are known to control the same, or similar, phenotypes in another species (Frary et al., 2000; Thornsberry et al., 2001; Szalma et al., 2005; Wang et al., 2005; Konishi et al., 2006; Li et al., 2006; Salvi et al., 2007). While top-down approaches have been used to successfully dissect phenotypic variation in a variety of taxa, including trait transitions that occurred during crop evolution, they are not without their drawbacks. For example, positional cloning is both costly and labor-intensive, and such efforts have resulted in only a handful of successes in crop systems (reviewed in Doebley et al., 2006). Moreover, while association mapping holds great promise when researchers have a priori knowledge of the genes that are likely to be regulating a trait of interest, such studies can produce a biased picture of the types of genes that are responsible for phenotypic evolution.

By contrast, bottom-up approaches involve the generation and statistical evaluation of population genetic data from across the genome to identify likely targets of past selection. Because selection acts in a locus-specific manner, whereas the effects of migration, inbreeding, and genetic drift are manifested throughout the genome, selective sweeps reduce genetic variation at and around the target locus while leaving the remainder of the genome unaffected (Maynard-Smith and Haigh, 1974; Slatkin, 1995; Innan and Kim, 2004). As such, functionally important genes can, at least in principle, be identified based on observed patterns of genetic variation even in the absence of information as to which trait(s) they regulate. Such bottom-up approaches provide a more or less unbiased view of the molecular basis of phenotypic evolution, though the phenotypically agnostic nature of such analyses means that follow-up investigations are typically required to identify the trait(s) regulated by loci exhibiting the signature of selection.

Genomic scans for selection have previously been used to search for regions of the genome that were targeted by selection

[1] Current address: Aerobiology Unit, c/o Biology Department, University of Leicester, Adrian Building, University Road, Leicester, LE1 7RH, UK.
[2] Address correspondence to jmburke@uga.edu.
The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) is: John M. Burke (jmburke@uga.edu).
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.108.059808

during the evolution of both maize (*Zea mays*) and sorghum (*Sorghum bicolor*) and have been met with mixed success. In the case of maize, population genetic analyses of gene-based simple sequence repeats (SSRs) and DNA sequence variation have resulted in the identification of ~50 genes that show evidence of selection during the evolution of maize (Vigouroux et al., 2002; Wright et al., 2005; Yamasaki et al., 2005). By contrast, a screen of 74 anonymous SSR loci in sorghum suggested that variation at ~15% of such loci may have been influenced by selection during domestication (Casa et al., 2005), though subsequent sequence-based investigations have failed to identify candidates for selectively important genes (Hamblin et al., 2004, 2006). In this article, we report the results of a genomic scan for selection during the evolution of cultivated sunflower (*Helianthus annuus*) using a large collection of gene-based SSR markers.

Sunflower is a globally important oilseed crop and also a major source of confectionery seeds and ornamental flowers (Putt, 1997). Derived from wild *H. annuus*, cultivated sunflower was domesticated more than 4000 years ago in what is now the central United States (Heiser et al., 1969; Smith, 1989; Rieseberg and Seiler, 1990; Harter et al., 2004). Although they are considered to be members of the same species, wild and cultivated sunflower exhibit a number of striking phenotypic differences. For example, the self-incompatible common sunflower is characterized by many branches along its entire stem, each with numerous small heads and small achenes (i.e., single-seeded fruits). When disturbed, mature heads release their achenes, or shatter. By contrast, the self-compatible cultivated sunflower is typically characterized by an unbranched growth form topped by a single, large head. Cultivated sunflower achenes, which are relatively large, are retained in the head until harvest and also exhibit greatly reduced germination dormancy. Previous genetic analyses have revealed that these trait differences are influenced by a relatively large number of loci, each of which has a small to moderate phenotypic effect (Burke et al., 2002; Wills and Burke, 2007). This stands in stark contrast with the majority of QTL-based investigations of the evolution of other crops, in which a small number of large-effect loci are responsible for the majority of trait differences (Koinange et al., 1996; Xiong et al., 1999; Cai and Morishima, 2002; Doerge, 2002).

In terms of genetic diversity, recent work has revealed that wild sunflower harbors at least as much nucleotide diversity as has been reported in other wild plant taxa and that cultivated sunflower has retained 40 to 50% of the sequence diversity present in the wild (Liu and Burke, 2006). As might be expected of an obligate outcrosser, linkage disequilibrium (LD) appears to decay extremely rapidly in wild sunflower, reaching negligible levels within a few hundred base pairs. In the case of cultivated sunflower, nonrandom associations appear to persist for 1 to 2 kb (Liu and Burke, 2006; Kolkman et al., 2007). While selection can have a major effect on the extent of LD in specific genomic regions (Palaisa et al., 2004; Olsen et al., 2006), the apparently rapid decay of LD in sunflower suggests that genes bearing the signature of selection may themselves have been targeted by selection, as opposed to simply marking larger genomic regions containing selectively important genes. Here, we describe a detailed analysis of genetic diversity in sunflower based on data from a collection of ~500 gene-based SSRs. Beyond providing insight into genome-wide patterns of genetic diversity in wild and cultivated sunflower, these data allow us to identify candidates for selectively important genes that may have been involved in the evolution of cultivated sunflower. Using a stratified sampling strategy involving wild sunflower, primitive landraces, and improved cultivars, we are further able to investigate the timing of selection and to make inferences regarding the relative proportion of the genome that was targeted by selection during domestication versus improvement.

RESULTS

Genome-Wide Levels of Diversity

A total of 492 EST-SSR loci were amplified from a set of 192 sunflower individuals comprising four individuals from each of 24 wild sunflower populations from across the species range, eight primitive landraces, and 16 improved lines (Table 1). As expected, the average genetic diversity per locus was highest in the wild lines and lowest in the improved lines. Mean expected heterozygosity and allelic richness per locus were 0.65 ± 0.01 (mean ± SE) and 6.58 ± 0.16 in the wild population, 0.43 ± 0.01 and 3.25 ± 0.07 in the primitive lines, and 0.32 ± 0.01 and 2.48 ± 0.05 in the improved lines (Table 2). Forty-three of the 492 loci were monomorphic in the primitive lines, and 85 were monomorphic in the improved lines. Thus, while the wild versus primitive (W-P) comparisons (below) were based on the full set of 492 loci, the primitive versus improved (P-I) comparisons were necessarily based on a reduced set of 449 loci.

Relationship between Wild, Primitive, and Improved Sunflower

The occurrence of multiple domestications would complicate the detection of selection (Yamasaki et al., 2007). The neighbor-joining tree generated from the SSR data (Figure 1), however, demonstrates that wild and cultivated sunflower are genetically distinct and is consistent with the view that sunflower domestication occurred only once (Harter et al., 2004; Wills and Burke, 2006). It also appears that the Havasupai and Hopi landraces form a genetically distinct cluster that is well-differentiated from the remainder of the domesticated populations.

Evidence for Selection

The reduction of variance in repeat number and gene diversity in the W-P and P-I population comparisons were calculated using the lnRV and lnRH statistics developed by Schlötterer (2002) and Schlötterer and Dieringer (2005), respectively. These are both diversity-based ranking statistics that are intended to identify loci in the tails of their respective distributions. In both cases, the pool of negative outliers is expected to be enriched for genes that have experienced selective sweeps. These tests implicitly account for the genome-wide reduction in diversity based on the domestication bottleneck and have been shown to be robust to the violation of a number of assumptions, including deviations from the stepwise mutation model (Schlötterer, 2002; Schlötterer

Table 1. Overview of Accessions Used in This Study

Name	Status	Collection Locale	Plant ID
Ames 14400	Wild	Arizona	PI 649851
Ann-1114	Wild	Arkansas	PI 613727
Ann-995	Wild	California	PI 613732
Ann-2298	Wild	Canada-Alberta	PI 592308
Ann-2310	Wild	Canada-Saskatchewan	PI 592317
Ann-2153	Wild	Colorado	PI 586840
Ann-2093	Wild	Illinois	PI 547168
Ann-1753	Wild	Iowa	PI 597895
A-1473	Wild	Kansas	PI 413027
A-1516	Wild	Mexico-Espana	PI 413067
A-1572	Wild	Mexico-Mayo	PI 413123
Ann-1661	Wild	Minnesota	PI 613745
A-1455	Wild	Missouri	PI 413011
2002	Wild	Montana	PI 531032
Ann-2188	Wild	Nebraska	PI 586865
Ann-2106	Wild	North Dakota	PI 586810
Ames 23238	Wild	Ohio	PI 649853
Ann-886	Wild	Oklahoma	PI 435619
Ames 23940	Wild	South Dakota	PI 649854
Ann-646	Wild	Tennessee	PI 435552
Ames 7442	Wild	Texas	PI 649845
1963	Wild	Utah	PI 531009
1975	Wild	Washington	PI 531016
Ann-2128	Wild	Wyoming	PI 586822
Arikara	Primitive		PI 369357
Havasupai	Primitive		PI 369358
Hidatsa	Primitive		PI 600721
Hopi	Primitive		PI 432504
Maiz de Tejas	Primitive		PI 650646
Maiz Negro	Primitive		PI 650761
Mandan	Primitive		PI 600717
Seneca	Primitive		PI 369360
cmsHA89	Improved		PI 650572
Damaya	Improved		PI 496263
Dong Feng	Improved		PI 496264
Jupiter	Improved		PI 296289
Klein Casares	Improved		PI 650817
Mammoth Russian	Improved		PI 478653
Mennonite	Improved		PI 650650
Peredovik	Improved		PI 372173
Pervenets	Improved		PI 483077
Sundak	Improved		Ames 4114
Sunrise	Improved		PI 162454
Tchernianka Select W-13	Improved		PI 343794
VIR 847	Improved		PI 386230
VK-47	Improved		PI 650467
VNIIMK 1646	Improved		PI 650385
VNIIMK 8931	Improved		PI 340790

Information includes accession name, improvement status, collection locale (where applicable), and USDA plant introduction number.

and Dieringer, 2005). The W-P comparison was used to identify candidates for genes that experienced selection during domestication (i.e., domestication-related genes). As mentioned above, of the 492 loci, 449 retained some level of polymorphism in the primitive population and were thus carried over to the P-I comparison. This second comparison allowed for the identification of candidates for genes that experienced selection during the more recent improvement of sunflower (i.e., in the time since domestication; improvement-related genes).

The nonstandardized lnRV and lnRH values for both the W-P and P-I comparisons were, on average, negative (see Supplemental Data Set 1 online), reflecting an overall loss of diversity across the wild-primitive and primitive-improved transitions. These two parameters also exhibited a significant positive correlation with each other (Figure 2). While it is possible, at least in principle, to identify loci that harbor an excess of variation (i.e., positive outliers), and are thus candidates for genes experiencing balancing selection, the primary goals of this study were to identify candidates for genes that experienced a selective sweep during domestication and/or improvement. As such, our focus was primarily on the identification of negative outliers (i.e., genes that show strongly reduced variation in the derived populations). Because lnRV and lnRH measure different aspects of variation at a particular locus, the joint application of these statistics can reduce the false positive rate by a factor of three (Schlötterer and Dieringer, 2005). Thus, while all significant negative outliers are reported, much of the discussion is restricted to a subset for which there is stronger evidence of selection.

For the W-P comparison, 28 significant ($P \leq 0.05$) lnRV outliers and 30 significant ($P \leq 0.05$) lnRH outliers were identified. Of these, 26 and 22 were negative outliers and were thus candidates for having been the target of a selective sweep. These numbers are well in excess of the number of negative outliers expected by chance (i.e., $0.025 \times 492 = 12.3$ in each tail at $\alpha = 0.05$). Overall, seven genes were identified as negative outliers in both tests (Table 3, Figure 2; see Supplemental Data Set 1 online). For the P-I comparison, 33 significant lnRV outliers and 27 significant lnRH outliers were identified. Of these, 27 and 21 were negative outliers (again, substantially more than the $0.025 \times 449 = 11.2$ expected by chance). Eleven of these genes were identified as negative outliers in both tests (Table 3, Figure 2; see Supplemental Data Set 1 online). Consistent with the hypothesis that these genes have experienced differential selective pressures, F_{ST} was significantly higher for the putatively selected loci identified here versus all other loci for both the W-P and P-I comparisons (*t* test, $P < 0.001$; Figure 3).

Because of the potential for differential selection to produce elevated levels of a population structure (Barton and Bengtsson, 1986; Charlesworth et al., 1997), we also used a distance-based simulation of population differentiation (F_{ST}) to identify candidates for genes under selection. This analysis revealed 12 candidates for positive selection during domestication (W-P comparison) and five during improvement (P-I comparison) at $P \leq 0.05$ (see Supplemental Data Set 1 online). Thirteen of the 17 candidate genes identified by F_{ST} were also identified as candidates by one or both of the two previous tests.

Table 3 lists the genes that were identified as outliers at the 95% significance level in at least one of the three statistical tests and at the 90% significance level in at least one other test. In our view, this list contains the best candidates for genes that experienced selection during the evolution of cultivated sunflower An equal number of genes, 18, were identified as candidates for selection during domestication and improvement (note that

Table 2. Summary of Population Genetic Data

		Wild (n = 96)				Primitive (n = 32)				Improved (n = 64)			
		Mean	SE	Min	Max	Mean	SE	Min	Max	Mean	SE	Min	Max
Domestication	AR	7.43	0.62	3.46	12.86	1.59	0.13	1.00	3.31	1.35	0.09	1.00	1.98
	H_e	0.74	0.02	0.62	0.89	0.06	0.02	0.00	0.36	0.03	0.01	0.00	0.14
Improvement	AR	5.83	0.53	3.11	11.79	2.97	0.17	2.00	4.54	1.24	0.09	1.00	2.14
	H_e	0.67	0.03	0.31	0.88	0.39	0.03	0.15	0.67	0.01	0.01	0.00	0.08
Neutral	AR	6.57	0.16	1.70	19.65	3.33	0.07	1.00	9.64	2.58	0.05	1.00	7.56
	H_e	0.64	0.01	0.05	0.95	0.44	0.01	0.00	0.87	0.34	0.01	0.00	0.80
Total	AR	6.58	0.16	1.70	19.65	3.25	0.07	1.00	9.64	2.48	0.05	1.00	7.56
	H_e	0.65	0.01	0.05	0.95	0.43	0.01	0.00	0.87	0.32	0.01	0.00	0.80

Results are presented for the selection candidates (domestication and improvement), the apparently neutral genes, and the total data set. AR, allelic richness; H_e, expected heterozygosity

c2873 and c3113 are actually derived from the same gene). These genes are referred to as "selection candidates" below, though it is important to recognize that there are a number of other genes that were identified as outliers in just one test (see Supplemental Data Set 1 online).

For the selection candidates from the W-P comparison, gene diversity (and allelic richness) dropped from 0.74 ± 0.02 (7.43 ± 0.62 alleles/locus) to 0.06 ± 0.02 (1.59 ± 0.13 alleles/locus), whereas for those from the P-I comparison, these values dropped from 0.39 ± 0.03 (2.97 ± 0.17 alleles/locus) to 0.01 ± 0.01 (1.24 ± 0.09 alleles/locus) (Table 2, Figure 4). Within the list of selection candidates, those loci that are significant at $P \leq 0.05$ under both lnRV and lnRH (shaded in Table 3 and Figure 2) are considered to be the strongest candidates for genes that were targeted by selection because, as noted above, the joint application of these tests dramatically reduces the likelihood of false-positive results (Schlötterer and Dieringer, 2005). Nonetheless, validation of our results is necessary because factors other than selection (e.g., demography) could be responsible for the extreme lnRH, lnRV, and/or F_{ST} values in some cases.

To confirm that the loci identified on the basis of SSR polymorphism showed evidence for selection at the nucleotide level, we arbitrarily selected three domestication and three improvement candidates for further investigation (Table 4), collected sequence data for each from a panel of wild, primitive, and improved sunflower lines (as well as an outgroup; *Helianthus petiolaris*), and analyzed the resulting data using the maximum likelihood HKA (MLHKA) approach of Wright and Charlesworth (2004). Seven additional loci, selected from those that showed no SSR evidence of selection, were included in this analysis as neutral control loci. For each of the selection candidates under consideration, a strictly neutral model was compared with one in which the candidate locus was deemed under selection. To determine the timing of selection, comparisons were made between the outgroup and wild, primitive, or improved sunflower individuals.

For all three loci that were identified as candidates for selection during crop improvement on the basis of our initial SSR screen (c1236, c1406, and c1921), the MLHKA test confirmed the occurrence of selection during improvement (all $P \leq 0.01$; Table 4). For the three domestication-related genes, one (c4973) was found to have experienced selection during domestication ($P < 0.01$), and the remaining two (c1666 and c5898) showed marginally significant evidence of selection during domestication ($0.05 < P < 0.10$), though both were significant when comparing improved lines against the outgroup (both $P < 0.05$). These latter results may actually be due to ongoing selection across the various stages of the evolution of cultivated sunflower.

Inferred Functions and Gene Ontology Classification

To assess whether certain types of genes were overrepresented in the set of putatively selected loci, the distributions of Gene Ontology (GO) terms (molecular function, biological process, and cellular component) were compared between the full set of loci and the 36 selection candidates. Of the 492 loci, *Arabidopsis thaliana* orthologs were identified for 313, which included 31 of the 36 selected loci. Using Gene-Merge (Castillo-Davis and Hartl, 2003), it was evident that genes encoding proteins with a lyase function (molecular function GO:0016829; P = 0.050) as well as those involved in amino acid metabolic processes (biological process GO:0006520; P = 0.008) were significantly overrepresented in the selection candidates. In addition, genes that encode proteins that are targeted to the mitochondrion were significantly overrepresented (cellular component GO:0005739; P = 0.044). While the Bonferroni corrected P values were not significant, such corrections assume independence of categories, which is clearly not the case when it comes to polyhierarchical databases such as GO. It has thus been argued that such corrections are overly conservative to the point of being counterproductive in these sorts of analyses, making it exceedingly difficult to detect true positives (Zeeberg et al., 2003; Osier et al., 2004). It is noteworthy that a comparable analysis in maize found a pattern very similar to that documented here, with the pool of candidates for selectively important genes being enriched for loci that are thought to play a role in amino acid biosynthesis and/or protein catabolism (Wright et al., 2005).

The top BLAST hits for the 36 selection candidates are listed in Table 3. Some of the putative functions are particularly interesting in relation to sunflower or, more generally, crop evolution. For example, at least three loci are potentially involved in the regulation of flowering time, with an additional two loci potentially being involved in both pathogen response and early seed development.

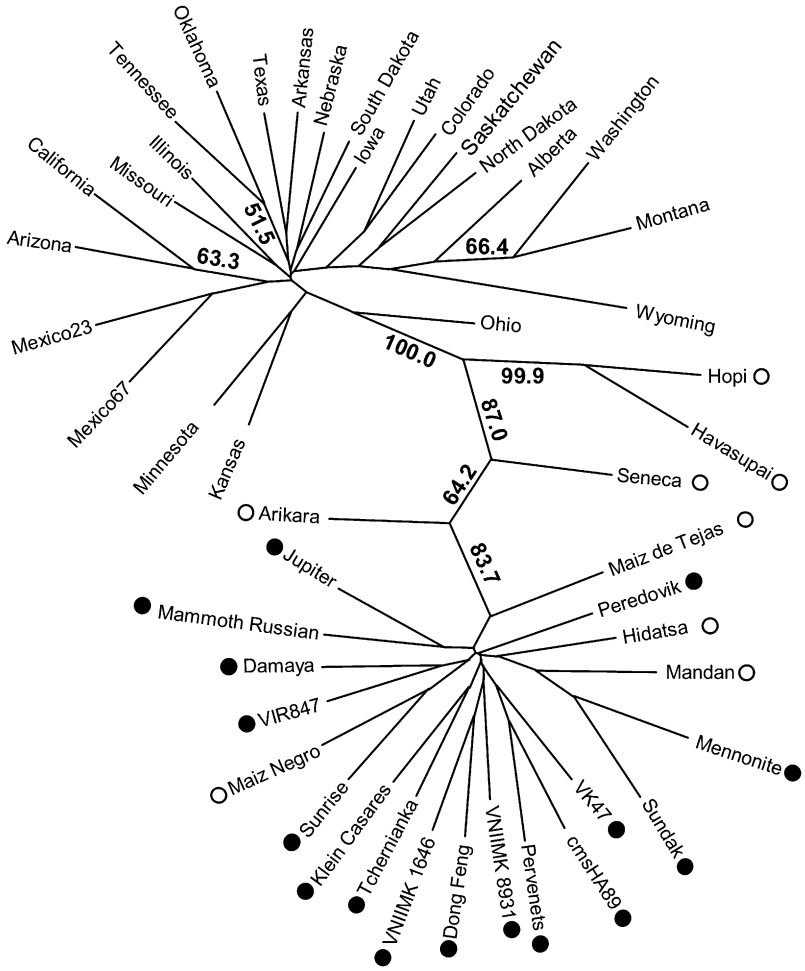

Figure 1. Neighbor-Joining Tree Showing the Relationships between the 48 Sunflower Accessions under Consideration Based on 492 SSR Loci.

Primitive and improved accessions are indicated by open and closed circles, respectively. Numbers alongside branches represent bootstrap values >50% (1000 replicates).

Genetic Mapping

Because a number of domestication- and improvement-related QTL have previously been mapped in sunflower (Burke et al., 2002, 2005; Tang et al., 2006; Wills and Burke, 2007; Baack et al., 2008), the map positions of the selection candidates identified herein were of great interest. Chromosomal locations of 31 of the 36 selection candidates were determined via genetic mapping and are listed in Table 3. The 31 loci were distributed across 12 of the 17 linkage groups (LGs) with between one and six loci on each LG. A subset of our candidate genes were found in the same genomic interval as previously mapped QTL (Burke et al., 2002, 2005; Tang et al., 2006; Wills and Burke, 2007). In fact, 23 of the 27 genes whose positions could be determined relative to the QTL mapped by Wills and Burke (2007) based on shared markers, colocalized with at least one QTL based on their one-LOD (log of the odds) confidence intervals. Given that an estimated 43% of the genome is covered by QTL (again, based one-LOD confidence intervals; Wills and Burke 2007), this finding indicates that the selection candidates are significantly associated with QTL clusters ($P < 0.0001$). One particularly intriguing example is the region surrounding markers ORS331 and ORS143 on LG7, which harbors five selection candidates and also contains QTL for flowering time and the number of main-stem leaves produced (Figure 5; Burke et al., 2002; Wills and Burke 2007). Additionally, four selection candidates mapped to the interval between markers ORS878 and ORS613 on LG10, a region that contains QTL for seed size in three different mapping populations as well as numerous other traits (Burke et al., 2002; Tang et al., 2006; Wills and Burke, 2007; Figure 5).

DISCUSSION

Genetic Diversity and Relatedness

Population bottlenecks are predicted to result in a genome-wide reduction in genetic diversity in domesticated species (Tanksley

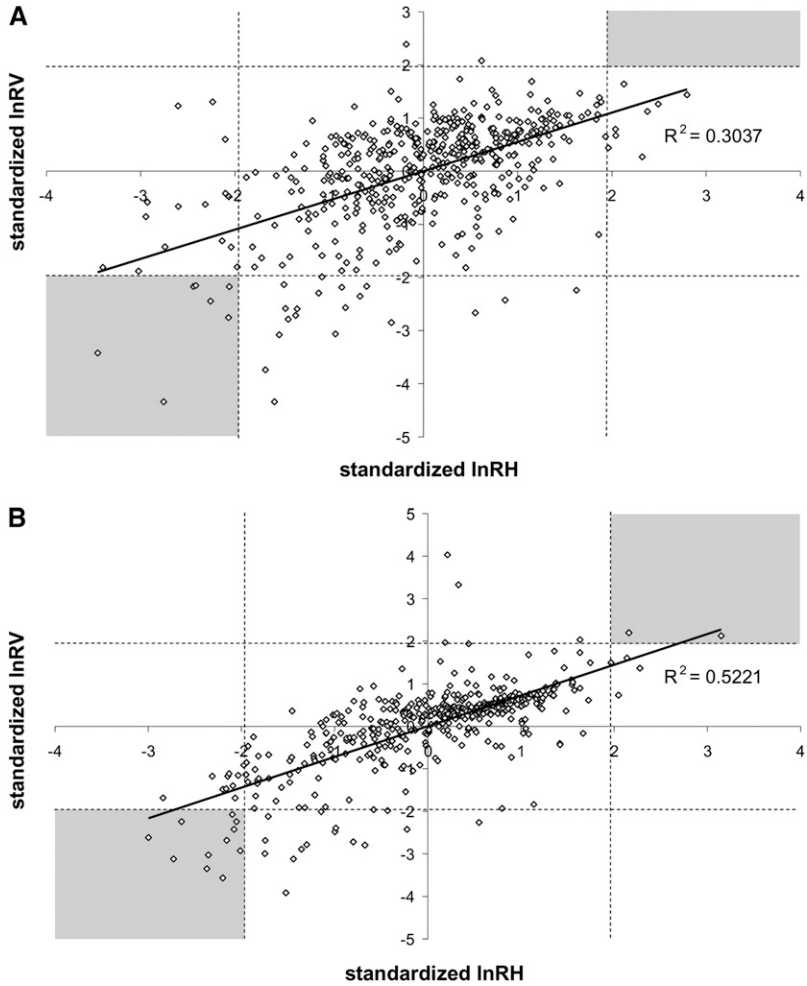

Figure 2. The Relationship between Standardized lnRH and Standardized lnRV Values.

(A) and **(B)** depict the values for the wild-primitive (n = 492 loci) and primitive-improved (n = 449 loci) comparisons, respectively. Broken lines indicate significance at the 95% level, and the gray boxes indicate the regions in which loci are significant outliers at the 95% level for both tests. lnRH is the ratio of heterozygosity in the derived and ancestral populations; lnRV is the ratio of variance in repeat number.

and McCouch, 1997; Yamasaki et al., 2005; Burke et al., 2007). Consistent with this expectation, and with previous findings both in sunflower (Tang and Knapp, 2003; Liu and Burke, 2006) and in other crops (Olsen and Schaal, 2001; Vigouroux et al., 2002; Casa et al., 2005; Caicedo et al., 2007; Sangiri et al., 2007; Zhu et al., 2007), we found that genetic diversity was highest among the wild lines and lowest among the improved lines, with the primitive domesticates being intermediate. There was, however, substantial variation in the amount of diversity lost across loci. This latter observation is presumably due to both sampling variation and differential selective pressures in different genomic regions. The DNA sequence polymorphism data also exhibited a reduction in diversity in wild versus primitive and primitive versus improved comparisons (Table 4).

In terms of SSR differentiation, the primitive and improved lines were genetically more similar to one another than the primitive lines were to their wild counterparts, as evidenced by the lower F_{ST} value in the former comparison relative to the latter (0.071 ± 0.004 [mean ± SE] versus 0.140 ± 0.006 for unselected loci; Figure 3). In terms of phylogenetic relationships among lines, the cultivars all fell into a single clade with 100% bootstrap support (Figure 1), which is in accordance with the view that sunflower is the product of a single domestication (Harter et al., 2004; Wills and Burke, 2006).

Evidence of Selection

The premise underlying our screen for selection candidates is that genes that were targeted by selection during the evolution of cultivated sunflower should exhibit a significantly greater reduction in diversity compared with neutral genes and that directional selection will result in elevated levels of differentiation when comparing between ancestral and derived populations. By comparing wild versus primitive and primitive versus improved lines, we were further able to make inferences regarding the timing of selection. This is an important point in the context of

Table 3. Overview of the Selection Candidates

Locus	Timing	lnRV	lnRH	*F*st	LG	Result of Homology Search
c1666	**D**	***	***	**	**14**	**AK117824.1 *Arabidopsis* putative Ser/Thr protein kinase**
c5898	**D**	***	***	*	**10**	**NM_120482.2 *Arabidopsis* unknown protein**
N21O05	**D**	***	**	–	–	**AF182079 *Matricaria chamomilla* thiol protease**
c3115	**D**	**	**	*	**12**	**AY214336.1 *Medicago truncatula* nicotinate phosphoribosyltransferase-like protein**
c4973	**D**	**	**	–	–	**NM_103779.4 *Arabidopsis* EMBRYO DEFECTIVE 1144; chorismate synthase**
H4B03	**D**	**	**	–	–	**NM_120065.2 *Arabidopsis* kinesin-related protein (MKRP2)**
c1533	**D**	**	**	–	**7**	**DQ661742.1 *Populus tremula* microtubule-associated protein (MAP20)**
c0211	D	*	***	**	14	NM_125464.2 *Arabidopsis* zinc finger (C3HC4-type RING finger) family protein
c1357	D	*	***	*	16	NM_117991.2 *Arabidopsis* pentatricopeptide repeat-containing protein
c2873[a]	D	*	**	**	10[b]	DQ256293 *Sesamum indicum* 11S globulin precursor
c3113[a]	D	**	–	**	NA	DQ256293 *S. indicum* 11S globulin precursor
c0097	D	***	*	–	5	No significant similarity
c2963	D	*	***	–	14	AF406702.1 *Solanum tuberosum* BEL1-related homeotic protein 29
M23M12	D	*	**	–	3	AY490253.1 *Solanum lycopersicum* CONSTANS 3
N5M02	D	*	**	–	7	NM_102939 *Arabidopsis* secretory carrier membrane protein
G13K16	D	***	–	**	14	NM_113590.3 *Arabidopsis* catalytic mRNA
c1649	D	–	**	*	11	NP_190988.1 *Arabidopsis* putative protein
G4G12	D	–	**	*	13[b]	No significant similarity
B12L21	D	*	–	**	10[b]	No significant similarity
c1700	**I**	***	***	–	**10**	**AM236862.1 *Arabidopsis* mitochondrial dicarboxylate carrier**
c2150	**I**	***	***	–	**9[b]**	**NP_175583.1 *Arabidopsis* NADP-specific glutatamate dehydrogenase**
c1236	**I**	**	***	–	**15**	**NP_564307.1 *Arabidopsis* NSL1 (NECROTIC SPOTTED LESIONS1)**
c1774	**I**	***	**	–	**1**	**No significant similarity**
c1921	**I**	***	**	–	**7**	**DQ857278.1 *Glycine max* Dof27**
c0019	**I**	***	**	–	**12**	**NM_105746.3 *Arabidopsis* unknown protein**
c2588	**I**	***	**	–	**7**	**NM_112234.2 *Arabidopsis* ATIDD11 (INDETERMINATE-DOMAIN11)**
J22O06	**I**	***	**	–	**11**	**NM_124978.3 *Arabidopsis* unknown protein**
c1144	**I**	**	**	–	**3**	**NM_118712.4 *Arabidopsis* calmodulin-binding protein**
c1406	**I**	**	**	–	**7**	**AY395743.1 *Vitis aestivalis* protein kinase-like protein**
c5666	**I**	**	**	–	**14**	**AY208699.1 *Artemisia annua* peroxidase 1 (POD1)**
c3070	I	*	***	–	–	AB007819.1 *Citrus unshiu* gene for Gly-rich RNA binding protein
I3D18	I	***	*	–	–	No significant similarity
L2K11	I	***	*	–	10	NPL428214 *Nicotiana plumbaginifolia* mRNA for SDL-1 protein
c1258	I	**	*	–	4	AF091842 *S. indicum* strain Tainan 1 11S globulin precursor
J8F14	I	*	**	–	10	No significant similarity
M18F17	I	–	**	*	1	No significant similarity
N2K13	I	**	–	***	16	NM_179410.3 *Arabidopsis* vacuolar protein sorting 55 family protein

Information on the apparent timing of selection (D = domestication, I = improvement), results of the lnRV, lnRH, and F_{ST} tests, mapped LG, and the results of the homology search wherein the best nonsunflower BLAST hits are reported. Rows in bold indicate loci with $P < 0.05$ for both lnRV and lnRH. ***$P < 0.01$; **$P < 0.05$; *$P < 0.1$.

[a] Two loci that correspond to the same gene.

[b] A locus that was previously mapped in a different cross (see text for details).

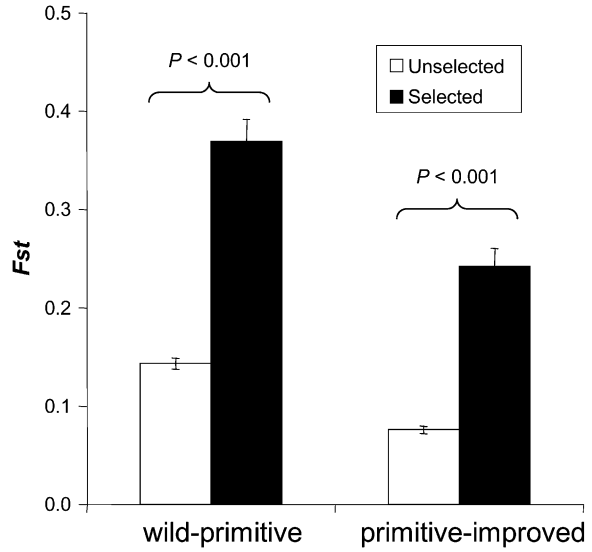

Figure 3. F_{ST} Values for the Selection Candidates Identified on the Basis of a Loss of SSR Diversity versus Unselected Loci in the Wild-Primitive and Primitive-Improved Comparisons.

Standard errors are shown (n = 41 selected and 451 unselected loci in the W-P comparison and n = 37 selected and 412 unselected loci in the P-I comparison). F_{ST} values, genetic differentiation between populations.

ongoing efforts aimed at sunflower improvement, in that improvement-related genes should still be segregating for functional variation in primitive landraces, whereas for domestication-related genes, one would have to look to the wild progenitor (or another related species) for novel alleles (Tanksley and McCouch, 1997; Yamasaki et al., 2005).

Given the large number of loci under consideration, an important caveat is that there are almost certainly some false-positives among the significant outliers, especially those identified as outliers in only one test. This is not, however, a major cause for concern, as our primary goal was to identify candidates for selectively important genes that are worthy of further study. Nonetheless, in the interest of caution, we restrict the balance of the discussion to the so-called selection candidates (Table 3), which are the loci that were identified as outliers in multiple tests. It is noteworthy that our sequence-based analyses of a subset of these selection candidates confirmed the occurrence of selection on all six loci tested, suggesting that the majority of our selection candidates were, in fact, targeted by selection during the evolution of cultivated sunflower (Table 4).

An important consideration in the interpretation of our data is the possible role that genetic hitchhiking may have had in producing the observed results. Because all of the SSRs under consideration were derived from ESTs, we immediately have a good candidate gene that is both known to be expressed and is tightly linked to the SSR in question. Moreover, recent analyses have revealed that LD persists over relatively short distances in sunflower, decaying to negligible levels within ~2 kb in cultivated lineages (Liu and Burke, 2006). This finding suggests that the signature of selection should be closely associated with the actual locus under selection and, by extension, that the selection candidates that we have identified are themselves likely to have been the targets of selection. While selective sweeps can result in a transient increase in the extent of LD in specific genomic regions (Palaisa et al., 2004), previous investigations in other species with low overall levels of LD have largely confirmed that outliers identified in gene-based SSR-based genome scans were themselves targeted by selection (Harr et al., 2002; Vigouroux et al., 2002; DuMont and Aquadro, 2005).

The issue of genetic hitchhiking is equally important when viewed in the context of the genomic distribution of our selection candidates. Do these loci mark 36 independent selective sweeps, or are there clusters of markers associated with a smaller number of sweeps? While some degree of clustering was evident, the loci in question sometimes showed evidence of selection during different time periods, making it unlikely that such instances arose through a single selective event. Moreover, based on an estimated genome size of ~3.5 Gb (Baack et al.,

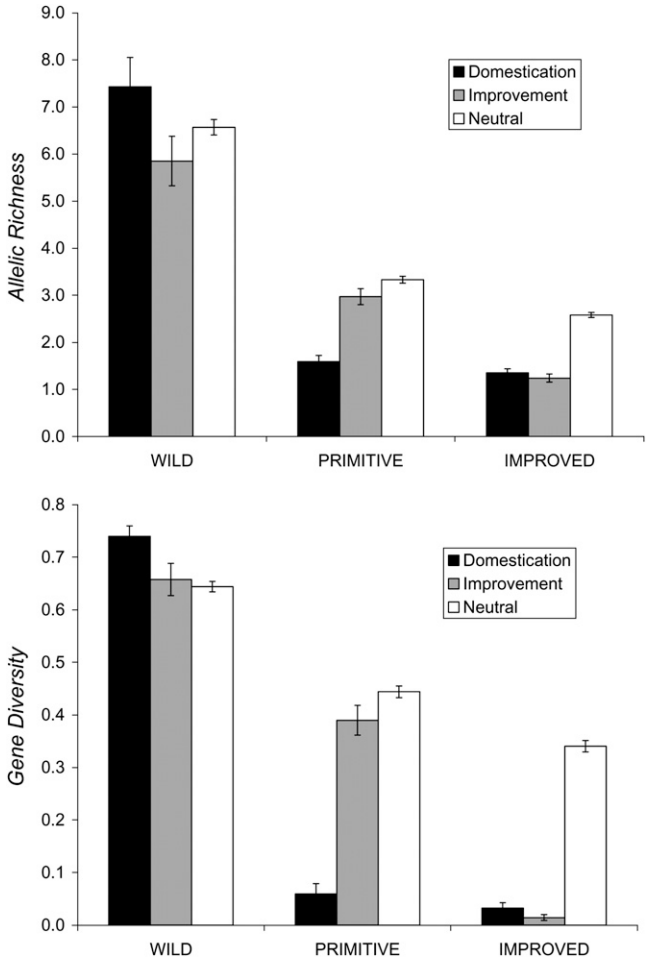

Figure 4. Allelic Richness and Gene Diversity (Expected Heterozygosity) in the Wild, Primitive, and Improved Accessions of Sunflower.

For each population, the values are given for the candidate domestication- (19 loci) and improvement-related (18 loci) loci as well as for the remainder of the loci (neutral; 455 loci). Values reflect mean ± SE.

Table 4. Results of the Sequence-Based Test for Selection

			Wild			Primitive			Improved		
Locus	Candidate Status	*L*	*n*	θ	P	*n*	θ	P	*n*	θ	P
Neutral		535 (82)	15.4 (0.4)	0.0160 (0.0033)	NA	11.7 (0.3)	0.0111 (0.0025)	NA	11.7 (0.3)	0.0086 (0.0020)	NA
c1666	D	677	16	0.0134	0.108	12	0.0139	0.082	12	0.0000	0.045
c4973	D	627	16	0.0084	0.911	12	0.0000	0.005	10	0.0000	0.006
c5898	D	461	16	0.0161	0.770	12	0.0015	0.074	12	0.0007	0.027
c1236	I	1069	16	0.0119	0.540	10	0.0068	0.568	12	0.0000	0.005
c1406	I	1146	16	0.0219	0.207	12	0.0194	0.163	12	0.0000	0.005
c1921	I	1019	16	0.0154	0.948	12	0.0088	0.638	10	0.0010	0.010

Candidate status indicates whether the SSR test suggested selection during domestication (D) or improvement (I). The values for seven putatively neutral genes are averaged (±SE). *L*, length of sequence (bp); *n*, number of sequences; θ, Watterson's estimator of diversity; P, MLHKA P value; NA, not applicable.

2005) and 80 to 85% coverage of the genome in the genetic map that we employed (Baack et al., 2008), 1 centimorgan corresponds to ~2.5 Mb of DNA. Thus, for all but the most tightly linked loci, it seems highly unlikely that genetic hitchhiking played a major role in producing our results.

Insights into the Nature and Frequency of the Selected Genes

While genomic scans of the sort described herein do not provide any direct insight into the phenotypes influenced by genes found to be under selection, there are two types of data available for making inferences about what these genes do. First, we can look at sequence similarity to genes of known effect. Second, we can look at genomic locations relative to previously mapped QTL. The selection candidates that we identified exhibit a wide range of putative functions, including kinases, transferases, transcription factors, and structural proteins (Table 3). The identification of a handful of genes of unknown function is an important point, as these genes would have been entirely missed by a traditional candidate gene-based approach, wherein only those genes showing similarity to genes of known effect are chosen for analysis.

When combining the genetic map locations with putative functions based on homology, two particularly interesting cases emerged. First, two of our selection candidates that show homology to proteins that are known to affect flowering time (c2588 and c1921) map to a region on LG7 that harbors QTL for flowering time in multiple crosses (Figure 5). More specifically, c2588 and c1921 show homology to genes that encode a protein with an INDETERMINATE domain and a Dof-like protein, respectively. Maize *INDETERMINATE1* has previously been shown to regulate the transition to flowering (Colasanti et al., 2006), and a Dof-like protein in *Arabidopsis* represses *CONSTANS*, thereby regulating flowering time (Imaizumi et al., 2005). Second, four selection candidates map to the region surrounding markers ORS878 and ORS613 on LG10. This region harbors QTL for seed size as well as numerous other traits in three different mapping populations (Burke et al., 2002; Tang et al., 2006; Wills and Burke, 2007; Figure 5) and is also known to harbor the classically defined *B* locus, which influences apical branching (Tang et al., 2006). Two of the candidates that map to this region (c2873 and L2K11) have homology to proteins of known function. The former is homologous to a sunflower seed storage protein and was found to be under selection during domestication, whereas the latter is homologous to *seedling lethal-1*, which is necessary for normal seedling growth in *Nicotiana* and *Arabidopsis* (Majira et al., 2002; Pagant et al., 2002; Lertpiriyapong and Sung, 2003) and was found to be under selection during improvement.

Interestingly, both our investigation and a genomic scan for selection focusing on the evolution of maize (Wright et al., 2005) found that loci that putatively play a role in amino acid biosynthesis or protein catabolism were overrepresented in the candidate genes relative to the genome as a whole. Taken together, these findings suggest that selection on amino acid composition may be a general feature of the evolution of crop plants. While the sampling strategy employed by Wright et al. (2005) did not allow them to determine the timing of selection during maize evolution, our analysis shows that these genes were evenly split between domestication and improvement in sunflower.

Future Directions

Despite the apparent success of our approach, a clear limitation of using an SSR-based screen for identifying genes under selection is that SSRs are only present in a fraction of all genes (~9%; range = 2.5 to 21.1%; Ellis and Burke, 2007). Given that our results have validated the utility of bottom-up scans for evidence of selection in sunflower, a promising avenue for future research would be to extend this type of investigation using high-throughput methods aimed directly at assaying nucleotide polymorphism. Such approaches have the potential to rapidly identify a large number of candidates for selectively important genes in a more or less unbiased fashion and could thus greatly enhance our understanding of the genetic changes that occurred during the evolution of cultivated sunflower.

To gain a better understanding of the role that our selection candidates may have played in the evolution of cultivated sunflower, a natural follow-up will be to investigate their patterns of expression. In a recent study in maize, for example, Hufford et al.

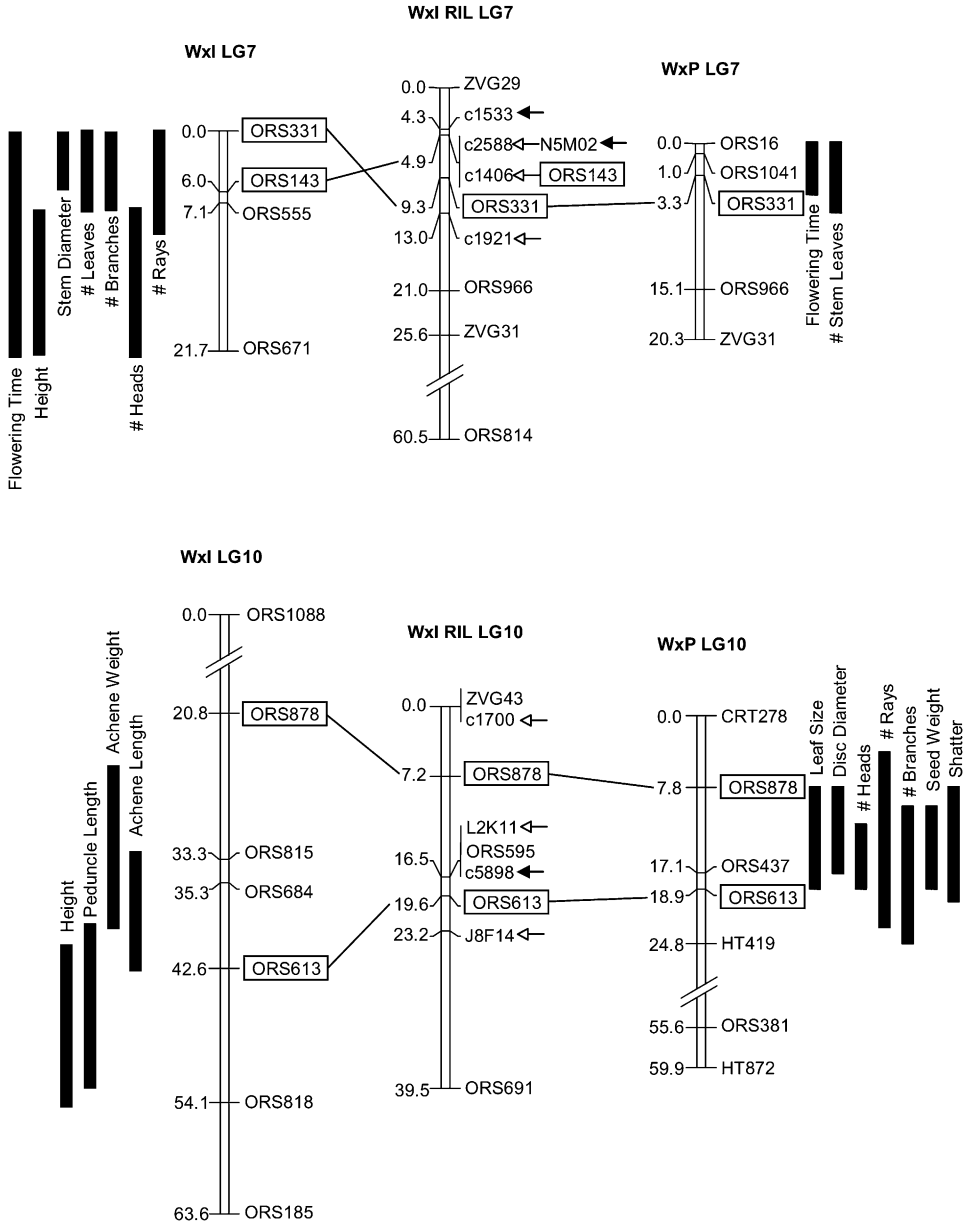

Figure 5. Examples of Candidate Gene and QTL Colocalization.

Linkage groups 7 and 10 of the wild × improved sunflower recombinant inbred lines (WxI RIL; center). Candidate domestication- and improvement-related loci are indicated with solid and open arrows, respectively. The same linkage groups from Burke et al. (2002; WxI, left) and Wills and Burke (2007; WxP, right) are shown to illustrate the QTL present (black boxes) on these linkage groups. Shared markers between maps are indicated with boxes and connected by lines. Data from Burke et al. (2002) and Wills and Burke (2007) are reproduced with permission of the Genetics Society of America.

(2007) investigated the expression of a set of genes that showed evidence of selection during domestication and/or improvement and found that they were significantly overexpressed in the maize ear relative to other tissues. This result is consistent with the likely occurrence of strong selection on ear morphology during the teosinte/maize divergence. It could also be that spatial and/or temporal expression differences are evident when one makes direct comparisons between wild, primitive, and improved lines. This sort of work will be especially enlightening for the genes for which no current function has been ascribed based on sequence similarity.

Finally, a particularly intriguing line of inquiry relates to the finding of apparent selection on genes involved in amino acid biosynthesis and protein catabolism. Given that this pattern has now been documented in both maize and sunflower, it does not appear to be a grass-specific phenomenon. But does this pattern

hold for crop plants in general? Or is it specific to seed crops? And why might these sorts of genes been targeted by selection? It could be that this pattern arose due to conscious selection for increased palatability. Alternatively, selection on these sorts of genes could be a byproduct of unconscious selection on other traits, such as seed dormancy/germination or seedling vigor (Heiser, 1988; Chibani et al., 2006; Reuzeau and Cavalié, 1997). The investigation of other crops, including leaf, tuber, and fruit crops, as well as the functional characterization of the selection candidates identified in such studies, has the potential to unlock these mysteries.

METHODS

Sampling Strategy and Plant Materials

The 48 sunflower (*Helianthus annuus*) accessions used in this study were obtained as seed from the USDA North Central Regional Plant Introduction Station (NCRPIS; Table 1). In an effort to capture as much of the genetic variability in wild sunflower as possible, these accessions were selected from a geographically broad area across North America, including 20 accessions from throughout the US, two accessions from Mexico, and two from Canada. The cultivated accessions consisted of eight Native American landraces, representing the most primitive sunflower domesticates available (Heiser, 1951; Rieseberg and Seiler, 1990), and 16 improved lines. Taken together, these represented 10 of the 12 subsets that make up the NCRPIS *H. annuus* core collection. Because most of these lines exhibit residual heterozygosity, multiple individuals were sampled per accession. Seeds from each accession were clipped on their cotyledon end to enhance germinability and then germinated on moist filter paper before being transferred to potting soil. Total DNA was extracted from four seedlings per accession using the Qiagen DNeasy plant mini kit.

Marker Development and SSR Genotyping

Genes containing SSRs were identified by mining the sunflower portion of the Compositae Genome Project EST Database (CGPDB; http://cgpdb.ucdavis.edu/). The unigene set from the first phase of sunflower EST sequencing was downloaded and searched for SSRs using SSRIT (Temnykh et al., 2001). Our search criteria resulted in the identification of 2360 unigenes (i.e., contigs or singletons) that contained $\geq$5 di-, $\geq$4 tri-, or $\geq$3 tetranucleotide repeats. Primers flanking SSRs were designed for 1728 unigenes using primer3 (Rozen and Skaletsky, 2000). To select loci for amplification across the full panel of 192 individuals, we performed two tests to determine the utility of each primer pair. First, we attempted to amplify each of the 1728 loci (see below for PCR conditions) from a subset of 12 wild sunflower individuals selected from throughout the species' range. Primer pairs that failed to produce an amplicon in at least two-thirds of the tested individuals (as evident from agarose gel electrophoresis) were discarded, as were those that obviously amplified multiple loci. For the remaining loci, we tested for polymorphism by running the samples on an automated DNA sequencer. Loci that were monomorphic across the sample of 12 wild sunflower DNAs were discarded, as were those that provided inconsistent amplification or produced unscorable banding patterns.

All loci were amplified using PCR. Instead of directly labeling each primer for visualization, a modified version of the three primer method of Schuelke (2000) was used (Wills et al., 2005). Each reaction contained 10 ng of template DNA, 30 mM Tricine pH 8.4-KOH, 50 mM KCl, 2 mM $MgCl_2$, 100 μM each deoxynucleotide triphosphate, 0.02 μM forward primer (with an M13 -29 sequence tail [CACGACGTTGTAAAACGACA]), 0.1 μM reverse primer, 0.1 μM fluorescently labeled M13 primer, and one unit of *Taq* DNA polymerase. The fluorescent labels included HEX, 6FAM, VIC, and TET. Cycling conditions followed a touchdown protocol as follows: initial denaturation at 95°C for 3 min; followed by 10 cycles of 30 s at 94°C, 30 s at 65°C (annealing temperature was reduced by 1° per cycle), and 45 s at 72°C; followed by 30 cycles of 30 s at 94°C, 30 s at 55°C, and 45 s at 72°C; and a final extension time of 20 min at 72°C. Amplicons were diluted 1:50 or 1:150 (depending on product intensity in the original screen) in deionized water and visualized on a BaseStation automated DNA sequencer (MJ Research) or an ABI 3730xl DNA sequencer (Applied Biosystems) with MapMarker 1000 ROX size standards (BioVentures) included in each lane to allow for accurate fragment size determination. Alleles were called using the software package CARTOGRAPHER (MJ Research) or GeneMarker (SoftGenetics). Once suitable markers were identified, they were used to genotype the full set of 192 wild and cultivated sunflower individuals. This approach resulted in the generation of genotypic data from 492 loci (primer information is listed in Supplemental Data Set 2 online).

Population Genetic Analyses

For each locus in each set of lines (wild, primitive, and improved), the number of alleles (A) and expected heterozygosity (H_e; also referred to as gene diversity) were determined using Genetic Data Analysis (P.O. Lewis and D. Zaykin; http://lewis.eeb.uconn.edu/lewishome/software.html), allelic richness (*AR*; a sample-size adjusted measure of the number of alleles) was calculated using HP-RARE (version 1.0; Kalinowski, 2005), and the variance in repeat number (*V*) was estimated using Microsatellite Analyzer (MSA; Dieringer and Schlötterer, 2003). For loci that were monomorphic in the primitive population (and in the improved population when diversity was present in the primitive population) a single heterozygous genotype was added to the data following the methods of Kauer et al. (2003). This results in a meaningful (albeit conservative for our purposes) value of H_e and *V* and also serves as a sample-size correction.

Identification of Putatively Selected Loci

The two statistics for detecting selection based on a loss of diversity were calculated as follows: $\ln RV = \ln(V_{der}/V_{anc})$ (Schlötterer, 2002); $\ln RH = \ln(((1/(1\text{-}H_{der}))^2 - 1)/((1/(1 - H_{anc}))^2 - 1))$ (Schlötterer and Dieringer, 2005), where *V* and *H* correspond to the variance in repeat number and gene diversity, respectively, and *der* and *anc* refer to the derived and ancestral populations being compared (i.e., for W-P, wild = ancestral, primitive = derived; and for P-I, primitive = ancestral and improved = derived). Because these statistics are approximately normally distributed, the probability that a given locus deviates from neutrality can be determined from the density function of a standard normal distribution.

Following the methods of Kauer et al. (2003), the lnRV and lnRH values were standardized by the mean and standard deviation for each comparison, such that the standardized distributions had a mean of zero and a standard deviation of one. After standardization, 95% of loci are expected have values between 1.96 and -1.96, with 2.5% of the loci falling above and 2.5% below these values, with significant outliers being candidates for genes under selection.

Because variation in mutation rates across loci can produce spurious results in tests for outlier loci (Schlotterer et al., 2002), and because di-, tri-, and tetranucleotide repeat motif SSRs may exhibit different mutation rates (Chakraborty et al., 1997), we were concerned about potential biases due to mutation rate variation. We thus tested whether or not our selection candidates differ from the balance of the loci under consideration in terms of their SSR motifs. The results of this test were nonsignificant (χ^2 test; $P > 0.3$ for both domestication- and improvement-related

loci), suggesting that possible differences in mutation rates across repeat motifs did not bias our results.

In addition to identifying candidates based on a loss of diversity (above), a distance-based method was also employed. More specifically, bayesfst.c (Beaumont and Balding, 2004; available from http://www.reading.ac.uk/Statistics/genetics/software.html) was used to investigate population structure and to identify significant F_{ST} outliers (Storz, 2005; Vasemagi et al., 2005). For each locus, 2000 Markov chain Monte-Carlo simulations were performed and outlier loci were identified following the methods of Beaumont and Balding (2004). As for the lnRV and lnRH tests, two versions of this test were performed (i.e., one each for the W-P and P-I comparisons).

Functional Annotation of Putatively Selected Loci

The likely functions of the selection candidates were investigated based on sequence similarity to genes of known function from other study systems using BLASTn and discontinuous megablast searches of the nonredundant National Center for Biotechnology Information Genbank database (Altschul et al., 1997). For some of these genes, no significant similarity was found following our BLAST searches, presumably because the sequence available from the CGPDB was either too short to provide a significant match or consisted primarily (or exclusively) of untranslated region, which is less conserved between species than coding regions (Makalowski et al., 1996; Larizza et al., 2002). To resolve this, genome walking was performed to obtain sufficient coding sequence for problematic loci as follows.

Each locus was amplified from an inbred sunflower line (cmsHA89; PI 650572) using the primers and PCR conditions outlined above, and treated with 4 units Exonuclease I and 0.8 units Shrimp Alkaline Phosphatase (USB) at 37°C for 45 min followed by enzyme denaturation at 80°C for 15 min to prepare for sequencing. BigDye v3.1 (Applied Biosystems) was used for the sequencing reaction following the manufacturer's protocol. Unincorporated dyes were removed from the sequencing reactions via Sephadex cleanup (Amersham), and the sequences were resolved on an ABI 3730xl (Applied Biosystems). From this sequence, primers were designed for the genome walking reactions (see Supplemental Table 1 online), which used a genome walking library of cmsHA89 that was constructed using the GenomeWalker Universal kit (BD Biosciences; now available from Clontech) following the manufacturer's instructions with minor modifications (half-sized reaction volumes and touchdown PCR conditions, as above). PCR products obtained from the genome walking were TA-cloned into pGEM-T vectors (Promega), transformed into competent *Escherichia coli*, and screened for presence of an insert. Positive colonies were sequenced as above except that vector primers (T7 and SP6) were used. In some cases, more than one genome walk was necessary to obtain enough coding region to provide a satisfactory hit to a GenBank sequence.

The selection candidates were further investigated by comparing their putative gene functions with those found in the full set of loci to determine if certain types of genes were overrepresented in our collection of outliers. Initially, the top BLAST hit for all loci was retrieved from the CGPDB (see Supplemental Data Set 1 online). For loci where no hit was recorded, or where the BLAST hit was not to an *Arabidopsis thaliana* protein, additional BLAST searches were performed as above. For all sunflower loci with putative *Arabidopsis* orthologs identified, Gene-Merge (Castillo-Davis and Hartl, 2003) was used to assign the *Arabidopsis* genes a molecular function, cellular component, and biological process following GO terminology (Ashburner et al., 2000). Gene-Merge was then used to calculate the probability that a certain class of genes was overrepresented among the putatively selected genes. Because of the relatively small total number of candidate genes that were identified, the two classes (domestication-related and improvement-related) were combined for this analysis.

Genetic Mapping

All 36 of the selection candidates from the W-P and P-I comparisons were screened for polymorphism using a subset of eight lines from a recombinant inbred line (RIL) population generated by S.J.K. and R.L. Brunick (Oregon State University) from an initial cross between wild sunflower (Ann1238 from Keith Co., Nebraska) and an inbred oilseed cultivar (cmsHA89; for further information, see Burke et al., 2002; Baack et al., 2008). The initial screening followed the SSR genotyping protocol detailed above. In cases where no length polymorphism was detected, each locus was sequenced in an attempt to identify DNA sequence polymorphisms that could be used in PCR-RFLP (restriction fragment length polymorphism) or single-strand conformation polymorphism analyses. In some cases, it was necessary to carry out genome walking (as described above) and generate additional sequence data to find polymorphisms that could be mapped. Once a polymorphism was detected, a given locus was amplified from the full set of 184 RILs and scored as either a length variant (18 loci) or via PCR-RFLP (seven loci) or single-strand conformation polymorphism (two loci) (primer sequences are given in Supplemental Table 1 online). Loci were added to the previously published linkage map using the data and methods of Baack et al. (2008). For nine loci, no polymorphism was detected in the RILs (despite gathering up to 3 kb of sequence data for each locus); however, for four of these, map locations have previously been determined in another mapping population (RHA280 × RHA801; for details, see Tang et al., 2002), and approximate positions could be inferred across maps based on shared markers.

Phylogenetic Analysis

To assess the relationships between the 48 sunflower accessions employed in this study, we used a bootstrapping approach in MSA (Dieringer and Schlötterer, 2003) to generate 1000 distance matrices between all pairs of accessions based on the distance measure of Nei et al. (1983). These matrices were then analyzed in PHYLIP (version 3.67; Felsenstein, 2005) using the NEIGHBOR and CONSENSE functions to generate a bootstrapped neighbor-joining tree.

Sequence-Based Test of Selection

Six of the selection candidates were randomly selected for sequence analysis (three each from both the domestication-related and improvement-related candidate pools). These genes were sequenced from a panel of wild, primitive, and improved individuals, plus *Helianthus petiolaris*, which was included as an outgroup (see Supplemental Table 2 online). In addition, seven genes for which we had no a priori evidence of selection were included as presumptively neutral controls. Each of these seven loci was tested against the other six to confirm neutrality prior to testing the selection candidates (see method below). Primer sequences (see Supplemental Table 1 online) were designed to amplify ~500 to 1200 bp from each locus, and PCR conditions followed the general protocol outlined for SSR genotyping except that, in some instances, the annealing temperature was increased to 60°C and the extension time to 90 s. Sequencing and cloning (where necessary) were performed as above.

Sequence alignments were constructed in Genedoc (K.B. Nicholas and H.B. Nicholas, Jr.; www.psc.edu/biomed/genedoc) and exported to DnaSP version 4.50.2 (Rozas et al., 2003). Individuals containing ambiguous bases were resolved into haplotypes using the PHASE algorithm in DnaSP. Coding and noncoding regions were annotated using the original EST sequences and BLASTn hits to *Arabidopsis* and *Vitis* genome sequences. DnaSP was then used to calculate the number of segregating sites, S, nucleotide diversity (π), number of haplotypes, and Watterson's (1975) estimate of diversity (θ). All sequences have been deposited in the Genbank database (see below).

Tests for departures from neutrality in the candidate loci were performed using the MLHKA (Hudson et al., 1987) test of Wright and Charlesworth (2004). For each locus, three pairs of tests (100,000 simulations each) were performed, each of which involved the seven neutral genes plus one putatively selected locus. The three pairs of tests involved polymorphism data from either the wild versus *H. petiolaris*, primitive versus *H. petiolaris*, or improved versus *H. petiolaris* comparisons. In all cases, a neutral model in which all eight loci (i.e., the selection candidate plus the seven neutral controls) were assumed to be evolving neutrally was run first. In a second run, the seven control loci were assumed to by evolving neutrally, whereas the eighth locus (the selection candidate of interest) was deemed under selection. In each case, significance was evaluated by calculating twice the difference in the likelihoods of the two models. This value is approximately χ^2 distributed with one degree of freedom (Wright and Charlesworth, 2004). Following the methods of Yamasaki et al. (2005) and Hufford et al. (2007), selection during the evolution of cultivated sunflower was tested only when the wild versus *H. petiolaris* test was nonsignificant. This was the case for all six genes tested. Selection during improvement was evidenced by a significant result in just the improved versus *H. petiolaris* test, whereas selection during domestication was evidenced by a significant result in both the primitive versus *H. petiolaris* and improved versus *H. petiolaris* tests, but not in the wild versus *H. petiolaris* test.

Accession Numbers

Sequence data from this article can be found in the GenBank/EMBL data libraries under the following accession numbers: locus c25, FJ373512-FJ373535; locus c1111, FJ373536-FJ373563; locus c1236, FJ373564-FJ373584; locus c1351, FJ373585-FJ373613; locus c1406, FJ373614-FJ373641; locus c1666, FJ373642-FJ373671; locus c1921, FJ373672-FJ373703; locus c2016, FJ373704-FJ373733; locus c2307, FJ373734-FJ373760; locus c4973, FJ373761-FJ373784; locus c5369, FJ373785-FJ373822; locus c5456, FJ373823-FJ373851; locus c5898, FJ373852-FJ373879.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Genetic Map Positions of 12 Domestication-Related Loci and 15 Improvement-Related Loci in a Wild × Improved Sunflower RIL Population.

Supplemental Table 1. Primers Employed in Mapping and Sequencing Experiments.

Supplemental Table 2. Individuals Employed in the Sequence Analysis.

Supplemental Data Set 1. Genetic Giversity and Selection Results.

Supplemental Data Set 2. Primer Sequences.

ACKNOWLEDGMENTS

We thank Natasha Sherman, David Wills, David Baum, and four anonymous reviewers for helpful comments that greatly improved the manuscript and Daniel Feckoury, Melissa Hester, Sarah Kimball, and Matt Wilkins for assistance in the SSR screening. This work was funded by grants to J.M.B. from the National Science Foundation Plant Genome Research Program (DBI-0332411) and the Plant Genome Program of the USDA Cooperative State Research, Education, and Extension Service–National Research Initiative (03-35300-13104).

Received April 1, 2008; revised October 22, 2008; accepted November 4, 2008; published November 18, 2008.

REFERENCES

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Ashburner, M., et al. (2000). Gene Ontology: Tool for the unification of biology. Nat. Genet. **25:** 25–29.

Baack, E.J., Sapir, Y., Chapman, M.A., Burke, J.M., and Rieseberg, L.H. (2008). Selection on domestication traits and quantitative trait loci in crop-wild sunflower hybrids. Mol. Ecol. **17:** 666–677.

Baack, E.J., Whitney, K.D., and Rieseberg, L.H. (2005). Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. New Phytol. **167:** 623–630.

Barton, N., and Bengtsson, B.O. (1986). The barrier to genetic exchange between hybridizing populations. Heredity **57:** 357–376.

Beaumont, M.A., and Balding, D.J. (2004). Identifying adaptive genetic divergence among populations from genome scans. Mol. Ecol. **13:** 969–980.

Burke, J.M., Burger, J.C., and Chapman, M.A. (2007). Crop evolution: From genetics to genomics. Curr. Opin. Genet. Dev. **17:** 525–532.

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

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

Cai, H.W., and Morishima, H. (2002). QTL clusters reflect character associations in wild and cultivated rice. Theor. Appl. Genet. **104:** 1217–1228.

Caicedo, A.L., Williamson, S.H., Hernandez, R.D., Boyko, A., Fledel-Alon, A., York, T.L., Polato, N.R., Olsen, K.M., Nielsen, R., McCouch, S.R., Bustamante, C.D., and Purugganan, M.D. (2007). Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLOS Genetics **3:** 1745–1756.

Casa, A.M., Mitchell, S.E., Hamblin, M.T., Sun, H., Bowers, J.E., Paterson, A.H., Aquadro, C.F., and Kresovich, S. (2005). Diversity and selection in sorghum: Simultaneous analyses using simple sequence repeats. Theor. Appl. Genet. **111:** 23–30.

Castillo-Davis, C.I., and Hartl, D.L. (2003). GeneMerge - Post-genomic analysis, data mining, and hypothesis testing. Bioinformatics **19:** 891–892.

Chakraborty, R., Kimmel, M., Stivers, D.N., Davison, L.J., and Deka, R. (1997). Relative mutation rates at di-, tri-, and tetranucleotide microsatellite loci. Proc. Natl. Acad. Sci. USA **94:** 1041–1046.

Charlesworth, B., Nordborg, M., and Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. **70:** 155–174.

Chibani, K., Ali-Rachedi, S., Job, C., Job, D., Jullien, M., and Grappin, P. (2006). Proteomic analysis of seed dormancy in *Arabidopsis*. Plant Physiol. **142:** 1493–1510.

Colasanti, J., Tremblay, R., Wong, A.Y.M., Coneva, V., Kozaki, A., and Mable, B.K. (2006). The maize INDETERMINATE1 flowering time regulator defines a highly conserved zinc finger protein family in higher plants. BMC Genomics **7**: 158.

Dieringer, D., and Schlötterer, C. (2003). Microsatellite Analyser (MSA): A platform independent analysis tool for large microsatellite data sets. Mol. Ecol. Notes **3:** 167–169.

Doebley, J.F., Gaut, B.S., and Smith, B.D. (2006). The molecular genetics of crop domestication. Cell **127:** 1309–1321.

Doerge, R.W. (2002). Mapping and analysis of quantitative trait loci in experimental populations. Nat. Rev. Genet. **3:** 43–52.

DuMont, V.B., and Aquadro, C.F. (2005). Multiple signatures of positive selection downstream of Notch on the tip X chromosome in *Drosophila melanogaster.* Genetics **171:** 639–653.

Ellis, J.R., and Burke, J.M. (2007). EST-SSRs as a resource for population genetic analyses. Heredity **99:** 125–132.

Felsenstein, J. (2004). PHYLIP (Phylogeny Inference Package) Version 3.72. (Seattle: University of Washington).

Frary, A., Nesbitt, T.C., Frary, A., Grandillo, S., van der Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K.B., and Tanksley, S.D. (2000). *fw2.2*: A quantitative trait locus key to the evolution of tomato fruit size. Science **289:** 85–88.

Hamblin, M.T., Casa, A.M., Sun, H., Murray, S.C., Paterson, A.H., Aquadro, C.F., and Kresovich, S. (2006). Challenges of detecting directional selection after a bottleneck: Lessons from *Sorghum bicolor.* Genetics **173:** 953–964.

Hamblin, M.T., Mitchell, S.E., White, G.M., Gallego, J., Kukatla, R., Wing, R.A., Paterson, A.H., and Kresovich, S. (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.

Harr, B., Kauer, M., and Schlötterer, C. (2002). Hitchhiking mapping: A population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster.* Proc. Natl. Acad. Sci. USA **99:** 12949–12954.

Harter, A.V., Gardner, K.A., Falush, D., Lentz, D.L., Bye, R.A., and Rieseberg, L.H. (2004). Origin of extant domesticated sunflowers in eastern North. Am. Nat. **430:** 201–205.

Heiser, C.B., Jr. (1951). The sunflower among North American Indians. Proc. Am. Philos. Soc. **95:** 432–448.

Heiser, C.B., Jr. (1988). Aspects of unconscious selection and the evolution of domesticated plants. Euphytica **37:** 77–81.

Heiser, C.B., Smith, D.M., Clevenger, S., and Martin, W.C. (1969). The North American sunflowers *Helianthus.* Memoirs of the Torrey Botanical Club **22:** 1–218.

Hudson, R.R., Kreitman, M., and Aguade, M. (1987). A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

Hufford, K.M., Canaran, P., Ware, D.H., McMullen, M.D., and Gaut, B.S. (2007). Patterns of selection and tissue-specific expression among maize domestication and crop improvement loci. Plant Physiol. **144:** 1642–1653.

Imaizumi, T., Schultz, T.F., Harmon, F.G., Ho, L.A., and Kay, S.A. (2005). FKF1F-BOX protein mediates cyclic degradation of a repressor of CONSTANS in Arabidopsis. Science **309:** 293–297.

Innan, H., and Kim, Y. (2004). Pattern of polymorphism after strong artificial selection in a domestication event. Proc. Natl. Acad. Sci. USA **101:** 10667–10672.

Kalinowski, S.T. (2005). HP-RARE 1.0: A computer program for performing rarefaction on measures of allelic richness. Mol. Ecol. Notes **5:** 187–189.

Kauer, M.O., Dieringer, D., and Schlötterer, C. (2003). A microsatellite variability screen for positive selection associated with the "Out of Africa" habitat expansion of *Drosophila melanogaster.* Genetics **165:** 1137–1148.

Koinange, E.M.K., Singh, S.P., and Gepts, P. (1996). Genetic control of the domestication syndrome in common bean. Crop Sci. **36:** 1037–1045.

Kolkman, J.M., Berry, S.T., Leon, A.J., Slabaugh, M.B., Tang, S., Gao, W.X., Shintani, D.K., Burke, J.M., and Knapp, S.J. (2007). Single nucleotide polymorphisms and linkage disequilibrium in sunflower. Genetics **177:** 457–468.

Konishi, S., Izawa, T., Lin, S.Y., Ebana, K., Fukuta, Y., Sasaki, T., and Yano, M. (2006). An SNP caused loss of seed shattering during rice domestication. Science **312:** 1392–1396.

Larizza, A., Makalowski, W., Pesole, G., and Saccone, C. (2002). Evolutionary dynamics of mammalian mRNA untranslated regions by comparative analysis of orthologous human, artiodactyl and rodent gene pairs. Comput. Chem. **26:** 479–490.

Lertpiriyapong, K., and Sung, Z.R. (2003). The *elongation defective1* mutant of *Arabidopsis* is impaired in the gene encoding a serine-rich secreted protein. Plant Mol. Biol. **53:** 581–595.

Li, C.B., Zhou, A.L., and Sang, T. (2006). Rice domestication by reducing shattering. Science **311:** 1936–1939.

Liu, A.Z., and Burke, J.M. (2006). Patterns of nucleotide diversity in wild and cultivated sunflower. Genetics **173:** 321–330.

Majira, A., Domin, M., Grandjean, O., Gofron, K., and Houba-Herin, N. (2002). Seedling lethality in *Nicotiana plumbaginifolia* conferred by *Ds* transposable element insertion into a plant-specific gene. Plant Mol. Biol. **50:** 551–562.

Makalowski, W., Zhang, J.H., and Boguski, M.S. (1996). Comparative analysis of 1196 orthologous mouse and human full-length mRNA and protein sequences. Genome Res. **6:** 846–857.

Maynard-Smith, J., and Haigh, J. (1974). Hitch-hiking effect of a favorable gene. Genet. Res. **23:** 23–35.

Nei, M., Tajima, F., and Tateno, Y. (1983). Accuracy of estimated phylogenetic trees from molecular data. 2. Gene-frequency data. J. Mol. Evol. **19:** 153–170.

Olsen, K.M., Caicedo, A.L., Polato, N., McClung, A., McCouch, S., and Purugganan, M.D. (2006). Selection under domestication: Evidence for a sweep in the rice *Waxy* genomic region. Genetics **173:** 975–983.

Olsen, K.M., and Schaal, B.A. (2001). Microsatellite variation in cassava (*Manihot esculenta*, Euphorbiaceae) and its wild relatives: Further evidence for a southern Amazonian origin of domestication. Am. J. Bot. **88:** 131–142.

Osier, M.V., Zhao, H.Y., and Cheung, K.H. (2004). Handling multiple testing while interpreting microarrays with the Gene Ontology database. BMC Bioinformatics **5**: 124.

Pagant, S., Bichet, A., Sugimoto, K., Lerouxel, O., Desprez, T., McCann, M., Lerouge, P., Vernhettes, S., and Hofte, H. (2002). KOBITO1 encodes a novel plasma membrane protein necessary for normal synthesis of cellulose during cell expansion in Arabidopsis. Plant Cell **14:** 2001–2013.

Palaisa, K., Morgante, M., Tingey, S., and Rafalski, A. (2004). Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. Proc. Natl. Acad. Sci. USA **101:** 9885–9890.

Putt, E.D. (1997). Early history of sunflower. In Sunflower Technology and Production, A.A. Schneiter, ed (Madison, WI: American Society of Agronomy), pp. 1–19.

Reuzeau, C., and Cavalié, G. (1997). Changes in RNA and protein metabolism associated with alterations in the germination efficiency of sunflower seeds. Ann. Bot. (Lond.) **80:** 131–137.

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

Ross-Ibarra, J., Morrell, P.L., and Gaut, B.S. (2007). Plant domestication, a unique opportunity to identify the genetic basis of adaptation. Proc. Natl. Acad. Sci. USA **104:** 8641–8648.

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496–2497.

Rozen, S., and Skaletsky, H.J. (2000). Primer3 on the WWW for general users and for biologist programmers. In Bioinformatics Methods and Protocols: Methods in Molecular Biology, S. Krawetz and S. Misener, eds (Totowa, NJ: Humana Press), pp. 365–386.

Salvi, S., et al. (2007). Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus m maize. Proc. Natl. Acad. Sci. USA **104:** 11376–11381.

Sangiri, C., Kaga, A., Tomooka, N., Vaughan, D., and Srinives, P. (2007). Genetic diversity of the mungbean (*Vigna radiata*, Leguminosae) genepool on the basis of microsatellite analysis. Aust. J. Bot. **55:** 837–847.

Schlotterer, C. (2002). A microsatellite-based multilocus screen for the identification of local selective sweeps. Genetics **160:** 753–763.

Schlötterer, C., and Dieringer, D. (2005). A novel test statistic for the identification of local selective sweeps based on microsatellite gene diversity. In Selective Sweep, D. Nurminsky, ed (Boston: Kluwer Academic Publishers), pp. 55–64.

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. Nat. Biotechnol. **18:** 233–234.

Slatkin, M. (1995). Hitchhiking and associative overdominance at a microsatellite locus. Mol. Biol. Evol. **12:** 473–480.

Smith, B.D. (1989). Origins of agriculture in Eastern North America. Science **246:** 1566–1571.

Storz, J.F. (2005). Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol. Ecol. **14:** 671–688.

Szalma, S.J., Buckler, E.S., Snook, M.E., and McMullen, M.D. (2005). Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. Theor. Appl. Genet. **110:** 1324–1333.

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

Tang, S., Yu, J.K., Slabaugh, M.B., Shintani, D.K., and Knapp, S.J. (2002). Simple sequence repeat map of the sunflower genome. Theor. Appl. Genet. **105:** 1124–1136.

Tang, S.X., Leon, A., Bridges, W.C., and Knapp, S.J. (2006). Quantitative trait loci for genetically correlated seed traits are tightly linked to branching and pericarp pigment loci in sunflower. Crop Sci. **46:** 721–734.

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

Temnykh, S., DeClerck, G., Lukashova, A., Lipovich, L., Cartinhour, S., and McCouch, S. (2001). Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): Frequency, length variation, transposon associations, and genetic marker potential. Genome Res. **11:** 1441–1452.

Thornsberry, J.M., Goodman, M.M., Doebley, J., Kresovich, S., Nielsen, D., and Buckler IV, E.S. (2001). *Dwarf8* polymorphisms associate with variation in flowering time. Nat. Genet. **28:** 286–289.

Vasemagi, A., Nilsson, J., and Primmer, C.R. (2005). Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). Mol. Biol. Evol. **22:** 1067–1076.

Vigouroux, Y., McMullen, M., Hittinger, C.T., Houchins, K., Schulz, L., Kresovich, S., Matsuoka, Y., and Doebley, J. (2002). Identifying genes of agronomic importance in maize by screening microsatellites for evidence of selection during domestication. Proc. Natl. Acad. Sci. USA **99:** 9650–9655.

Wang, H., Nussbaum-Wagler, T., Li, B.L., Zhao, Q., Vigouroux, Y., Faller, M., Bomblies, K., Lukens, L., and Doebley, J.F. (2005). The origin of the naked grains of maize. Nature **436:** 714–719.

Watterson, G.A. (1975). On the number of segregating sites in genetic models without recombination. Theor. Popul. Biol. **7:** 256–276.

Wills, D.M., and Burke, J.M. (2006). Chloroplast DNA variation confirms a single origin of domesticated sunflower (*Helianthus annuus* L.). J. Hered. **97:** 403–408.

Wills, D.M., and Burke, J.M. (2007). QTL analysis of the early domestication of sunflower. Genetics **176:** 2589–2599.

Wills, D.M., Hester, M.L., Liu, A., and Burke, J.M. (2005). Chloroplast SSR polymorphisms in the Compositae and the mode of organellar inheritance in *Helianthus annuus.* Theor. Appl. Genet. **110:** 941–947.

Wright, S.I., Bi, I.V., Schroeder, S.G., Yamasaki, M., Doebley, J.F., McMullen, M.D., and Gaut, B.S. (2005). The effects of artificial selection on the maize genome. Science **308:** 1310–1314.

Wright, S.I., and Charlesworth, B. (2004). The HKA test revisited: A maximum-likelihood-ratio test of the standard neutral model. Genetics **168:** 1071–1076.

Wright, S.I., and Gaut, B.S. (2004). Molecular population genetics and the search for adaptive evolution in plants. Mol. Biol. Evol. **22:** 506–519.

Xiong, L.X., Liu, K.D., Dai, X.K., Xu, C.G., and Zhang, Q.F. (1999). Identification of genetic factors controlling domestication-related traits of rice using an F_2 population of a cross between *Oryza sativa* and *O. rufipogon.* Theor. Appl. Genet. **98:** 243–251.

Yamasaki, M., Tenaillon, M.I., Bi, I.V., Schroeder, S.G., Sanchez-Villeda, H., Doebley, J.F., Gaut, B.S., and McMullen, M.D. (2005). A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. Plant Cell **17:** 2859–2872.

Yamasaki, M., Wright, S.I., and McMullen, M.D. (2007). Genomic screening for artificial selection during domestication and improvement in maize. Ann. Bot. (Lond.) **100:** 967–973.

Zeeberg, B.R., et al. (2003). GoMiner: A resource for biological interpretation of genomic and proteomic data. Genome Biol. **4:** R28.

Zhu, Q.H., Zheng, X.M., Luo, J.C., Gaut, B.S., and Ge, S. (2007). Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: Severe bottleneck during domestication of rice. Mol. Biol. Evol. **24:** 875–888.