American Journal of Botany

American Journal of Botany 97(5): 831–840. 2010.

Population genetic analysis of safflower (*Carthamus tinctorius*; Asteraceae) reveals a Near Eastern origin and five centers of diversity[1]

Mark A. Chapman, John Hvala, Jason Strever, and John M. Burke[2]

Department of Plant Biology, University of Georgia, Athens, Georgia 30602 USA

Analyses of genetic variation in crop gene pools are a powerful tool for investigating the origin and early evolution of crop lineages. Such analyses also have the potential to identify unique genetic resources for continued crop improvement. The oilseed crop safflower (*Carthamus tinctorius*) is believed to have been domesticated in the Fertile Crescent region, but up to 10 geographic centers of similarity throughout the world have been proposed based on morphology. Nuclear microsatellite analysis of accessions from each of the 10 proposed centers of similarity, as well as individuals of the progenitor species, suggested the presence of five genetic clusters (1, Europe; 2, Turkey–Iran–Iraq–Afghanistan; 3, Israel–Jordan–Syria; 4, Egypt–Ethiopia; and 5, the Far East–India–Pakistan). North American accessions, products of a secondary introduction from the native range, suggest that a subset of the native accessions harbor unique genetic diversity that could be useful in future breeding efforts. Overall, a Near Eastern origin of safflower was confirmed based on the genetic similarity between the progenitor and the Near Eastern safflower accessions, as well as previous archaeological finds. Genetic differentiation between geographical clusters of accessions is evident, although not to the degree proposed based on morphology.

Key words: Asteraceae; *Carthamus*; crop evolution; domestication; genetic variation; population structure; safflower.

Population genetic analyses of crop gene pools can provide insight into the origin and subsequent evolution of crop plants and can also result in the identification of novel sources of genetic variation for the continued improvement of crop plants. Because of the importance of the Fertile Crescent region in the early origins of agriculture ca. 10 000 yr ago (Zohary and Hopf, 2000), numerous such studies have focused on species that were domesticated in this region. For example, crops such as barley (*Hordeum vulgare* L.; Badr et al., 2000), Emmer wheat (*Triticum dicoccum* Schübl.; Ozkan et al., 2002), Einkorn wheat (*Triticum monococcum* L.; Heun et al., 1997), garden pea (*Pisum sativum* L.; Palmer et al., 1985), chickpea (*Cicer arietinum* L.; Abbo et al., 2003), and flax (*Linum usitatissimum* L.; Zohary and Hopf, 2000) were all domesticated in this region, as were several animal species (Pedrosa et al., 2005; Driscoll et al., 2007; Naderi et al., 2008). The oilseed crop safflower (*Carthamus tinctorius* L., Asteraceae) is also believed to have been domesticated in the Fertile Crescent, albeit more recently than most other crops from this region (Knowles and Ashri, 1995).

Safflower is a thistle-like, herbaceous crop that thrives in hot, dry climates. Because it is phenotypically differentiated from its progenitor species, grown over a much larger land area, and has a long history of cultivation, safflower has been described as a "strongly domesticated" species (Dempewolf et al., 2008). Currently grown as an oilseed crop (safflower oil is very low in saturated fatty acids; Knowles 1958), safflower has also been proposed as a platform for the production of plant-made pharmaceuticals (Lacey et al., 1998). It was, however, likely first domesticated for its flowers, which have traditionally served as a source of dye (carthamine) for various cultural uses. Floral extracts have also been used as a food additive (in place of saffron) and have been valued for their various medicinal properties (Weiss, 1971). The earliest *Carthamus* seed remains, which are thought to be the remnants of early safflower cultivation, are from northern Syria (dating to ca. 2500 yr BC; van Zeist and Walterbolk-van Rooijen, 1992), making it considerably younger than most other Fertile Crescent domesticates. Additional safflower seeds have been found in the ca. 1325 BC tomb of Tutankhamen in Egypt, and safflower garlands have also been found adorning Egyptian mummies dating to ca. 1600 BC (Zohary and Hopf, 2000 and references therein).

A recent DNA sequence-based analysis of *Carthamus* sect. *Carthamus* has revealed that the progenitor species of safflower is most likely *C. palaestinus* Eig, a self-compatible species native to the region spanning southern Israel to western Iraq (Chapman and Burke, 2007). Moreover, another species that was previously implicated in the origin of safflower, *C. oxyacanthus* Bieb. (Ashri and Knowles, 1960), was found to be more distantly related than originally thought (Chapman and Burke, 2007). These three species have been shown to be genetically distinct from one another based on variation at a series of simple-sequence repeat (SSR) markers (Chapman et al., 2009), but relatively little is known about differentiation within the safflower gene pool.

Early researchers proposed the existence of a number of centers of similarity within the safflower gene pool. Accessions within each of the proposed centers are known to be quite similar to one another with respect to certain attributes such as height, branching, head size, and flower color, whereas differences

[1] Manuscript received 15 May 2009; revision accepted 11 March 2010.

The authors thank the greenhouse staff at UGA for maintenance of the safflower plants and J. Dechaine-Berkas, S. Pearl, and N. Sherman for comments on an earlier version of this manuscript. This work was supported 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).

[2] Author for correspondence (e-mail: jmburke@uga.edu)

doi:10.3732/ajb.0900137

are evident between the centers. Knowles (1969) initially identified seven such centers (Table 1A): (1) the Far East, (2) India–Pakistan, (3) the Middle East, (4) Egypt, (5) Sudan, (6) Ethiopia, and (7) Europe. In contrast, Ashri (1975) used additional morphological descriptors to split the Middle East center into three centers (Iran–Afghanistan, Near East, and Turkey), and he added Kenya as a separate center (Table 1B). The morphological differentiation among the proposed centers of similarity suggests that genetic discontinuities might likewise exist between centers. Unfortunately, the extent to which accessions within centers are genetically more similar to one another than accessions from different centers remains unclear.

To date, analyses of the partitioning of genetic variation across the safflower gene pool have been limited to surveys based on random (and largely dominant) nuclear markers (i.e., RAPDs, ISSRs, and AFLPs). Genetic structuring is evident between accessions (Sehgal and Raina, 2005), though Amini et al. (2008) argued that genetic and morphological similarity are only weakly correlated. Johnson et al. (2007) included accessions from all of Knowles' (1969) seven centers in an analysis of AFLP variation, but they did not investigate the correspondence between these morphologically defined groups and underlying patterns of genetic differentiation. Rather, accessions from Sudan, Egypt, and Kenya were grouped as East Africa, and accessions from Egypt, Syria, Turkey, and other European countries were grouped as Mediterranean. Nonetheless, three of the nine groupings obtained using a Bayesian clustering program contained a predominant population group that corresponded to Knowles' Far East, Middle East, and European centers (Johnson et al., 2007), suggesting that at least some of the proposed centers are biologically real.

Here, we describe a survey of genetic diversity across a broad collection of *Carthamus* accessions, including a diverse, globally representative array of safflower lineages as well as accessions of both *C. palaestinus* and *C. oxyacanthus*. All sampled individuals were genotyped with a suite of codominant, gene-based nuclear markers along with a pair of chloroplast DNA (cpDNA) markers. As such, the resulting data provided us with a detailed understanding of the partitioning of genetic variation across the safflower gene pool with respect to the geographic origin of the cultivars and allowed us to test the hypothesis that the proposed centers of similarity do, in fact, correspond to genetically distinctive subunits. Beyond providing us with novel evolutionary insights, these data also allowed us to identify lineages that are likely to harbor untapped diversity.

TABLE 1. Summaries of the two previous classifications of the safflower gene pool based on morphological variation (A and B) and a comparison with the results of this study based on inferred genetic similarity (C). For ease of cross-referencing elsewhere in this paper, accessions were assigned a two-letter prefix corresponding to their geographic region of origin following Ashri's (1975) hypothesized 10 centers of similarity (B).

A) Knowles (1969)	B) Ashri (1975)	C) This study
	Near East (NE)	Near East
Middle East	Iran/Afghanistan (IA)	Iran & Afghanistan,
	Turkey (Tu)	Turkey
Egypt	Egypt (Eg)	Egypt, Ethiopia, (Sudan)
Ethiopia	Ethiopia (Et)	
Sudan	Sudan (Su)	
Far East	Far East (FE)	Far East, India/Pakistan,
India/Pakistan	India/Pakistan (IP)	(Sudan)
Europe	Europe (Eu)	Europe
	Kenya (Ke)	

MATERIALS AND METHODS

Seed sources and DNA extraction—Seeds from 76 accessions (populations) of *Carthamus tinctorius*, one of *C. palaestinus*, and two of *C. oxyacanthus* were obtained from the USDA Agricultural Research Service Germplasm Resources Information Network (ARS-GRIN; http://www.ars-grin.gov/). Seventy of the safflower accessions represented Ashri's (1975) 10 putative centers of similarity (populations were coded as per Table 1B), with four to eight accessions from each (Appendix 1). These 70 accessions are referred to throughout the manuscript as native safflower accessions because they are not thought to represent secondary introductions. The remaining six, on the other hand, were North American accessions that were not included in the primary analyses because they represent secondary introductions that occurred in the last ca. 100 yr (Knowles, 1958). Because just one *C. palaestinus* accession was available, three additional DNA extractions from *C. palaestinus* herbarium samples (Chapman and Burke, 2007) were also included in the analyses outlined below.

Seeds from all accessions were clipped to break dormancy and germinated on damp filter paper. Seedlings were then transferred to soil and allowed to grow for 2–3 wk prior to DNA extraction. Extractions were performed on young leaf tissue using the standard CTAB procedure (Doyle and Doyle, 1990); however, dichloromethane was used instead of chloroform/isoamyl alcohol. DNA was extracted from three to five individuals of each safflower accession, 12 *C. palaestinus* individuals from the USDA accession, and six individuals from each of the *C. oxyacanthus* accessions.

Nuclear markers—Twenty-four expressed sequence tag (EST)-SSR primer pairs (Appendix 2) developed by Chapman et al. (2009) were used to amplify DNA from all sampled individuals. Forward primers had a universal "tail" appended to their 5′ end such that inclusion of a third (fluorescently labeled) primer in the reaction produced a fluorescently labeled PCR product (Schuelke, 2000). Each PCR contained 10 ng of template DNA, 30 mM tricine (pH 8.4)-KOH, 50 mM KCl, 2 mM $MgCl_2$, 100 μM each dNTP, 0.02 μM forward primer (with a 5′-CACGACGTTGTAAAACGAC-3′ tail), 0.1 μM reverse primer, 0.1 μM fluorescently labeled primer (5′-CACGACGTTGTAAAACGAC-3′ labeled with HEX or FAM), and one unit of *Taq* DNA polymerase. Cycling conditions followed a "touchdown" protocol as follows: initial denaturation at 95°C for 3 min; 10 cycles of 30 s at 94°C, 30 s at 65°C (annealing temperature was reduced by one degree per cycle), and 45 s at 72°C; 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 from three to six primer pairs were combined based on nonoverlapping fragment sizes and/or different fluorescent labels (HEX or FAM) and diluted 1 : 15 in water. Then 3 μL of this mixture was added to 7 μL of formamide containing 0.2 μL the ROX-labeled MapMarker1000 ladder (BioVentures, Murfreesboro, Tennessee, USA) and resolved on an ABI 3730xl DNA sequencer/genotyper (Applied Biosystems, Foster City, California, USA). Alleles were called using GeneMarker ver. 1.51 (SoftGenetics, State College, Pennsylvania, USA), and patterns of population genetic variation were analyzed as follows.

The program GDA (Lewis and Zaykin, 2001) was used to calculate the following statistics: A (number of alleles per locus) and H_e (expected heterozygosity [i.e., gene diversity]). This same program was also used after the Bayesian cluster analysis (see below) to calculate the A and H_e on a per-cluster basis. One thousand randomized between-population distance matrices were then computed in the program MSA (Dieringer and Schlötterer, 2003) using Nei's genetic distance (Nei et al., 1983), and these matrices were used to construct a bootstrapped neighbor-joining (NJ) tree with the program PHYLIP (Felsenstein, 2004; programs CONSENSE and NEIGHBOR).

Principal coordinate analysis (PCO) was carried out using the program GenAlEx (Peakall and Smouse, 2002) to transform the multidimensional genetic distances between populations into a two-dimensional representation that explains as much of the observed variance as possible. Because GenAlEx cannot handle raw genotypic data derived from more than ~125 individuals, the between-population distance matrix was first computed in MSA (above). For this analysis, each of the native safflower accessions was treated as a population, as were the three herbarium samples of *C. palaestinus*. Note that *C. oxyacanthus* was not included in this analysis because the divergent nature of this species significantly reduced the ability to distinguish clusters within safflower.

The Bayesian clustering program STRUCTURE (Pritchard et al., 2000) was then used to identify distinct genetic subgroups within the data set, independent of any prior information concerning the geographic origin of the accessions. The most likely value of *K* (i.e., the estimated number of clusters) was determined following Evanno et al. (2005). After preliminary analyses revealed that the inclusion of *C. oxyacanthus* within the data set always resulted in the most likely number of clusters being two (*C. oxyacanthus* vs. *C. palaestinus* and the native safflower accessions), thereby obscuring any possible substructuring within safflower, this species was removed. Five runs (50 000 replicates after a 10 000 replicate burn-in) for each value of *K* between two and 12 were performed. Once the most likely value of *K* was determined, the proportion of membership in each of the *K* clusters was averaged across runs.

The STRUCTURE analysis was then repeated to investigate the origin of the North American safflower accessions. Given the results of the initial STRUCTURE analysis (see below), this follow up analysis was performed with $K = 5$ clusters. The original data set was culled to include only those native safflower accessions that showed >50% membership in a single cluster, and the six North American accessions were added. Individuals assigned to each of the five clusters were then treated as "learning samples" with the USEPOPINFO feature turned on. Individuals from the North American accessions were treated as unknowns, and STRUCTURE was used to infer their ancestry. This analysis was otherwise performed as described above, and the results were found to be consistent across runs.

cpDNA sequence variation—Ten primer pairs were used to amplify intergenic and intronic regions of the chloroplast genome from eight individuals of diverse origin (one each from Knowles' seven centers; accessions Eu895, Tu624, NE285, Et473, IP889, FE342, Eg613, plus one *C. palaestinus* individual). The *trnC-trnD* region was amplified in three parts using primers *trn*C-*pet*N2R, *pet*N1-*psb*M2R, and *psb*M2-*trn*D (Lee and Wen, 2004), the *atpB-rbcL* spacer was amplified using primers *atp*B-1 and *rbc*L-1 (Chiang et al., 1998), the *trnK-rps16* spacer was amplified using primers *trn*K5′r (Johnson and Soltis, 1995) and *rps*16-4547mod (Kress et al., 2005), the *trnY-rpoB*, *trnL-rpl32*, and *ndhC-trnV* spacers were amplified using primers in (Timme et al., 2007), the *trnL* gene/*trnL-F* spacer was amplified using primers C and F (Taberlet et al., 1991), and the *trnS-psbC* spacer was amplified using primers of Demesure et al. (1995). The last of these regions was previously shown to be polymorphic in safflower by Sehgal et al. (2008). Some of the remainder were selected because of previous reports of high interspecific variation (Kress et al., 2005; Timme et al., 2007).

PCR conditions were the same as for the nuclear SSRs except that the final annealing temperature was increased to 60°C for the *trnL* gene/*trnL-F* spacer and the *trnS-psbC* spacer. In addition, three internal sequencing primers were designed for the *trnS-psbC* spacer (trnS2, 5′-ATTCCTTTTTCAAATCCTGCTG-3′; psbC2, 5′-CTTTTAGGACCTGAGACGCTTG-3′; and trnS3, 5′-TGGGTCCGTAAAACTCACTAGG-3′). PCR products were also obtained for five of the 10 primer pairs using the herbarium DNA (*trn*C-*petN2R*, *petN1-psbM2R*, *trnK-rps16*, *trnL-rpl32*, and *trnL*/*trnL-F*). Finally, given the results of Sehgal et al. (2008; see above), the *trnS-psbC* spacer was also sequenced in six individuals of *C. oxyacanthus*. PCR purification and sequencing followed previously published protocols (Chapman and Burke, 2007), and all sequences were deposited in the NCBI GenBank database (accession numbers GU990407–GU990520).

Sequence alignments were constructed in the program Genedoc (Nicholas and Nicholas Jr, 1997), and polymorphisms were identified by eye. Two of the regions that contained polymorphisms corresponding to restriction site differences (i.e., *petN1-psbM2R* and *trnL-rpl32*) were then used for PCR-RFLP analysis on one individual from each of the 76 safflower accessions, all 12 *C. palaestinus* individuals, and the 12 *C. oxyacanthus* individuals. For these analyses, the PCR primers and conditions were the same as those described above. Five microliters of the resulting PCR was then mixed with 5 μL cocktail containing 2 units of the appropriate restriction enzyme (Table 2), 1 μL enzyme buffer, 0.1 μL bovine serum albumin (if applicable) and ddH_2O. Reactions were incubated at 37°C for 3 h, followed by enzyme inactivation at 80°C for 15 min. RFLPs were then resolved on 1.5% agarose gels stained with ethidium bromide.

RESULTS

Nuclear SSR variation—Across species, the 24 EST-SSR loci amplified between 3 and 15 alleles per locus (mean 7.5), and expected heterozygosity ranged from 0.10 to 0.84 (mean 0.44). Considering just the native safflower accessions, these numbers fell to between 2 and 14 alleles per locus (mean 5.9), and expected heterozygosity ranged from 0.03 to 0.84 (mean 0.39). The NJ analysis (Fig. 1A) revealed that *C. oxyacanthus* is genetically distinct from *C. palaestinus* and the native safflower accessions (100% bootstrap support). The two *C. palaestinus* populations were found at the base of the safflower clade, with one of them falling within a small, unsupported clade comprised of two safflower accessions from Sudan and one from Kenya. The positioning of *C. palaestinus* relative to safflower is thus consistent with a progenitor-derivative species relationship. An anomalous Turkish accession of safflower (Tu273) was found along the branch separating *C. oxyacanthus* from *C. palaestinus*/safflower. Bootstrap support (BS) was low for the majority of branches within safflower; however, a notable exception is a clade with 93% BS comprising all four Ethiopian accessions (Fig. 1A).

Although the PCO plot exhibits overlap between accessions from different geographic regions (Fig. 1B), some geographic structuring is apparent. For the most part, this structuring is similar to that revealed by the NJ analysis. For example, accessions from Iran–Afghanistan and Turkey are closely allied in both the PCO1 vs. PCO2 (Fig. 1B) and PCO1 vs. PCO3 (data not shown) plots, as are Egypt and Ethiopia. The majority of accessions from the Near East, India–Pakistan, and Europe also form fairly tight clusters in the PCO1 vs. PCO2 plot, but with some outlier populations. In contrast, the accessions from Kenya and Sudan are less tightly clustered.

The STRUCTURE analysis revealed that the most likely number of clusters (when *C. oxyacanthus* was excluded; see above) was five. In each of these five clusters, accessions from one or two of the 10 presumptive centers of similarity (Table 1) predominated (Fig. 2). Cluster 1 contained the majority of Near East accessions. The two *C. palaestinus* populations (i.e., the USDA accession as well as the three herbarium specimens, which were grouped together) also had the highest percentage membership in this cluster. Cluster 2 contained six of seven Turkish accessions (the seventh did not cluster with the remainder in the NJ tree or PCO analysis, either) and all accessions from Iran–Afghanistan. Cluster 3 contained seven of the eight Egyptian accessions and all four Ethiopian accessions. Cluster 4 primarily consisted of accessions from Europe, whereas cluster 5 was primarily composed of accessions from India–Pakistan and the Far East. Accessions from Sudan and Kenya were split between clusters 3 and 5. While accessions with mixed ancestry could result either from the occurrence of individuals with mixed ancestry or from the averaging together of "pure" individuals that show an affinity for different clusters, we found that the former scenario predominated, In other words, individuals within accessions were genetically quite similar to one another.

TABLE 2. Polymorphisms and restriction enzymes used to distinguish the three cpDNA haplotypes.

Region	Haplotype A	Haplotype B	Haplotype C	Restriction enzyme
petN1-psbM2R	aatAaattgatt	aatCaattgatt	aatAaattgatt	*Mfe*I (CAATTG)
trnL-rpl32	catCttagaag	catTttagaag	catTttagaag	*Dde*I (CTNAG)

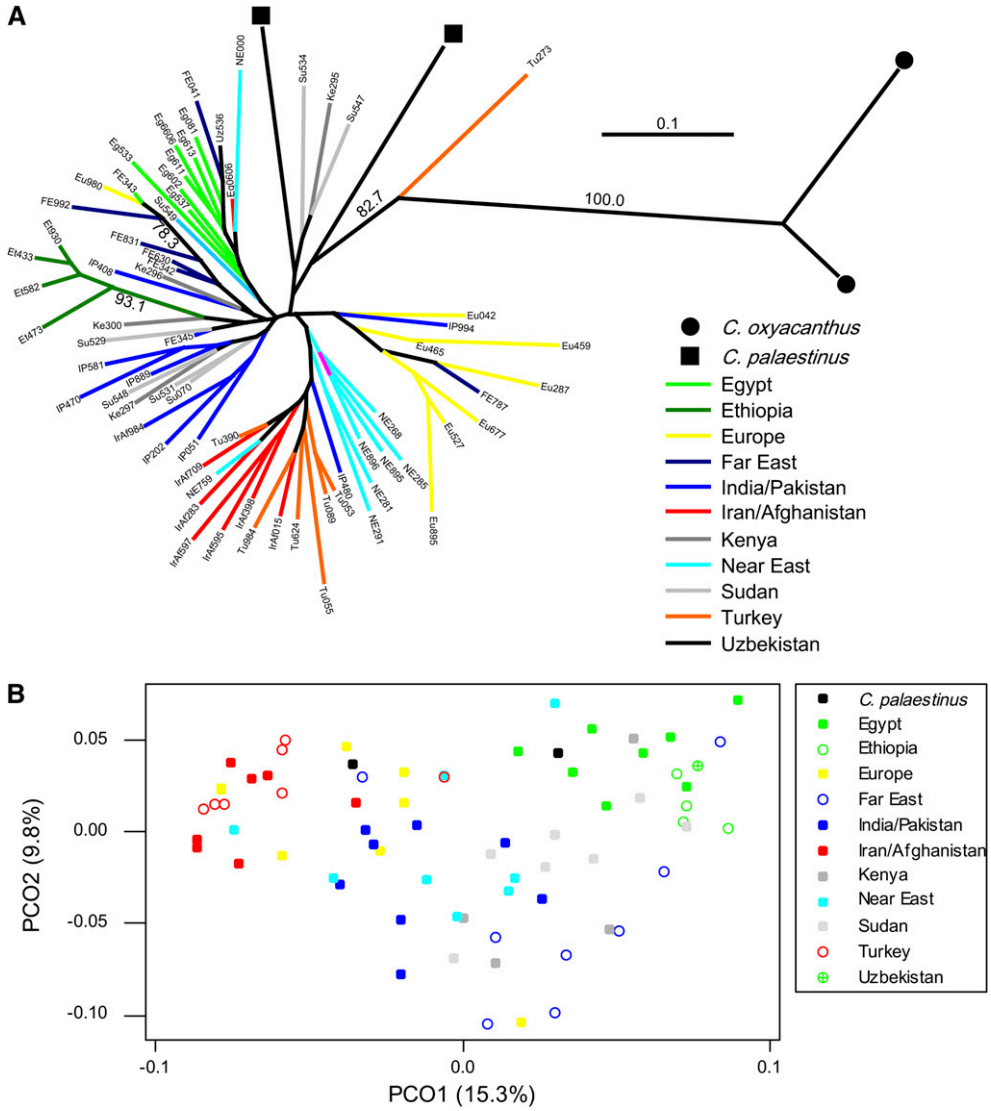

Fig. 1. Relationships between *Carthamus* accessions. (A) Neighbor-joining tree of the 70 native safflower accessions, two *C. palaestinus* accessions, and two *C. oxyacanthus* accessions. Bootstrap percentages greater than 75% are indicated. (B) Principal coordinate analysis (PCO) of the 70 native safflower accessions plus two accessions of *C. palaestinus* based on genotypic information from 24 nuclear SSR loci. The geographic origin of each species is indicated by symbols.

An analysis of the native safflower accessions on a per-cluster basis (and excluding those accessions without >50% membership in any one cluster) revealed that expected heterozygosity was highest in the Near East cluster (cluster 1; 0.348) and lowest in Europe (cluster 4; 0.265) and India–Pakistan (cluster 5; 0.287). The reanalysis of these data using STRUCTURE (and including the six North American safflower accessions) revealed that the North American accessions belong to clusters 2, 3, or 4 (one accession from the USA belongs to cluster 2, one accession from Mexico belongs to cluster 3, two accessions from the USA and one from Canada belong to cluster 4, and the final Canadian accession is split between clusters 3 and 4; Appendix 1).

Chloroplast DNA variation—Ten polymorphic cpDNA sites (nine single nucleotide polymorphisms [SNPs] and one length variation in the poly A region; Table 3) were identified from a total of ~9.1 kb of sequence collected from eight individuals. Five of these regions (totaling 3.8 kb) were also sequenced from the three herbarium accessions of *C. palaestinus*. These variants resolved into three haplotypes within the sequenced individuals, with the Tu624 and IP889 individuals having haplotype A, one *C. palaestinus* individual (herbarium extraction pal98) having haplotype C and the remaining eight individuals (five safflower plus three *C. palaestinus*) having haplotype B. Despite a report of a ca. 60-bp indel in the *trnS-psbC* spacer of one *C. tinctorius* and one *C. oxyacanthus* individual (Sehgal et al., 2008), only one single nucleotide polymorphism (SNP) was resolved in this region even after an additional six *C. oxyacanthus* individuals were sequenced.

Two of the polymorphisms that could distinguish the three haplotypes (Table 2) were selected for RFLP investigation on a single individual from each of the sampled native safflower accessions, plus all *C. palaestinus* and *C. oxyacanthus* individuals.

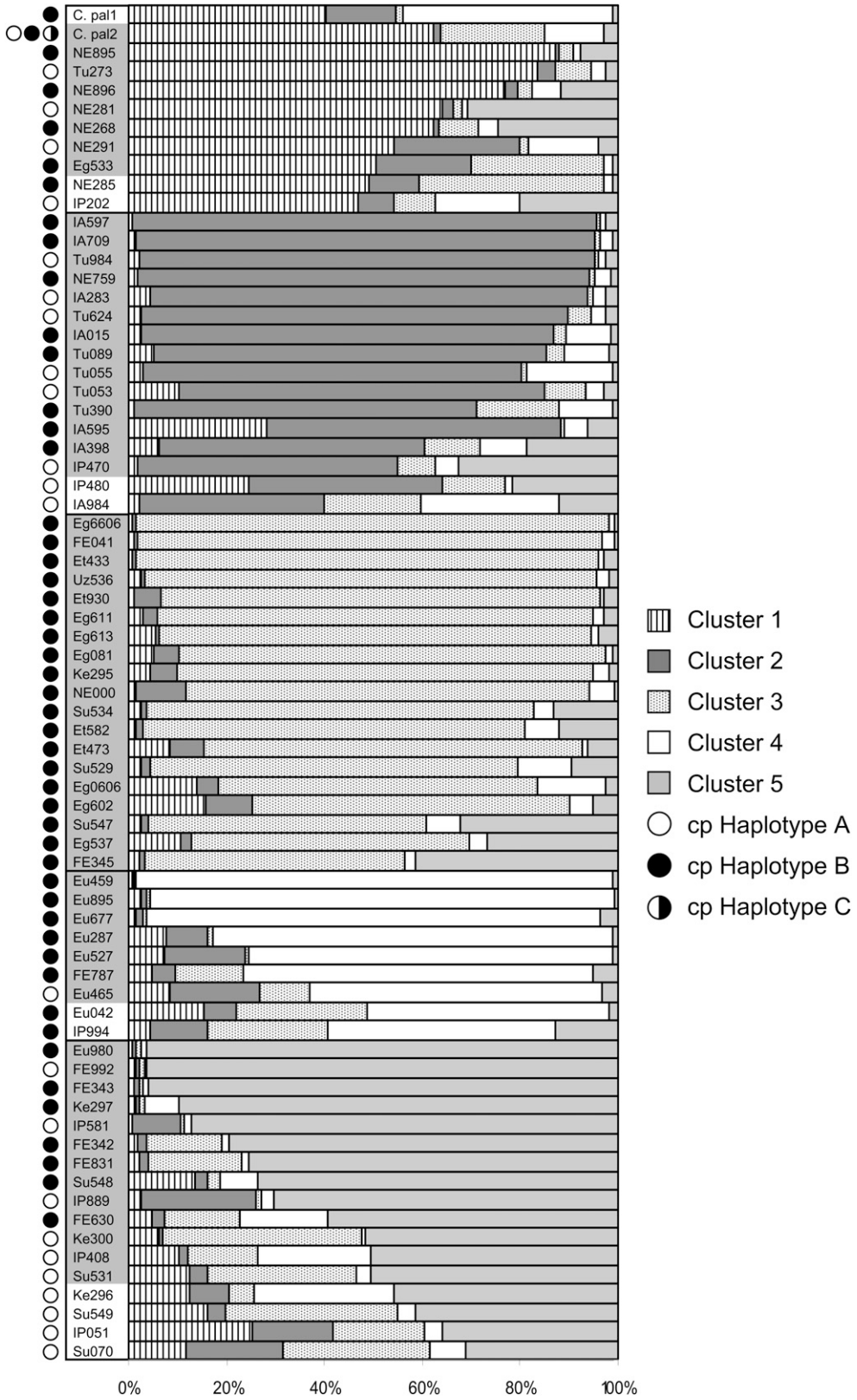

Fig. 2. Results of STRUCTURE analysis. Populations are ordered (top to bottom) according to (1) cluster with greatest proportion of membership and (2) percentage of membership to that cluster. The names of populations with >50% membership to the same cluster are shaded. Population names are prefixed to identify their place of origin (see Table 1). Chloroplast haplotypes are depicted next to the population name.

TABLE 3. Summary of aligned sequence length, number of polymorphisms and sequences of the three haplotypes for 10 cpDNA regions sequenced in a panel of seven safflower and one to four *C. palaestinus* individuals.

Region	Sequence length	No. of polymorphisms	Haplotype A	Haplotype B	Haplotype C
Monomorphic loci					
atpB-rbcL	831	0			
trnK-rps16 [a]	811	0			
trnY-rpoB	1049	0			
Polymorphic loci					
trnC-petN2R [a]	584	1	A_9	A_{10}	A_{10}
petN1-psbM2R [a]	691	1	A	C	A
psbM2-trnD	736	2	T, G	A, T	? [b]
trnL-rpl32 [a]	883	3	G, G, C	A, A, T	A, A, T
ndhC-trnV	1202	1	C	A	? [b]
trnL/trnL-F spacer [a]	826	1	T	C	T
trnS-psbC	1513	1	G	A	? [b]
Total	9126	10			

[a] This cpDNA region was sequenced from the *C. palaestinus* herbarium DNA extractions.

[b] Haplotype C was found in a single herbarium individual of *C. palaestinus* from which PCR products could not be obtained for all cpDNA regions.

Haplotypes are mapped onto the STRUCTURE results (Fig. 2). The three herbarium DNAs each had a different cpDNA haplotype. Haplotypes A and C were also found in *C. oxyacanthus* (nine and three individuals, respectively). Haplotype B was present in all individuals of the USDA accession of *C. palaestinus*. Twenty-three native safflower accessions contained haplotype A, and 47 contained haplotype B (Fig. 2).

Comparing the cpDNA haplotypes with the STRUCTURE results reveals that clusters 1, 2, and 5 contain accessions with both haplotypes. Cluster 3 contains only accessions with haplotype B, and all but one accession from cluster 4 exhibit haplotype B. The results for the North American accessions were all consistent with these findings.

DISCUSSION

An understanding of the partitioning of genetic variation within crop gene pools can provide insight into the evolution of crop lineages and can also reveal untapped sources of genetic variation that could fuel future improvement efforts (Tanksley and McCouch, 1997; Yamasaki et al., 2005). Historically, researchers have relied on patterns of morphological variation to make inferences about crop origins as well as relationships among lineages within a given crop. More recently, these issues have been investigated at the genetic level. In some cases, the morphological and genetic data are in agreement, but in other cases, they are not. For example, the early view (based on morphology) was that maize had been domesticated on more than one occasion (reviewed in Matsuoka, 2005). Genetic analyses have, however, subsequently revealed that maize is the product of a single domestication (Matsuoka et al., 2002). Further, in common bean, which has long been known to have been domesticated twice (Becerra Velasquez and Gepts, 1994), morphological analyses suggested the presence of seven distinct races. In contrast, genetic analyses have revealed the presence of just four (Diaz and Blair, 2006; Blair et al., 2007).

In safflower, DNA sequence variation has been used to identify the progenitor species (*C. palaestinus*; Chapman and Burke, 2007), and random nuclear markers have been used to distinguish between cultivars (e.g., Sehgal and Raina, 2005; Johnson et al., 2007; Amini et al., 2008). Until now, however, there has been no detailed population genetic analysis of the safflower gene pool, and the centers of similarity hypothesis of Knowles (1969) and Ashri (1975) has remained untested at the molecular level. Based on the data presented herein, the safflower gene pool appears to be composed of five somewhat distinct centers of genetic similarity. Table 1 highlights the similarities and differences between the Knowles (1969) and Ashri (1975) classifications (see the introduction for details) and the results of this study (Figs. 1, 2).

Ashri (1975) proposed that Knowles' (1969) grouping of all Middle Eastern accessions of safflower was incorrect due to morphological differences between the Near East, Iran–Afghanistan, and Turkish accessions. Based on our data, however, the Iran/Afghanistan and Turkish accessions indeed appear to be genetically quite similar (Fig. 2; cluster 2), with the Near East accessions being distinct (cluster 1). Accession NE759 is the only Iraqi accession, and the fact that it groups with Iran/Afghanistan (Fig. 2) suggests that this group might extend farther west than previously believed, with the Near East cluster occurring west of Iraq. In addition, both Knowles (1969) and Ashri (1975) separated the Far East accessions from the India and Pakistan accessions, though they are similar enough to be grouped together based on our genetic data (cluster 5). The two India/Pakistan accessions that do not cluster with the balance of lines from that region are both from Bangladesh and thus might represent a secondary introduction, perhaps via trade, as opposed to lines native to Bangladesh.

Our STRUCTURE results placed the Egyptian and Ethiopian accessions (along with some from Sudan) into one group (cluster 3; Fig. 2), though the NJ tree separates these groups of accessions (Fig. 1A). Knowles (1969) noted that traditional Sudanese types were found in northern Sudan and southern Egypt, an area that was flooded by the construction of the Aswan High Dam in the 1960s. People from that region were displaced and are thought to have taken seeds of their crops north with them (Knowles, 1969). It is thus possible that a merging of the Sudanese and Egyptian types has occurred. In addition, American varieties were being field-tested in Egypt in the 1960s (Knowles, 1969) and were likely widely distributed during subsequent years, perhaps diluting the native Egyptian and former-Sudanese accessions. Regardless of the cause, our STRUCTURE results suggest that the safflower lines from Northeast Africa comprise a single, more-or-less distinct gene pool.

It is worth noting here that similar patterns have been seen in other crop plants. For example, an analysis of genetic structure of soybean in its native range (Li et al., 2008) revealed qualitatively similar results to those found in the present investigation. More specifically, Bayesian clustering revealed that (1) genetic clusters of accessions did not always correspond well to morphological groups of accessions and that (2) only a subset of these genetic clusters corresponded well with geographic origin. Both of these discordances are, to a greater or less extent, apparent within the safflower data set. Similarly, as noted above, morphological differentiation is not a good predictor of genetic differentiation in common bean (Diaz and Blair, 2006; Blair et al., 2007) and in flax, another oilseed crop that was domesticated in the Fertile Crescent (Allaby et al., 2005), the

geographic origin of accessions does not correspond strongly with genetic subdivisions (Fu, 2005).

Carthamus palaestinus, the progenitor species of safflower, is restricted to a small region of the Near East (southern Israel through western Iraq). It therefore seems likely that safflower was domesticated somewhere in this general vicinity. Consistent with this hypothesis, the STRUCTURE analysis grouped the Near East safflower accessions with *C. palaestinus* (Fig. 2). This finding also agrees with the archaeological data, which places the earliest safflower remains in Syria approximately 4500 yr BP (van Zeist and Walterbolk-van Rooijen, 1992). Beyond this, our estimates of genetic diversity were highest in the Near East group (as defined based on the STRUCTURE output), as might be expected for the ancestral domesticated population from which the other centers of similarity were subsequently derived. We cannot, however, discount the possibility that localized introgression from wild relatives produced this pattern (e.g., Ellstrand et al., 1999; Weissmann et al., 2005), nor can we rule out the possibility of a more complex origin of safflower (see below).

Unfortunately, we were unable to rigorously test the hypothesis of single vs. multiple origins due to limited availability of *C. palaestinus* (this study used one USDA accession as well as three herbarium samples that were collectively treated as a second population). The PCO and NJ results (Fig. 1) demonstrate that (1) there is some genetic differentiation between the USDA accession and the herbarium individuals and that (2) these two populations are not monophyletic in the NJ tree (Fig. 1A). The NJ tree, however, suffers from very low bootstrap support values; hence, the nonmonophyly of *C. palaestinus* is not well supported. In terms of cpDNA variation, two haplotypes were present in cultivated safflower, which are a subset of those found in *C. palaestinus*. Sehgal et al. (2008) likewise reported the presence of two distinct cpDNA haplotypes in safflower and further concluded that this was evidence that both *C. palaestinus* and *C. oxyacanthus* played a role in the origin of safflower. However, their work was based on very limited sampling—just two safflower individuals along with one individual from each of the wild species. A more reasonable interpretation is that there is extensive haplotype sharing among *Carthamus* species. Regardless, the finding of multiple cpDNA haplotypes is not necessarily indicative of multiple origins (or multiple progenitor species) because a number of crops that are known to have arisen just once have been found to harbor multiple cpDNA haplotypes (e.g., Wang et al., 1997; Ishii et al., 2001). Conversely, crops with multiple origins sometimes exhibit just one cpDNA haplotype (e.g., Clegg et al., 1984).

When combined with findings from the archaeological record as well as the geographic distribution of the progenitor species, our data are fully consistent with a single origin of safflower somewhere west of the Fertile Crescent followed by subsequent expansion into Africa, Europe, and the balance of Asia. It has, however, been argued that the domestication of crop plants in the Fertile Crescent region was a complex, protracted, and geographically diffuse process (Brown et al., 2009). Thus, while safflower is believed to have been domesticated much later than most other Fertile Crescent species, it could be that the observed differentiation within the contemporary gene pool is partly due to early safflower evolution having occurred in a similarly diffuse fashion. Despite the lingering uncertainty surrounding the early evolution of safflower, our results indicate that at least a subset of the native safflower accessions are likely to harbor unique genetic diversity that can be exploited for the continued improvement of safflower in North America. Most notably, the Near East accessions, which comprise the most genetically diverse subgroup that we identified, appear to be relatively distinct from the North American accessions included in our survey.

LITERATURE CITED

ABBO, S., D. SHTIENBERG, J. LICHTENZVEIG, S. LEV-YADUN, AND A. GOPHER. 2003. The chickpea, summer cropping, and a new model for pulse domestication in the ancient Near East. *Quarterly Review of Biology* 78: 435–448.

ALLABY, R. G., G. W. PETERSON, D. A. MERRIWETHER, AND Y. B. FU. 2005. Evidence of the domestication history of flax (*Linum usitatissimum* L.) from genetic diversity of the *sad2* locus. *Theoretical and Applied Genetics* 112: 58–65.

AMINI, F., G. SAEIDI, AND A. ARZANI. 2008. Study of genetic diversity in safflower genotypes using agro-morphological traits and RAPD markers. *Euphytica* 163: 21–30.

ASHRI, A. 1975. Evaluation of germplasm collection of safflower, *Carthamus tinctorius* L. V. Distribution and regional divergence for morphological characters. *Euphytica* 24: 651–659.

ASHRI, A., AND P. F. KNOWLES. 1960. Cytogenetics of safflower (*Carthamus* L.) species and their hybrids. *Agronomy Journal* 52: 11–17.

BADR, A., K. MULLER, R. SCHAFER-PREGL, H. EL RABEY, S. EFFGEN, H. H. IBRAHIM, C. POZZI, W. ROHDE, AND F. SALAMINI. 2000. On the origin and domestication history of barley (*Hordeum vulgare*). *Molecular Biology and Evolution* 17: 499–510.

BECERRA VELASQUEZ, V. L., AND P. GEPTS. 1994. RFLP diversity of common bean (*Phaseolus vulgaris*) in its centers of origin. *Genome* 37: 256–263.

BLAIR, M. W., J. M. DÍAZ, R. HIDALGO, L. M. DÍAZ, AND M. C. DUQUE. 2007. Microsatellite characterization of Andean races of common bean (*Phaseolus vulgaris* L.). *Theoretical and Applied Genetics* 116: 29–43.

BROWN, T. A., M. K. JONES, W. POWELL, AND R. G. ALLABY. 2009. The complex origins of domesticated crops in the Fertile Crescent. *Trends in Ecology & Evolution* 24: 103–109.

CHAPMAN, M. A., AND J. M. BURKE. 2007. DNA sequence diversity and the origin of cultivated safflower (*Carthamus tinctorius* L.; Asteraceae). *BMC Plant Biology* 7: 60.

CHAPMAN, M. A., J. HVALA, J. STREVER, M. MATVIENKO, A. KOZIK, R. W. MICHELMORE, S. TANG, ET AL. 2009. Development, polymorphism, and cross-taxon utility of EST-SSR markers from safflower (*Carthamus tinctorius* L.). *Theoretical and Applied Genetics* 120: 85–91.

CHIANG, T. Y., B. A. SCHAAL, AND C.-I. PENG. 1998. Universal primers for amplification and sequencing a non-coding spacer between the *atpB* and *rbcL* genes of chloroplast DNA. *Botanical Bulletin of Academia Sinica* 39: 245–250.

CLEGG, M. T., J. R. Y. RAWSON, AND K. THOMAS. 1984. Chloroplast DNA variation in pearl millet and related species. *Genetics* 106: 449–461.

DEMESURE, B., N. SODZI, AND R. J. PETIT. 1995. A set of universal primers for amplification of polymorphic noncoding regions of mitochondrial and chloroplast DNA in plants. *Molecular Ecology* 4: 129–131.

DEMPEWOLF, H., L. H. RIESEBERG, AND Q. C. CRONK. 2008. Crop domestication in the Compositae: a family-wide trait assessment. *Genetic Resources and Crop Evolution* 55: 1141–1157.

DIAZ, L. M., AND M. W. BLAIR. 2006. Race structure within the Mesoamerican gene pool of common bean (*Phaseolus vulgaris* L.) as determined by microsatellite markers. *Theoretical and Applied Genetics* 114: 143–154.

DIERINGER, D., AND C. SCHLÖTTERER. 2003. Microsatellite analyser (MSA): A platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* 3: 167–169.

DOYLE, J. J., AND J. L. DOYLE. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12: 13–15.

DRISCOLL, C. A., M. MENOTTI-RAYMOND, A. L. ROCA, K. HUPE, W. E. JOHNSON, E. GEFFEN, E. H. HARLEY, ET AL. 2007. The Near Eastern origin of cat domestication. *Science* 317: 519–523.

ELLSTRAND, N. C., H. C. PRENTICE, AND J. F. HANCOCK. 1999. Gene flow and introgression from domesticated plants into their wild relatives. *Annual Review of Ecology and Systematics* 30: 539–563.

EVANNO, G., S. REGNAUT, AND J. GOUDET. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology* 14: 2611–2620.

FELSENSTEIN, J. 2004. PHYLIP (Phylogeny Inference Package), version 3.72. Distributed by the author, Department of Genome Sciences, University of Washington, Seattle. Website http://evolution.gs.washington.edu/phylip.html.

FU, Y. B. 2005. Geographic patterns of RAPD variation in cultivated flax. *Crop Science* 45: 1084–1091.

HEUN, M., R. SCHÄFER-PREGL, D. KLAWAN, R. CASTAGNA, M. ACCERBI, B. BORGHI, AND F. SALAMINI. 1997. Site of einkorn wheat domestication identified by DNA fingerprinting. *Science* 278: 1312–1314.

ISHII, T., N. MORI, AND Y. OGIHARA. 2001. Evaluation of allelic diversity at chloroplast microsatellite loci among common wheat and its ancestral species. *Theoretical and Applied Genetics* 103: 896–904.

JOHNSON, L. A., AND D. E. SOLTIS. 1995. Phylogenetic inference in Saxifragaceae sensu stricto and *Gilia* (Polemoniaceae) using *matK* sequences. *Annals of the Missouri Botanical Garden* 82: 149–175.

JOHNSON, R. C., T. J. KISHA, AND M. A. EVANS. 2007. Characterizing safflower germplasm with AFLP molecular markers. *Crop Science* 47: 1728–1736.

KNOWLES, P. F. 1958. Safflower. *Advances in Agronomy* 10: 289–323.

KNOWLES, P. F. 1969. Centers of plant diversity and conservation of crop germplasm—Safflower. *Economic Botany* 23: 324–329.

KNOWLES, P. F., AND A. ASHRI. 1995. Safflower: *Carthamus tinctorius* (Compositae). *In* J. Smartt and N. W. Simmonds [eds.], Evolution of crop plants, 47–50. Longman, Harlow, UK.

KRESS, W. J., K. J. WURDACK, E. A. ZIMMER, L. A. WEIGT, AND D. H. JANZEN. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences, USA* 102: 8369–8374.

LACEY, D. J., N. WELLNER, F. BEAUDOIN, J. A. NAPIER, AND P. R. SHEWRY. 1998. Secondary structure of oleosins in oil bodies isolated from seeds of safflower (*Carthamus tinctorius* L.) and sunflower (*Helianthus annuus* L.). *Biochemical Journal* 334: 469–477.

LEE, C., AND J. WEN. 2004. Phylogeny of *Panax* using chloroplast *trn*C–*trn*D intergenic region and the utility of *trn*C–*trn*D in interspecific studies of plants. *Molecular Phylogenetics and Evolution* 31: 894–903.

LEWIS, P. O., AND D. ZAYKIN. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. Website http://www.eeb.uconn.edu/people/plewis/software.php.

LI, Y. H., R. X. GUAN, Z. X. LIU, Y. S. MA, L. X. WANG, L. H. LI, F. Y. LIN, ET AL. 2008. Genetic structure and diversity of cultivated soybean (*Glycine max* (L.) Merr.) landraces in China. *Theoretical and Applied Genetics* 117: 857–871.

MATSUOKA, Y. 2005. Origin matters: Lessons from the search for the wild ancestor of maize. *Breeding Science* 55: 383–390.

MATSUOKA, Y., Y. VIGOUROUX, M. M. GOODMAN, J. SANCHEZ G., E. BUCKLER, AND J. DOEBLEY. 2002. A single domestication for maize is shown by multilocus microsatellite genotyping. *Proceedings of the National Academy of Sciences, USA* 99: 6080–6084.

NADERI, S., H. R. REZAEI, F. POMPANON, M. G. B. BLUM, R. NEGRINI, H. R. NAGHASH, O. BALKIZ, ET AL. 2008. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences, USA* 105: 17659–17664.

NEI, M., F. TAJIMA, AND Y. TATENO. 1983. Accuracy of estimated phylogenetic trees from molecular data. 2. Gene-frequency data. *Journal of Molecular Evolution* 19: 153–170.

NICHOLAS, K. B., AND H. B. NICHOLAS JR. 1997. Genedoc: A tool for editing and annotating multiple sequence alignments. Computer program distributed by the author, website http://www.nrbsc.org/gfx/genedoc/.

OZKAN, H., A. BRANDOLINI, R. SCHAFER-PREGL, AND F. SALAMINI. 2002. AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. *Molecular Biology and Evolution* 19: 1797–1801.

PALMER, J. D., R. A. JORGENSEN, AND W. F. THOMPSON. 1985. Chloroplast DNA variation and evolution in *Pisum*—Patterns of change and phylogenetic analysis. *Genetics* 109: 195–213.

PEAKALL, R., AND P. E. SMOUSE. 2002. GenAlEx v5.04: Genetic analysis in Excel. Population genetic software for teaching and research. Australian National University, Canberra, Australia. Website http://www.anu.edu.au/BoZo/GenAlEx/.

PEDROSA, S., M. UZUN, J. J. ARRANZ, B. GUTIERREZ-GILL, F. S. PRIMITIVO, AND Y. BAYON. 2005. Evidence of three maternal lineages in near eastern sheep supporting multiple domestication events. *Proceedings of the Royal Society of Longdon, B, Biological Sciences* 272: 2211–2217.

PRITCHARD, J. K., M. STEPHENS, AND P. DONNELLY. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.

SCHUELKE, M. 2000. An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* 18: 233–234.

SEHGAL, D., AND S. N. RAINA. 2005. Genotyping safflower (*Carthamus tinctorius*) cultivars by DNA fingerprints. *Euphytica* 146: 67–76.

SEHGAL, D., V. R. RAJPAL, AND S. N. RAINA. 2008. Chloroplast DNA diversity reveals the contribution of two wild species to the origin and evolution of diploid safflower (*Carthamus tinctorius* L.). *Genome* 51: 638–643.

TABERLET, P., L. GIELLY, G. PAUTOU, AND J. BOUVET. 1991. Universal primers for amplification of 3 noncoding regions of chloroplast DNA. *Plant Molecular Biology* 17: 1105–1109.

TANKSLEY, S. D., AND S. R. MCCOUCH. 1997. Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* 277: 1063–1066.

TIMME, R. E., J. V. KUEHL, J. L. BOORE, AND R. K. JANSEN. 2007. A comparative analysis of the *Lactuca* and *Helianthus* (Asteraceae) plastid genomes: Identification of divergent regions and categorization of shared repeats. *American Journal of Botany* 94: 302–312.

VAN ZEIST, W., AND W. WALTERBOLK-VAN ROOIJEN. 1992. Two interesting floral finds from third millennium B.C. Tell Hamman et-Turkman, northern Syria. *Vegetation History and Archaeobotany* 1: 157–161.

WANG, G. Z., N. T. MIYASHITA, AND K. TSUNEWAKI. 1997. Plasmon analyses of *Triticum* (wheat) and *Aegilops*: PCR single-strand conformational polymorphism (PCR-SSCP) analyses of organellar DNAs. *Proceedings of the National Academy of Sciences, USA* 94: 14570–14577.

WEISS, E. A. 1971. Castor, sesame and safflower. Barnes and Noble, New York, New York, USA.

WEISSMANN, S., M. FELDMAN, AND J. GRESSEL. 2005. Sequence evidence for sporadic intergeneric DNA introgression from wheat into a wild *Aegilops* species. *Molecular Biology and Evolution* 22: 2055–2062.

YAMASAKI, M., M. I. TENAILLON, I. V. BI, S. G. SCHROEDER, H. SANCHEZ-VILLEDA, J. F. DOEBLEY, B. S. GAUT, AND M. D. MCMULLEN. 2005. A large-scale screen for artificial selection in maize identifies candidate agronomic loci for domestication and crop improvement. *Plant Cell* 17: 2859–2872.

ZOHARY, D., AND M. HOPF. 2000. Domestication of plants in the old world: The origin and spread of cultivated plants in West Asia, Europe and the Nile Valley. Oxford University Press, Oxford, UK.

APPENDIX 1. List of *Carthamus* accessions from the USDA Agricultural Research Service Germplasm Resources Information Network used in this investigation.

Accession	Species	Country of origin	Center of origin	PI number
Eg0606	*C. tinctorius*	Egypt	Egypt	250606
Eg081	*C. tinctorius*	Egypt	Egypt	250081
Eg533	*C. tinctorius*	Egypt	Egypt	250533
Eg537	*C. tinctorius*	Egypt	Egypt	250537
Eg602	*C. tinctorius*	Egypt	Egypt	306602
Eg611	*C. tinctorius*	Egypt	Egypt	250611
Eg613	*C. tinctorius*	Egypt	Egypt	306613
Eg6606	*C. tinctorius*	Egypt	Egypt	306606
Et433	*C. tinctorius*	Ethiopia	Ethiopia	262433
Et473	*C. tinctorius*	Ethiopia	Ethiopia	193473
Et582	*C. tinctorius*	Ethiopia	Ethiopia	257582
Et930	*C. tinctorius*	Ethiopia	Ethiopia	343930
Eu042	*C. tinctorius*	Morocco	Europe	239042
Eu287	*C. tinctorius*	Romania	Europe	209287
Eu459	*C. tinctorius*	Portugal	Europe	613459
Eu465	*C. tinctorius*	Spain	Europe	613465
Eu527	*C. tinctorius*	France	Europe	253527
Eu677	*C. tinctorius*	Algeria	Europe	208677
Eu895	*C. tinctorius*	Morocco	Europe	195895
Eu980	*C. tinctorius*	Italy	Europe	576980
FE041	*C. tinctorius*	China	Far East	544041
FE342	*C. tinctorius*	Japan	Far East	279342
FE343	*C. tinctorius*	Japan	Far East	279343
FE345	*C. tinctorius*	Japan	Far East	279345
FE630	*C. tinctorius*	China	Far East	514630
FE787	*C. tinctorius*	China	Far East	568787
FE831	*C. tinctorius*	China	Far East	568831
FE992	*C. tinctorius*	N. Korea	Far East	576992
IP051	*C. tinctorius*	India	India/Pakist.	279051
IP202	*C. tinctorius*	Pakistan	India/Pakist.	250202
IP408	*C. tinctorius*	Pakistan	India/Pakist.	304408
IP470	*C. tinctorius*	Bangla.	India/Pakist.	401470
IP480	*C. tinctorius*	Bangla.	India/Pakist.	401480
IP581	*C. tinctorius*	India	India/Pakist.	401581
IP889	*C. tinctorius*	India	India/Pakist.	199889
IP994	*C. tinctorius*	Pakistan	India/Pakist.	259994
IA015	*C. tinctorius*	Iran	Iran/Afghan.	406015
IA283	*C. tinctorius*	Afghan.	Iran/Afghan.	220283
IA398	*C. tinctorius*	Iran	Iran/Afghan.	251398
IA595	*C. tinctorius*	Afghan.	Iran/Afghan.	304595
IA597	*C. tinctorius*	Afghan.	Iran/Afghan.	304597
IA709	*C. tinctorius*	Iran	Iran/Afghan.	250709
IA984	*C. tinctorius*	Iran	Iran/Afghan.	405984
Ke295	*C. tinctorius*	Kenya	Kenya	209295
Ke296	*C. tinctorius*	Kenya	Kenya	209296
Ke297	*C. tinctorius*	Kenya	Kenya	209297
Ke300	*C. tinctorius*	Kenya	Kenya	209300
NE000	*C. tinctorius*	Israel	Near East	292000
NE268	*C. tinctorius*	Jordan	Near East	251268
NE281	*C. tinctorius*	Israel	Near East	209281
NE285	*C. tinctorius*	Jordan	Near East	251285
NE291	*C. tinctorius*	Jordan	Near East	251291
NE759	*C. tinctorius*	Iraq	Near East	253759
NE895	*C. tinctorius*	Syria	Near East	253895
NE896	*C. tinctorius*	Syria	Near East	253896
Su070	*C. tinctorius*	Sudan	Sudan	271070
Su529	*C. tinctorius*	Sudan	Sudan	305529
Su531	*C. tinctorius*	Sudan	Sudan	305531
Su534	*C. tinctorius*	Sudan	Sudan	305534
Su547	*C. tinctorius*	Sudan	Sudan	237547
Su548	*C. tinctorius*	Sudan	Sudan	237548
Su549	*C. tinctorius*	Sudan	Sudan	237549
Tu053	*C. tinctorius*	Turkey	Turkey	301053
Tu055	*C. tinctorius*	Turkey	Turkey	301055
Tu089	*C. tinctorius*	Turkey	Turkey	340089
Tu273	*C. tinctorius*	Turkey	Turkey	170273
Tu390	*C. tinctorius*	Turkey	Turkey	167390
Tu624	*C. tinctorius*	Turkey	Turkey	407624
Tu984	*C. tinctorius*	Turkey	Turkey	251984
Uz536	*C. tinctorius*	Uzbekist.		305536
MEX111	*C. tinctorius*	Mexico	NA	537111
USA163	*C. tinctorius*	USA	NA	560163
USA418	*C. tinctorius*	USA	NA	572418
USA435	*C. tinctorius*	USA	NA	572435
CAN391	*C. tinctorius*	Canada	NA	592391
CAN207	*C. tinctorius*	Canada	NA	603207
pal1	*C. palaestinus*	Israel	NA	235663
ox2	*C. oxyacanthus*	Pakistan	NA	426428
ox753	*C. oxyacanthus*	Pakistan	NA	426488

Notes: Afganist. = Afganistan, Bangla. = Bangladesh, Pakist. = Pakistan, Uzbekist. = Uzbekistan

APPENDIX 2. Microsatellite loci, primer sequences, and levels of polymorphism analyzed across 70 native safflower accessions.

Locus	FWD primer	REV primer	No. of alleles	H_e
CT1.238	TCCACCTTCGGTACACCTTC	CGAGCTCTGTTGCTGAAATG	2	0.460712
CT1.406	TTCCTTCCATGGCTACAACAC	AAGAAGCGGCGGATAATAGG	3	0.160125
CT1.2042	TCTTCAACCACCACCAATTC	AAACCACAGCGAAGATCACC	5	0.299302
CT1.2160	AACCACCGATTCATCGTCTC	AAACACCACAGCTTCCGTTC	4	0.316798
CT1.2276	GTGGCTGGTTTGATTGATTG	TGCAGTTGTTTGGATTCACAG	5	0.357291
CT1.3012	CCCCTTCCTCTTCTTCAACC	GCGTAGCGACTGCCTTAATC	7	0.475829
CT1.3321	AATTCCATCACGGTTTCTGC	CCCATCCCTCTTCCTCTCTG	5	0.598188
CT1.3436	GGCTCAACTCGACTCATCATC	TTTCCTTCCTGCAACCTCTG	14	0.572186
CT1.3659	GCTGCGTTACCTCACTCCTC	ATTGAGCAGCAATCCTCTGC	7	0.743255
CT1.4463	CTCGATGCCATTTCTAACCAC	TCCTCCTGATCTCCACCATC	7	0.641982
CT1.5358	TGGCAGATTTGAAGGGAAAG	CACTGGAGGAGAAGGAGCAC	2	0.10853
CT1.7092	CAGAGGCTTCCATCGATCAG	GATCGTCCCAAATCATGACC	6	0.25856
EL384102	ACCACCGATGACTCACCTTC	GCTCATCTTCTTCGGACAGC	7	0.259389
EL390720	CCATCATCTCCTCCTCCATC	CCCACTTCAAACTTTTAGCACA	7	0.768849
EL393877	AACCCTCTCTCTACTCCCTAACG	TGTGAACCCTCATCCATCTG	5	0.113033
EL394865	AACAATGGAGGAGGATGACG	TTGTGAATTGCCAGACCTTG	2	0.029901
EL398063	ACGGAGTTCACAGGTGGTG	TGTCGGCAGGTACGAGTATG	5	0.317728
EL399497	CCGATTCGAGATCCTGTTTC	CCATTACCGATCGTTGTTTC	14	0.838093
EL401029	TCACCTTCTTCCTTACTTCTTTCC	TCGACACGCAAACAAATCTC	5	0.125429
EL404104	CAATTTCTCACACACTTTCAACC	ATGATTTCCTTTGGCAGCAG	6	0.696529
EL404597	TTCGAGTTGTGCCCTAAACC	GTTGCTCCTCCTTTGATTGC	6	0.672239
EL407741	TTCAGCCCTCTTCCAATCAG	CAAATTTCAGACTTTAGCATCACC	8	0.267747
EL410363	TGCTGCTTCTGCTTCAGTTTC	AAGATTCCATGGTGGTGGAG	4	0.347982
EL410627	TGGCAATACAACATGCACAC	TGGCTTTAACGACCTCAACC	5	0.110359