American Journal of Botany
Celebrating 100 years 1914–2014

American Journal of Botany 101(10): 1640–1650, 2014.

SPECIAL INVITED PAPER

Genetic diversity in *Carthamus tinctorius* (Asteraceae; safflower), an underutilized oilseed crop[1]

Stephanie A. Pearl[2,3] and John M. Burke[2,4]

[2]Department of Plant Biology, Miller Plant Sciences, University of Georgia, Athens, Georgia 30602, USA; and [3]Center for Applied Genetic Technologies, University of Georgia, Athens, GA 30602, USA

- *Premise of the study:* Underutilized crops are potentially valuable resources for meeting increasing food demands. Safflower, an oilseed crop, is an example of one such underutilized crop that thrives in moisture-limited areas. Characterization of the genetic diversity maintained within the gene pools of underutilized crops such as safflower is an important step in their further development.
- *Methods:* A total of 190 safflower individuals, including 134 USDA accessions, 48 breeding lines from two private North American safflower breeding companies, and eight wild safflower individuals, were genotyped using 133 single nucleotide polymorphism (SNP) markers. We then used the resulting data to assess the amount and distribution of genetic diversity within and among these collections of safflower.
- *Key results:* Although just a modest reduction in gene diversity was observed in the commercial breeding lines (relative to the other safflower groupings), safflower domestication was accompanied by a significant decrease in allelic richness. Further, our results suggest that most safflower breeding lines originated from a single pool of diversity within the Old World safflower germplasm.
- *Conclusions:* Taken together, our results suggest that both the safflower germplasm collection and related, wild species harbor previously undocumented genetic diversity that could help fuel future improvement efforts. Paired with analyses of functional diversity, the molecular resources described herein will be thus be useful in the continued development of safflower as an oilseed crop.

Key words: Asteraceae; *Carthamus tinctorius*; crop improvement; genetic diversity; germplasm; introgression; plant breeding; population structure; safflower; underutilized crop.

Underutilized crops are defined as those domesticated species whose genetic potential has not been fully realized (Padulosi and Hoeschle-Zeledon, 2004). These noncommodity crops are part of a "larger biodiversity portfolio" that tends to be underused by farmers and consumers for a variety of agronomic, economic, and cultural factors (Padulosi and Hoeschle-Zeledon, 2004). Given that food security is improved by the availability of a diverse assemblage of crop species, the development and production of underutilized crops has recently increased in priority (Mayes et al., 2012). Because these species are often adapted to cultivation on marginal lands, they also offer viable agricultural alternatives in response to climate change and provide farmers with additional options for maximizing land usage (Mayes et al., 2012). These crops also help satisfy an increasing demand for "natural" and environmentally friendly products while offering sources of diversified income to farmers and agricultural businesses (Thies, 2000).

The establishment and genetic characterization of germplasm collections is an important first step in securing and leveraging the resource base of underutilized crops (Padulosi et al., 1999). Such germplasm collections often include cultivated materials obtained from throughout the world and may also include closely related, wild species. These collections thus represent a potentially important source of genetic diversity for ongoing plant breeding efforts (Tanksley and McCouch, 1997). Unimproved landraces and wild germplasm may be particularly valuable sources of undiscovered alleles for the adaptation of crop plants to environmental challenges (McCouch et al., 2013). Unfortunately, little is often known about the genetic diversity within such collections, and their genetic potential often goes untapped.

Carthamus tinctorius L. (safflower; Asteraceae; $2n = 2x = 24$; Patel and Narayana, 1935) was domesticated approximately 4500 yr ago in the Fertile Crescent region from its putative wild progenitor, *Carthamus palestinus* Eig. (Van Zeist and Rooijen Waterbolk-Van, 1992; Knowles and Ashri, 1995; Chapman and Burke, 2007). Safflower was originally cultivated for the deep red pigments (carthamine) in its florets, which were used as a source of dye for various cultural purposes. Floral extracts have also been used as a food additive and are valued for their

[1]Manuscript received 26 February 2014; revision accepted 22 May 2014.

This work was supported by funding from the NSF Plant Genome Research Program (DBI-0820451 to J.M.B.). The authors thank the following people for help with obtaining plant materials and providing insights into safflower domestication history: R. Johnson and V. Bradley (USDA Western Regional Plant Introduction Station), A. Weisker and K. Ham (CalOils), J. Bergman (Safflower Technologies International), Y. Sapir (Tel Aviv University), A. Ashri (Hebrew University of Jerusalem), R. Hadas and S. Hadad (Israel Plant Gene Bank), and A. Susanna de la Serna (Botanical Institute of Barcelona). Also, the authors thank J. Roeder for assistance with DNA extractions, R. Nilsen and M. Belanger at the Georgia Genomics Facility for assistance with the SNP genotyping, and J. Lee and S. Ritter for help in the greenhouse.

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

doi:10.3732/ajb.1400079

supposed medicinal properties (Weiss, 2000). Following its domestication, safflower cultivation spread throughout the Middle East, northern Africa, India, and the Far East. In the late 1890s, safflower was introduced to North America where commercial production commenced in the 1950s. A previous study has described safflower as being "strongly domesticated" (Dempewolf et al., 2008) though, in our view, it exhibits only moderate phenotypic differentiation from its wild progenitor and has been the subject of relatively limited modern breeding.

Today, safflower is grown for its seed oils that are rich in unsaturated fatty acids and as a source of seeds for use in birdseed mixes. Its flowers are also occasionally sold in ornamental bouquets. However, safflower remains something of a niche crop with limited production in North America (FAO, 2013) and much of its production elsewhere is being done in the context of smallholder farms. In the mid-1990s, safflower was identified by the International Plant Genetic Resources Institute (IPGRI) and the German Agency for Technical Cooperation (GTZ) as one of 25 underutilized crops that should be the focus of further development (Dajue and Mündel, 1996; Thies, 2000). This interest in safflower was driven by its local and regional importance (e.g., both economically and as a staple food in underdeveloped countries such as India and Ethiopia), potential for socioeconomic and agricultural development throughout the world, adaptation to areas in which surface moisture is limited, and the danger of genetic erosion within the crop gene pool (Thies, 2000). Safflower is, however, somewhat of an "orphan" with respect to the genomics revolution (Varshney et al., 2012), and breeding efforts have consequently been hampered by a lack of molecular tools that could otherwise facilitate more rapid improvement. In recent years, however, this situation has begun to change (e.g., Johnson et al., 2007; Chapman et al., 2009; Mayerhofer et al., 2010; Pearl et al., 2014).

Herein, we describe the use of a collection of single nucleotide polymorphisms (SNPs) to characterize patterns and levels of nucleotide diversity across a broad cross section of the safflower gene pool. This cross section includes a representative, worldwide sampling of diversity from the USDA germplasm collection, lines from the major two private North American commercial safflower breeding programs, plus a set of wild safflower individuals. Using the resulting data, we explore the likely origins of the modern breeding materials and consider the utility of the available germplasm resources, including wild species, as possible sources of newly recognized genetic diversity for the advancement of safflower breeding programs.

MATERIALS AND METHODS

***Plant materials and genotyping*—**The focus of this study was a broad sampling of *Carthamus* germplasm ($N = 190$ individuals total), including eight wild (three *C. palaestinus* and five *C. persicus* from various sources; Appendix 1) and 182 cultivated safflower individuals. The latter included representatives of 96 geographically widespread Old World accessions and 38 New World (plus Australian) accessions obtained from the U. S. Department of Agriculture (USDA) Western Regional PI Station (Pullman, Washington; Appendix 2). The full set of 182 cultivated safflowers also included 48 lines donated by the two primary safflower breeding companies in North America (CalOils and Safflower Technologies International; referred to hereafter as CO and STI). There is a lack of consensus in the literature as to whether *C. persicus* and *C. palaestinus* are synonymous (Hanelt, 1963) or separate species (Garnatje et al., 2006), and there has even been speculation that *C. palaestinus* is a hybrid between *C. tinctorius* and *C. persicus* (Ashri and Knowles, 1960; Hanelt, 1963). However, authorities on Mediterranean floral taxonomy regard *C. palaestinus* as an invalid designation of *C. persicus* (Euro+Med, 2013); we thus treat *C. persicus* and *C. palaestinus* as synonymous for the remainder of this paper and generally refer to them as "wild safflower".

Because cultivated safflower is self-compatible and previous studies have shown that individual accessions within the USDA collection are genetically quite uniform (Johnson et al., 2007), just a single representative of each USDA accession or breeding line was used in our study. Of the 134 Old World and New World accessions, 81 are part of the USDA safflower core collection (Western Regional PI Station), and we included an additional 10 historically important New World accessions that were developed during the latter half of the 20th century (Appendix 2). Finally, the materials donated by safflower breeders included a total of 25 commercial varieties (including six dual use oil/birdseed cultivars with the balance being oil lines), 17 elite breeding lines, and 6 "germplasm conversion" lines that were produced by the breeding of CO lines with a diverse collection of cultivated safflower accessions.

Seeds were planted in the University of Georgia greenhouses, leaf tissue was collected from seedlings, and DNA was extracted using Qiagen (Valencia, California, USA) DNeasy Plant Mini Kits following the manufacturer's protocol. Single nucleotide polymorphisms (SNPs) were then genotyped using the Illumina (San Diego, California, USA) GoldenGate Assay described by Pearl et al. (2014) on an Illumina Bead Express at the Georgia Genomics Facility. Briefly, these SNPs were selected because they exhibited differences between the wild and cultivated mapping parents used in that earlier study. Finally, allele calls were obtained using the Illumina GenomeStudio software (ver. 2011.1).

***Population genetic and statistical analyses*—**For each of our groupings (wild, Old World, New World, CO, and STI), we estimated expected heterozygosity (Nei, 1978), observed heterozygosity, and the percentage polymorphic loci using the program GENALEX ver. 6.5b2 (Peakall and Smouse, 2006, 2012). To compare the number of private alleles and estimates of allelic diversity among each of our unequally sized study groups, we used rarefaction (Hurlbert and Jul, 1971; Petit et al., 1998; Kalinowski, 2004) as implemented in the program HP-RARE (Kalinowski, 2005). For each of these statistics, we estimated the significance of the differences among groups using Tukey's post hoc test, where group and locus were used as the model effects (Sokal and Rohlf, 1995). Additionally, we calculated these statistics for the pooled sets of Old and New World samples to obtain global estimates of diversity in wild vs. cultivated safflower, and we tested for significant differences between the two groups using Wilcoxon rank-sum tests. To estimate intrachromosomal pairwise linkage disequilibrium (LD) among markers used in this study, we generated a matrix of the squared allele frequency correlations (r^2) and plotted these values as a function of distance (in cM) using the R programming language (R Development Core Team, 2013). We then summarized the r^2 values using the "locpoly" and "dpill" functions in the R package "KernSmooth" (Wand, 2013).

Genetic structure among our five groups was assessed in GENALEX via analysis of molecular variation (AMOVA; Excoffier et al., 1992), which hierarchically partitioned genetic variation, estimated F_{ST} (Wright, 1949), and determined significance based on 999 permutations of the data. Additionally, population structure among the five safflower groupings was examined using the Bayesian, model-based clustering algorithm STRUCTURE ver. 2.3 (Pritchard et al., 2000). Initially, STRUCTURE was used to assign cluster membership of each sample without using geographic priors. To determine the most likely K (number of clusters), we followed the methods detailed by Evanno et al. (2005). STRUCTURE was used to perform 5 runs (each with 1 000 000 replicates following a 100 000 replicate burn-in) for each K from 1 through 12. For the most likely value of K, the proportion membership of each individual in each cluster was determined using the LargeKGreedy algorithm in CLUMPP v. 1.12, with up to 30 000 random input orders (Jakobsson and Rosenberg, 2007). Additionally, STRUCTURE results were depicted geographically in maps drawn using the R package "maptools" (Bivand and Lewin-Koh, 2013).

STRUCTURE was also used to investigate the origins of New World accessions and North American breeders' lines. To do this, the STRUCTURE analyses were repeated as detailed, but with only the Old World accessions as inputs. After determining the most likely number of clusters (K) within the Old World accessions, all individuals with greater than 80% membership in a single cluster were assigned to that cluster, and the results were used to inform a third analysis (Hubisz et al., 2009). This third analysis involved invoking the USEPOPINFO flag such that the aforementioned Old World individuals were used as a "training set" to assign the New World accessions and breeders' lines to likely Old World clusters of origin. This same procedure was also used to investigate which Old World cluster corresponded most closely to the wild samples.

Genetic relationships among populations were also visualized using a principal coordinate analysis (PCoA) in GENALEX, which involved using the

multilocus genotypes of all 190 individuals to create a standard genetic distance matrix (Nei, 1978). Principal coordinates were then estimated based on this matrix, and the first two coordinates were plotted in two-dimensional space. Relationships among all individuals were further assessed by constructing a neighbor-joining tree with the program POPTREE2 (Takezaki et al., 2010) using 500 bootstrap replicates. Trees were visualized using the program FigTree ver. 1.3.1 (Rambaut, 2006–2009).

RESULTS

Genetic diversity and linkage disequilibrium—Of the previously identified 244 SNPs with known genetic map positions (Pearl et al., 2014), we selected 133 that were clearly interpretable and did not colocalize with one another. Estimates of genetic diversity within our five groups of wild and cultivated safflowers were relatively high. Nei's unbiased heterozygosity (H_e) (which could not exceed 0.50 because all loci were biallelic) ranged from 0.126 (CO) to 0.271 (wild) (overall mean ± SE = 0.226 ± 0.017; Table 1). In contrast, the average observed heterozygosity was much lower, ranging from 0.013 (CO) to 0.048 (STI) (overall mean = 0.036 ± 0.002; Table 1). Although the observed heterozygosity estimate differed significantly when comparing the wild and pooled cultivated safflower samples (Z = −4.98, $P < 0.001$), the expected heterozygosity estimate did not (Table 2). The mean percentage of polymorphic loci across groups was 70.54% ± 5.8% (range = 55.64% [CO] to 81.2% [wild]; Table 1).

The private allelic richness (based on rarefaction) was significantly higher in the wild group when compared with all other groups individually, as well as when compared against the pooled sample of cultivars (Tables 1, 2). When considering only the four groups comprising the *C. tinctorius* subset, four loci had alleles private to the New World grouping and another four loci had alleles unique to the Old World grouping. However, these alleles were all present in the wild safflower samples, and we found a total of 23 additional private alleles in the wild safflower grouping. Further, 25 alleles not found in the wilds were private to the pooled cultivated safflowers. Meanwhile, the rarefied allelic richness (A_g) was significantly higher in the wild grouping when compared with all other groups individually, as well as in the pooled analysis (A_g ± SE = 1.82 ± 0.034), and the CO group had the lowest estimate of allelic richness (1.43 ± 0.037; Tables 1, 2).

Analysis of the subgroups within the CO breeding lines revealed that the germplasm conversion lines were significantly more homozygous than the elite lines and commercial varieties ($F_{2,130} = 1.66$, $P = 0.0003$), but were not significantly more diverse (Appendix S1, see Supplemental Data with online version of this article). Additionally, H_e, A_g, and private allelic richness were not significantly different among any of these groups (Appendix S1). Interestingly, the germplasm conversion lines had the lowest percentage of polymorphic loci (Appendix S1).

Overall levels of LD were generally low. The average intrachromosomal LD was less than 0.1, with the exceptions of linkage groups F and L. Although the extent of LD varied somewhat across linkage groups (Appendix S2, see online Supplemental Data), LD diminished to less than 0.1 within 9 cM and sometimes within 1 cM (linkage groups D and J). Note that, for some linkage groups (i.e., B and I), it was not possible to summarize LD via the KernSmooth function due to a paucity of SNPs. Interestingly, we found surprisingly high pairwise LD ($r^2 = 0.365$ and 0.304) between opposite ends of linkage group L, a distance spanning over 100 cM.

Population structure, relationships—Among the five predefined groups investigated in this study, F_{ST} as estimated from AMOVA ranged from 0.070 (between New World and Old World; $P = 0.001$) to 0.712 (between wild safflowers and the CO breeding lines; $P = 0.001$; online Appendix S3). Our STRUCTURE analyses of all wild and cultivated safflowers indicated $K = 2$ was the most likely number of clusters (online Appendix S4A, B), with one cluster largely corresponding to the wild individuals and Old World accessions and the other cluster mostly corresponding to the New World accessions and breeders' lines (online Appendix S5). Examination of the next most likely result ($K = 10$, Appendix S4A, B; Fig. 1) revealed a much more nuanced picture: the wild individuals formed their own cluster (cluster 1), the Old World accessions grouped into six different clusters (clusters 2 through 7), two of which jointly grouped with several New World accessions (clusters 6 and 7). Each of these clusters was generally characterized by a predominant geographic region: cluster 2 = Israel, Jordan, and Ethiopia; cluster 3 = Europe; cluster 4 = Iran, Afghanistan, and Turkey; cluster 5 = Sudan and southern Asia; cluster 6 = Far East; and cluster 7 = Egypt and Sudan (Figs. 1, 2A). Two additional clusters included one set of New World accessions plus the majority of the North American breeders' lines (cluster 8) and a separate cluster composed of a subset of New World accessions (cluster 9). Finally, a subset of the STI breeding lines formed their own cluster (cluster 10). An analysis of wild and cultivated safflowers on a per-cluster basis (and excluding those accessions with less than 50% membership in any one cluster) revealed that allelic richness was greatest in cluster 4 (corresponding to the Iran,

TABLE 1. Genetic diversity statistics for the wild, cultivated (Old World and New World) and commercial (CO and STI) safflower groupings.

Population	*N*	%P	A_g (±SE)	PAL	H_o	H_e
Wild	8	81.2	1.82^a (0.034)	0.25^a	0.040^a	0.271^a
Old World	96	79.7	1.68^b (0.037)	0^b	0.042^a	0.266^a
New World	38	78.95	1.67^{bc} (0.035)	0^b	0.038^a	0.247^a
CO	34	55.64	1.43^d (0.037)	0^b	0.013^b	0.126^b
STI	14	57.14	1.55^c (0.042)	0^b	0.048^a	0.221^a

Notes: *N* = number of plants sampled, %P = percent polymorphic loci, A_g = allelic richness (based on the rarefaction method) averaged across all loci ($F_{2,130} = 4.08$), PAL = private allelic richness (based on the rarefaction method) averaged across all loci ($F_{2,130} = 2.48$), H_o = observed heterozygosity averaged across all loci ($F_{2,130} = 2.17$), H_e = expected heterozygosity averaged across all loci ($F_{2,130} = 3.84$). Letters indicate differences in significance levels ($P < 0.001$).

TABLE 2. Genetic diversity statistics comparing wild safflower with cultivated accessions from the USDA (Old World and New World combined).

Population	*N*	%P	A_g (±SE)	PAL	H_o	H_e
Wild	8	82.71	1.82* (0.034)	0.31*	0.040	0.271
USDA cultivated (Old World + New World)	134	81.20	1.69* (0.036)	0.18*	0.041	0.256

Notes: *N* = number of plants sampled, %P = percent polymorphic loci, A_g = allelic richness (based on the rarefaction method) averaged across all loci (*Z* = 8.55), PAL = private allelic richness (based on the rarefaction method) averaged across all loci (*Z* = 7.83), H_o = observed heterozygosity averaged across all loci (*Z* = −4.98), H_e = expected heterozygosity averaged across all loci (*Z* = −0.50). * Values in categories are significantly different ($P < 0.001$).

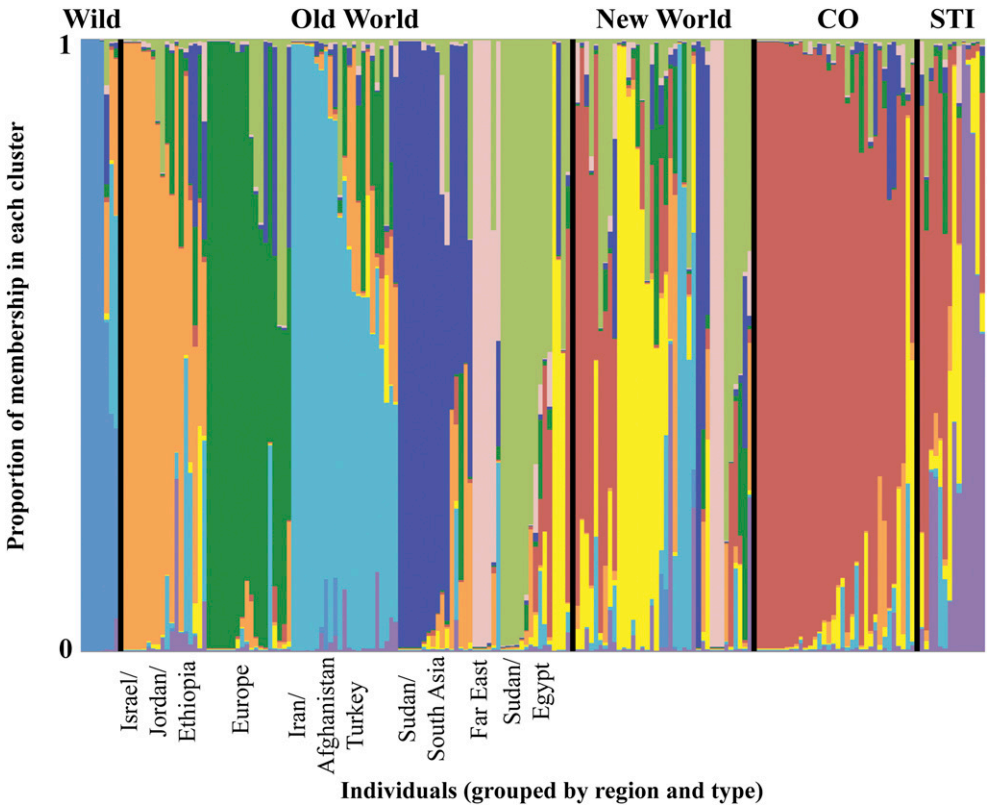

Fig. 1. STRUCTURE plot of 190 wild and cultivated safflower individuals. Black bars are used to separate predefined groupings (labeled at top of graph). Each vertical bar represents a single individual, and the proportion of membership in each cluster is indicated on the *y*-axis. Here, $K = 10$ clusters, with geographic origins of the majority of each cluster indicated along the bottom of the graph; dark blue = cluster 1, orange = cluster 2, green = cluster 3, light blue = cluster 4, royal blue = cluster 5, pink = cluster 6, light green = cluster 7, red = cluster 8, yellow = cluster 9, purple = cluster 10.

Afghanistan, Turkey cluster; rarefied $A_g = 1.54 \pm 0.035$) and lowest in cluster 6 (the Far Eastern cluster; $A_g = 1.15 \pm 0.029$; data not shown).

Our separate STRUCTURE analysis of Old World accessions to identify a "training set" for subsequent population assignment yielded a most likely result of $K = 4$ clusters (Appendix S4C, D). These clusters largely corresponded to the four Old World clusters described, with the exception that cluster 7 (Sudan and Egypt) grouped with cluster 2 (Israel, Jordan, and Ethiopia) and that the Far Eastern cluster (cluster 6) was split and had individuals grouping with either cluster 2 or 5 (Sudan and southern Asia; online Appendix S6). The subsequent population assignment analysis in STRUCTURE assigned each of the New World accessions and breeders' lines to a mixture of the four Old World clusters, though the greatest proportion of each individual corresponded to the joint Jordan/Israel/Ethiopia–Egypt/Sudan cluster (Fig. 3). Meanwhile, the Iran/Afghanistan/Turkey cluster showed the highest level of similarity with the wild individuals, followed by the joint Israel/Jordan/Ethiopia–Egypt/Sudan clusters (Fig. 3).

The results of the PCoA and the neighbor-joining analyses (Fig. 4 and online Appendix S7) were largely congruent with the STRUCTURE results and reflected our F_{ST} estimates, showing that the wild safflowers were largely distinct from cultivated safflowers. Interestingly, the longest branch in the neighbor-joining tree separated the wild individuals recently collected in Israel from the remaining *C. persicus* (Barcelona) and *C. palaestinus* (USDA) accessions (Appendix S6). In both the PCoA and neighbor-joining analyses, all wild samples clustered nearest to the Iran/Afghanistan/Turkey Old World population, which corresponds to safflower's putative center of origin. Finally, although Old World accessions from different geographic regions exhibited overlap (Fig. 4; Appendix S6), geographic structuring was still apparent.

DISCUSSION

An understanding of the amount and distribution of genetic diversity within crop germplasm collections can provide valuable insight into the evolutionary history of the species in question and help to guide future improvement efforts (Tanksley and McCouch, 1997). An important caveat in the present study was the limited availability of wild safflower samples, which likely resulted in an undersampling of the diversity present in wild safflower populations and may have driven the high F_{ST} values observed between the wild and all other cultivated safflower groupings (online Appendix S3). Despite this limited sampling, our analyses revealed an overall reduction in diversity in cultivated vs. wild safflower (Tables 1, 2). The low levels of observed heterozygosity were consistent with a history of inbreeding due to the self-compatibility of safflower and its wild relatives (Claassen, 1950) as well as the breeding history of cultivated safflower.

Within the CO breeding lines, the germplasm conversion lines were significantly more inbred than the varieties and elite lines, and surprisingly, introgression from the wild has failed to produce the expected infusion of molecular diversity

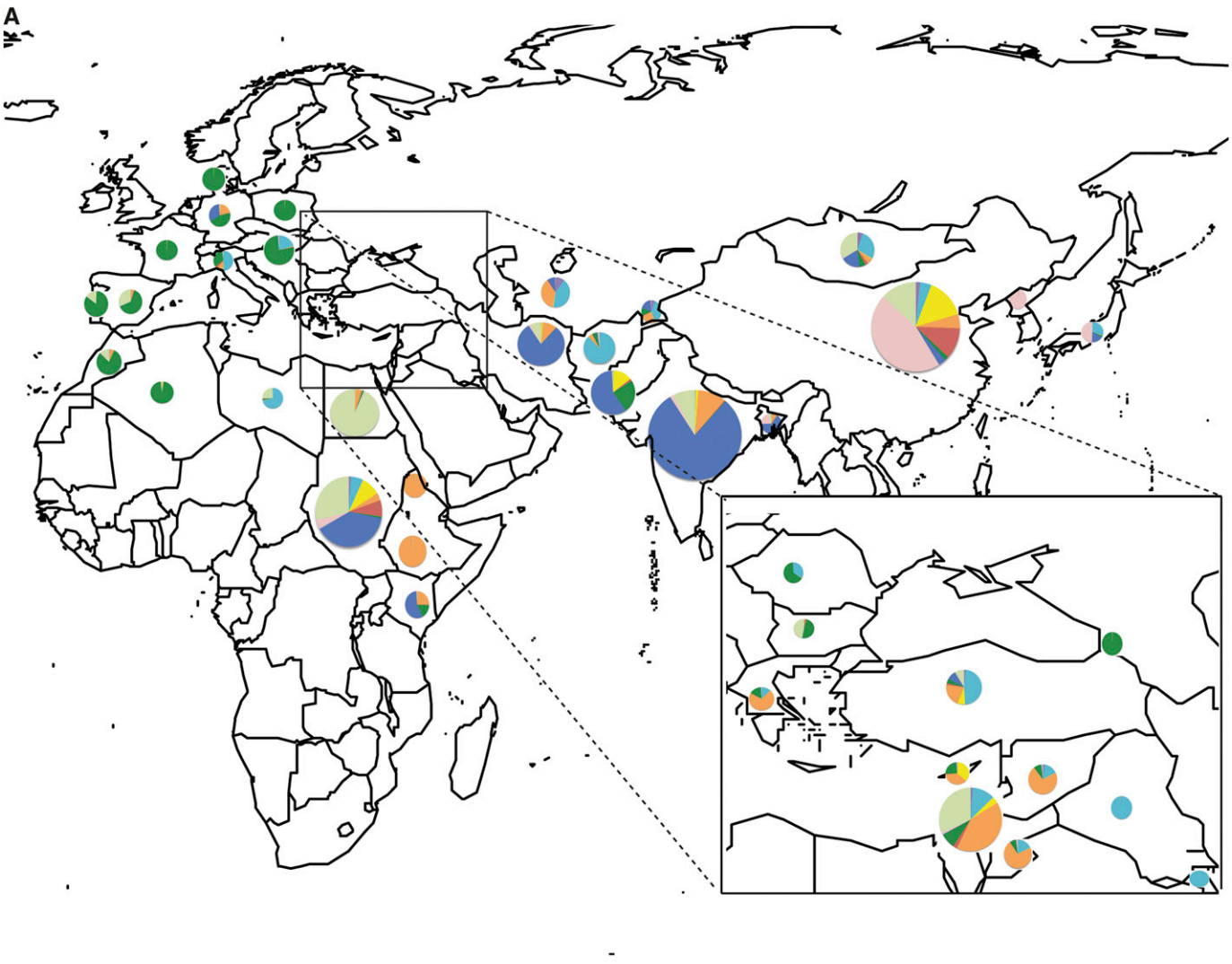

Fig. 2. Map of the sampling locations (A) Old World and (B) New World USDA accessions and corresponding assignment to each the 10 clusters depicted in Fig. 1. Pie charts are placed on the map to represent samples collected from each country, sizes of the pie charts represent the number of samples collected in each country (as indicated in the legend), and colors are consistent with Fig. 1.

(Appendix S1). Meanwhile, the STI breeding lines were more differentiated, falling predominantly into one of two distinct clusters within the STRUCTURE analysis and neighbor-joining tree (Fig. 1, Appendix S7). This division largely corresponded to differences in market type—i.e., most of the high oleic lines fell in one grouping, while the linoleic, birdseed, and subset of oleic lines clustered in the other.

Our STRUCTURE analysis of all 190 samples partitioned the data into as many as 10 genetically distinct clusters that largely corresponded with geography and/or breeding history (Figs. 1, 2). Within our Old World grouping of safflower accessions, STRUCTURE identified four clusters that corresponded to four different geographic regions (Appendix S6) that presumably represent somewhat distinct breeding pools. These clusters correspond quite closely with five centers of safflower diversity previously identified by Chapman et al. (2010). Our population assignment analysis investigating the origins of New World accessions and commercial breeding lines grouped these samples primarily with the cluster predominantly comprising individuals from Israel, Jordan, Ethiopia, Egypt, and Sudan (Fig. 3). Within the Old World grouping in the original STRUCTURE analysis (Fig. 1), a few accessions from China and Sudan (the predominantly yellow bars far to the right within that grouping) are largely assigned to the same cluster as many of the New World and North American breeding lines, perhaps exhibiting similarities to the original source material from which several of these New World and North American breeding lines were derived. The PCoA likewise showed that a Sudanese and a Chinese accession exhibited a large amount of overlap with North American breeding lines (Fig. 4).

Our original STRUCTURE analysis also revealed that wild safflowers formed a largely distinct cluster (Fig. 1), perhaps owing to the substantial number of private alleles that were found within this group. However, our population assignment analysis in STRUCTURE and subsequent PCoA and neighbor-joining analyses suggested that the wild safflowers shared the greatest similarity with the Iran-Afghanistan-Turkey cluster in the Old World (Figs. 3, 4; Appendix S7). This finding is consistent with

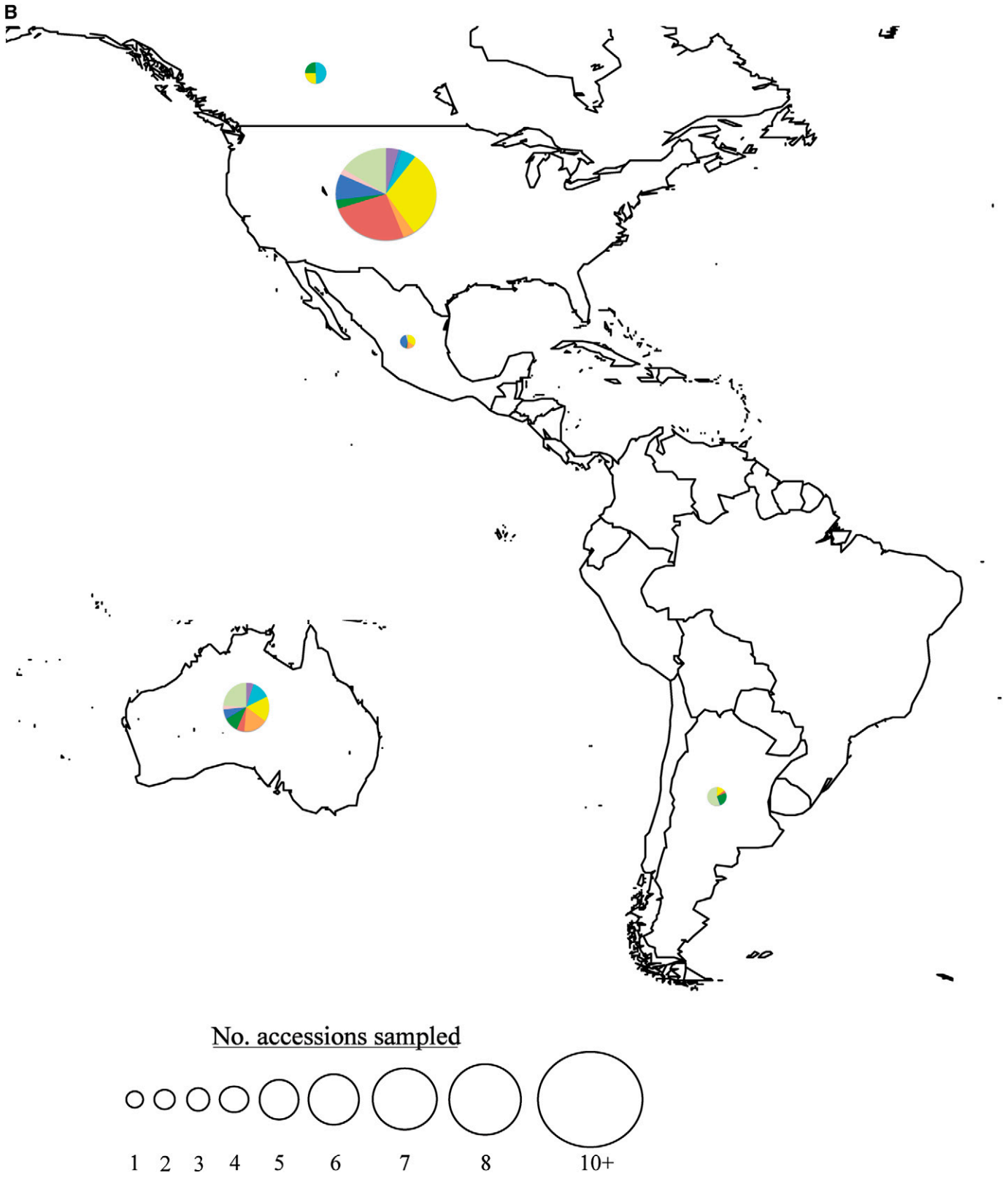

Fig. 2. Continued.

safflower's presumed Near Eastern center of origin (Van Zeist and Rooijen Waterbolk-Van, 1992; Chapman et al., 2010). Interestingly, the longest branch of the neighbor-joining tree (and the only branch with 100% bootstrap support) separated the wild safflowers collected in Israel (obtained from the Israel Plant Gene Bank, Dr. Yuval Sapir and Dr. Amram Ashri) from

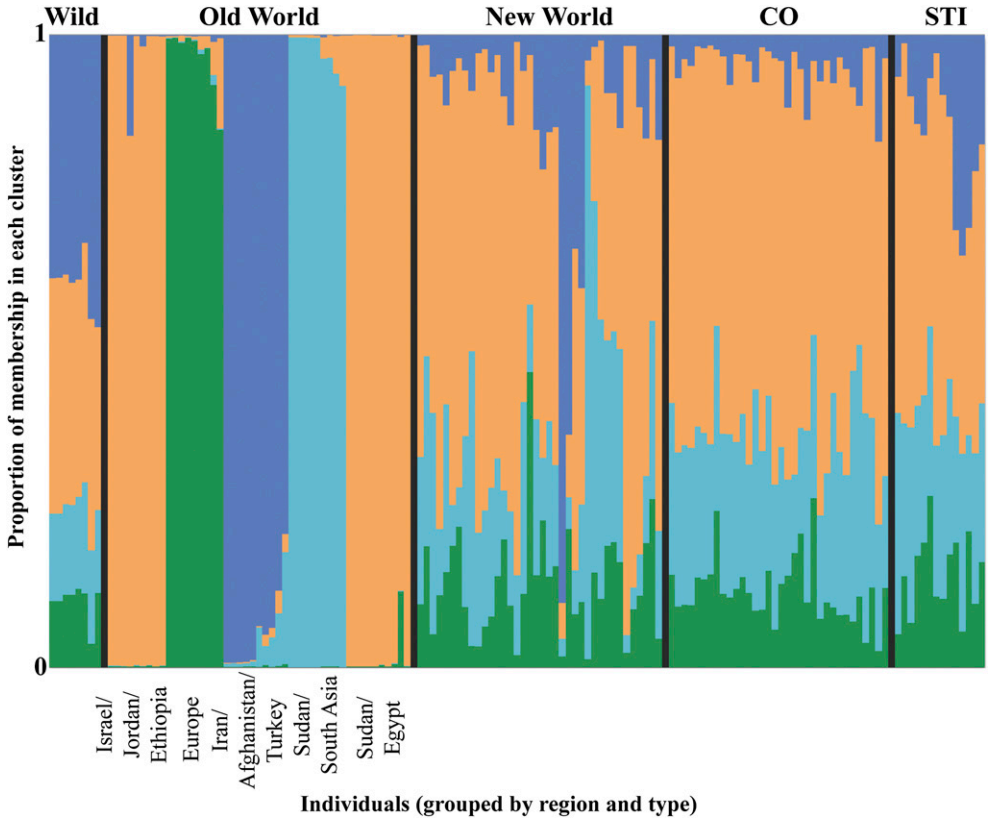

Fig. 3. STRUCTURE plot of 141 wild and cultivated safflower individuals in which wild, New World, CO, and STI accessions are assigned to one of four Old World populations. Black bars are used to separate predefined groupings (labeled at top of graph). Each vertical bar represents a single individual, and the proportion of membership in each population is indicated on the *y*-axis. Old World accessions with less than 80% membership in any of these four populations were excluded from the analysis. Colors correspond to Appendix S6, in which orange = cluster 2, 7, and part of 6; green = cluster 3, light blue = cluster 4, and royal blue = cluster 5 and part of 6.

the wild safflowers obtained from the Botanical Institute of Barcelona and the USDA (Appendix S7), with the former being placed at the distal end of that branch. It appears that the wild private alleles are concentrated within the IGB and recently collected wild individuals, driving this separation between sets of wild individuals in the neighbor-joining tree.

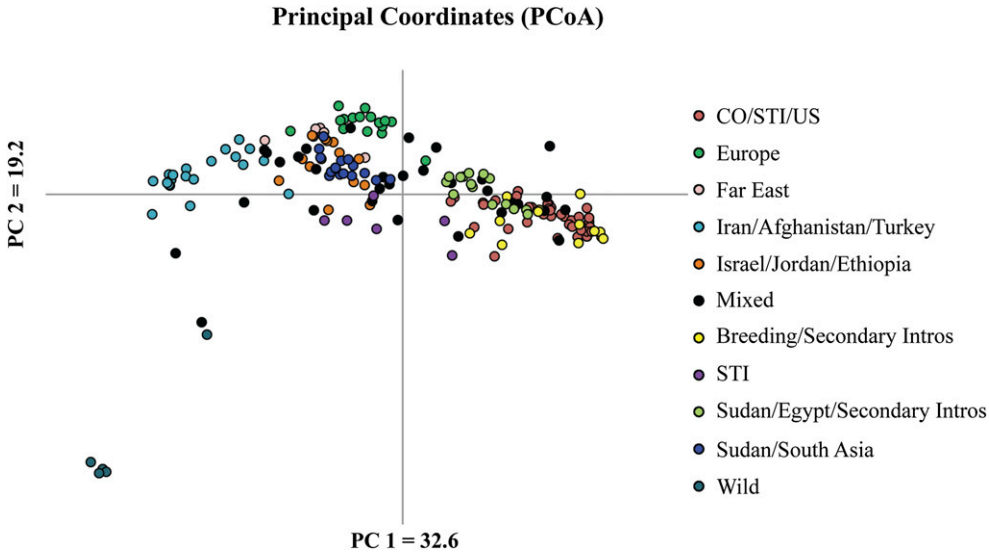

Fig. 4. Principal coordinate analysis (PCoA) of 190 *Carthamus* individuals. Color-coding corresponds to the 10 clusters depicted in Fig. 1. Black circles represent individuals that have less than 50% membership in any given cluster and therefore a greater proportion of their genotypes are mixes.

Although our analyses have revealed a rather limited reduction in overall genetic diversity within commercial breeding pools, this effect was likely underestimated due to the aforementioned limited sampling of the wild gene pool. Moreover, the elevated private allelic richness in the wild safflower grouping highlights their potential utility as a new source of allelic diversity. As such, molecular tools such as those described herein could help to guide the selection of germplasm for prebreeding efforts. Our results suggest that many of the allelic variants present in the wild were left behind during safflower domestication and subsequent breeding. In this light, it is also worth noting that a prior study of the genetic basis of domestication-related traits in safflower revealed the presence of numerous QTL with antagonistic effects (i.e., genomic regions in which the wild allele produces a more crop-like phenotype and vice versa; Pearl et al., 2014). As such, it appears that there are, indeed, agronomically favorable alleles present in wild safflower. While it is true that molecular diversity may not be an accurate predictor of phenotypic diversity (Reed and Frankham, 2001), it seems that expanded efforts to access wild genetic diversity would facilitate the continued improvement of safflower, as it has done for numerous other crops (e.g., tomato, potato, rice, and wheat; reviewed by Hajjar and Hodgkin, 2007). The continued generation of genomic resources for safflower could also facilitate its continued development in the same way that such resources have aided the improvement of the world's most important crops (e.g., maize, rice, and wheat; Varshney et al., 2012).

LITERATURE CITED

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

BIVAND, R., AND N. LEWIN-KOH. 2013. maptools: Tools for reading and handling spatial objects. Available at http://cran.r-project.org/package=maptools.

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, AND J. M. BURKE. 2010. Population genetic analysis of safflower (*Carthamus tinctorius*; Asteraceae) reveals a Near Eastern origin and five centers of diversity. *American Journal of Botany* 97: 831–840.

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.

CLAASSEN, C. E. 1950. Natural and controlled crossing in safflower, *Carthamus tinctorius* L. *Agronomy Journal* 42: 381–384.

DAJUE, L., AND H.-H. MÜNDEL. 1996. Safflower. *Carthamus tinctorius* L. Promoting the conservation and use of underutilized and neglected crops. Monograph 7. Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany/International Plant Genetic Resources Institute, Rome, Italy.

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.

EURO+MED. 2013. Euro+Med PlantBase. Available at http://ww2.bgbm.org/EuroPlusMed/.

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.

EXCOFFIER, L., P. E. SMOUSE, AND J. M. QUATTRO. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics* 131: 479–491.

FAO [FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS]. 2013. FAOSTAT. Available at http://faostat.fao.org.

GARNATJE, T., S. GARCIA, R. VILATERSANA, AND J. VALLÈS. 2006. Genome size variation in the genus *Carthamus* (Asteraceae, Cardueae): Systematic implications and additive changes during allopolyploidization. *Annals of Botany* 97: 461–467.

HAJJAR, R., AND T. HODGKIN. 2007. The use of wild relatives in crop improvement: A survey of developments over the last 20 years. *Euphytica* 156: 1–13.

HANELT, P. 1963. Monographische ubersicht der gattung *Carthamus* L. (Compositae). *Feddes Repertorium* 67: 41–180.

HUBISZ, M. J., D. FALUSH, M. STEPHENS, AND J. K. PRITCHARD. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9: 1322–1332.

HURLBERT, S. H., AND N. JUL. 1971. The nonconcept of species diversity: A critique and alternative parameters. *Ecology* 52: 577–586.

JAKOBSSON, M., AND N. A. ROSENBERG. 2007. CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23: 1801–1806.

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

KALINOWSKI, S. T. 2004. Counting alleles with rarefaction: Private alleles and hierarchical sampling designs. *Conservation Genetics* 5: 539–543.

KALINOWSKI, S. T. 2005. HP-Rare 1.0: A computer program for performing rarefaction on measures of allelic richness. *Molecular Ecology Notes* 5: 187–189.

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

MAYERHOFER, R., C. ARCHIBALD, V. BOWLES, AND A. G. GOOD. 2010. Development of molecular markers and linkage maps for the *Carthamus* species *C. tinctorius* and *C. oxyacanthus*. *Genome* 53: 266–276.

MAYES, S., F. J. MASSAWE, P.G. ALDERSON, J. A. ROBERTS, S. N. AZAM-ALI, AND M. HERMANN. 2012. The potential for underutilized crops to improve security of food production. *Journal of Experimental Botany* 63: 1075–1079.

MCCOUCH, S., G. BAUTE, J. BRADEEN, P. BRAMEL, P. BRETTING, E. S. BUCKLER, J. M. BURKE, ET AL. 2013. Agriculture: Feeding the future. *Nature* 499: 23–24.

NEI, M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.

PADULOSI, S., T. HODGKIN, J. T. WILLIAMS, AND N. HAQ. 1999. Underutilized crops : Trends, challenges and opportunities in the 21st Century. *In* J. Janick [ed.], Perspectives on new crops and new uses, 140–145. American Society for Horticultural Science (ASHS) Press, Alexandria, Virginia, USA.

PADULOSI, S., AND I. HOESCHLE-ZELEDON. 2004. Underutilized plant species: What are they? *Leisa Leusden* 20: 5–6.

PATEL, J., AND G. NARAYANA. 1935. Chromosome numbers in safflower. *Current Science* 4: 412.

PEAKALL, R., AND P. E. SMOUSE. 2006. GENALEX 6: Genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288–295.

PEAKALL, R., AND P. E. SMOUSE. 2012. GENALEX 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—An update. *Bioinformatics* 28: 2537–2539.

PEARL, S. A., J. E. BOWERS, S. REYES-CHIN-WO, R. MICHELMORE, AND J. BURKE. 2014. Genetic analysis of safflower domestication. *BMC Plant Biology* 14: 43.

PETIT, R., A. E. L. MOUSADIK, AND O. PONS. 1998. Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* 12: 844–855.

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

R DEVELOPMENT CORE TEAM. 2013. R: A language and environment for statistical computing. Available at http://www.r-project.org.

RAMBAUT, A. 2006–2009. FigTree: Tree figure drawing tool. Available at http://tree.bio.ed.ac.uk/software/figtree/.

REED, D. H., AND R. FRANKHAM. 2001. How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* 55: 1095–1103.

SOKAL, R., AND F. ROHLF. 1995. Biometry, 3rd. ed. W.H. Freeman, New York, New York, USA.

TAKEZAKI, N., M. NEI, AND K. TAMURA. 2010. POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Molecular Biology and Evolution* 27: 747–752.

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

THIES, E. 2000. Promising and underutilized species, crops and breeds. Deutsche Gesellschaft für Technische Zusammenarbeit (GTZ) GmbH, Eschborn, Germany.

VARSHNEY, R. K., J.-M. RIBAUT, E. S. BUCKLER, R. TUBEROSA, J. A. RAFALSKI, AND P. LANGRIDGE. 2012. Can genomics boost productivity of orphan crops? *Nature Biotechnology* 30: 1172–1176.

WAND, M. 2013. KernSmooth: Functions for kernel smoothing for Wand & Jones (1995). Available at http://cran.r-project.org/package=KernSmooth.

WEISS, E. A. 2000. Safflower. *In* E. A. Weiss [ed.], Oilseed crops, 93–129. Blackwell Science, London, UK.

WRIGHT, S. 1949. The genetical structure of populations. *Annals of Eugenics* 15: 323–354.

VAN ZEIST, W., AND W. ROOIJEN WATERBOLK-VAN. 1992. Two interesting floral finds from third millenium BC Tell Hammam et-Turkman, northern Syria. *Vegetation History and Archaeobotany* 1: 157–161.

APPENDIX 1. Collection information for wild safflower (*Carthamus*) samples used in this study.

Species	Identifier	*N*	Collection site	Source
C. palaestinus	PI 235663	2	N/A	USDA
C. palaestinus	Ashri	1	Revivim, Israel	Amram Ashri
C. persicus	S-2358	1	Elazig, Turkey	Botanical Institute of Barcelona
C. persicus	23666	2	Arava Valley, Israel	Israel Plant Gene Bank
C. persicus	Sapir	2	Negev Desert, Israel	Yuval Sapir

Notes: *N* = No. individuals sampled in this study.

APPENDIX 2. List of cultivated safflower accessions obtained from the USDA for use in this study.

Ac. No.	Country of origin	Notes [a]	Cluster [b]
PI 181866	Syria	Cultivated, Core	2
PI 193473	Ethiopia	Cultivated, Core	2
PI 198844	France	Cultivated	3
PI 198990	Israel		2
PI 199889	India	Cultivated, Core	5
PI 208677	Algeria		3
PI 209287	Romania	Cultivated, Core	3
PI 209297	Kenya	Cultivated, Core	5
PI 209300	Kenya	Cultivated, Core	5
PI 220647	Afghanistan	Cultivated, Core	4
PI 226993	Israel	Core	2
PI 235658	Australia	Cultivated	2
PI 237547	Sudan	Cultivated, Core	7
PI 237548	Sudan	Cultivated, Core	4/5
PI 239042	Morocco		3
PI 239226	Spain	Cultivated, Core	3
PI 242419	Australia		7/3/5
PI 243070	Jordan	Cultivated	2
PI 248625	Pakistan	Cultivated, Core	5
PI 250081	Egypt	Cultivated	7
PI 250202	Pakistan	Cultivated, Core	5
PI 250533	Egypt	Cultivated	7
PI 250537	Egypt	Cultivated, Core	7
PI 250611	Egypt	Cultivated	7
PI 250833	Iran	Cultivated, Core	4/3
PI 251262	Jordan	Wild (Knowles)	2
PI 251290	Israel	Cultivated	2
PI 251291	Jordan	Cultivated	4
PI 251398	Iran	Wild (Knowles), Core	4
PI 251984	Turkey	Cultivated, Core	4
PI 253386	Israel	Wild (Knowles)	4
PI 253523	Italy	Cultivated, Core	4/3
PI 253531	Bulgaria	Cultivated, Core	3/7
PI 253538	Armenia	Cultivated	3
PI 253540	Hungary	Cultivated, Core	4
PI 253541	Hungary	Cultivated, Core	3
PI 253543	Poland	Cultivated, Core	3
PI 253548	Denmark	Cultivated	3
PI 253559	Portugal	Cultivated	3
PI 253560	Morocco	Cultivated	3
PI 253759	Iraq	Cultivated, Core	4
PI 253908	Afghanistan	Wild (Knowles), Core	4
PI 254976	Greece	Cultivated, Core	2
PI 257582	Ethiopia	Core	2
PI 258420	Portugal		3
PI 259992	Pakistan	Cultivated, Core	3
PI 260637	India	Cultivated, Core	5
PI 262420	Australia	Cultivated	7/10/2
PI 262423	Australia	Cultivated	4/7
PI 262430	Syria	Core	2
PI 262435	Uzbekistan		2/4/5
PI 262444	Kazakstan	Core	4
PI 268374	Afghanistan	Wild (Harlan), Core	4
PI 271070	Sudan	Cultivar, Core	5
PI 273876	Eritrea	Cultivated, Core	2
PI 279051	India	Cultivated, Core	5
PI 279342	Japan	Cultivated	6
PI 283764	India	Core	5
PI 286199	Kuwait	Core	4
PI 291600	Argentina	Cultivated, Core	7
PI 292003	Israel		7
PI 301048	Turkey	Cultivated, Core	4/7/9
PI 304408	Pakistan	Wild (Knowles), Core	9
PI 304503	Turkey	Wild (Knowles), Core	2/5
PI 304595	Afghanistan	Cultivated, Core	4
PI 305529	Sudan	Core	7
PI 305531	Sudan	Core	5
PI 305534	Sudan	Core	7
PI 305540	Kazakhstan	Core	5
PI 306599	Egypt	Wild (Knowles),	7

APPENDIX 2. Continued.

Ac. No.	Country of origin	Notes [a]	Cluster [b]
PI 306686	Israel		7
PI 306974	India	Wild (Knowles), Core	5
PI 307055	India	Wild (Knowles), Core	5
PI 312275	Hungary	Cultivated, Core	3
PI 314650	Kazakhstan	Wild (Jones), Core	7
PI 343930	Ethiopia	Cultivated	2
PI 348915	Canada	Cultivated, Core	4
PI 369843	Uzbekistan		2
PI 369847	Tajikistan	Core	4/2/5
PI 369853	Uzbekistan	Cultivar	4
PI 380800	Iran	Cultivated, Core	4
PI 386174	Syria	Cultivar, Core	2/4
PI 393499	Libya		4
PI 401479	Bangladesh	Wild (Hobbs), Core	5
PI 405984	Iran	Cultivated, Core	4
PI 406015	Iran	Cultivated, Core	4
PI 407624	Turkey	Cultivated, Core	4
PI 426523	Pakistan	Cultivated, Core	3
PI 451956	India	Cultivated, Core	5/7
PI 506427	China	Cultivar, Core	8
PI 514630	China	Cultivar, Core	6
PI 525457	US	"Girard," Cultivar, Core, Historic	8
PI 532619	Cyprus	Cultivated	2/9
PI 537608	US	Breeding, Core	7
PI 537626	US	Breeding, Core	9/4/3
PI 537636	US	Breeding, Core	7
PI 537652	Mexico	Breeding	5/9
PI 537659	US	Breeding, Core	5
PI 537682	US	Breeding, Core	9/8
PI 537692	US	"Gila," Cultivar, Core, Historic	9
PI 537695	US	"Ole," Cultivar, Historic	8/7
PI 538779	US	"Centennial," Cultivar, Pureline	8
PI 543995	China	Cultivated, Core	6
PI 544006	China	Cultivated, Core	6
PI 544033	China	Cultivated, Core	2/4
PI 544041	China	Cultivated, Core	9
PI 544052	China	Cultivated, Core	7
PI 560172	US	Breeding, Core	9
PI 560175	US	Breeding, Core	8
PI 560177	US	"Oleic Leed," Breeding, historic	9
PI 560192	US	Breeding, Core	10
PI 560200	US	Breeding, Core	7
PI 560205	US	"Mexico Dwarf," Breeding	5
PI 561703	US	"San Jose," Cultivar, Historic	8/7
PI 562638	India	Core	5
PI 568864	China	Cultivated	6
PI 572415	US	Cultivar	8/5
PI 572420	US	Cultivar	9
PI 572428	US	"Nebraska 10," Cultivar, Historic	9
PI 572435	US	"Dart," Cultivar, Historic	8
PI 572436	US	"Leed," Cultivar, Historic	9
PI 576991	Germany		3/5
PI 576992	North Korea		6
PI 613394	US	Cultivated, Core	4/9/10
PI 592391	Canada	"AC Sunset," Cultivar, Pureline, Historic	9/3
PI 601166	US	"Oker," Cultivar, Historic	8/9
PI 603208	US	"Lesaf," Breeding, Historic	9
PI 613465	Spain	Cultivated, Core	3
PI 613498	US	Cultivated	8/7
PI 613514	Australia	Cultivated	9
W6 39446	US	Cultivated, Winter hardy	4
PI 651878	US	Breeding pureline, Winter hardy	6
PI 651879	US	Breeding pureline, Winter hardy	6
PI 651880	US	Breeding pureline, Winter hardy	6

[a] Notes regarding each accession are provided by the USDA and include accession improvement status: Wild = not collected in a field or cultivated area; Cultivar = named cultivar developed by scientific means; Cultivated = collected from a field planting; Breeding = lines developed by scientific means and used in breeding programs; Landrace = locally adapted variety.

[b] Numbers correspond to the primary cluster (>50% membership) assigned to each sample by STRUCTURE. Individuals with multiple clusters listed had no predominant cluster of ancestry, and the multiple dominating clusters are listed.