Theor Appl Genet (1997) 95: 532–545

R. Cronn · M. Brothers · K. Klier · P.K. Bretting
J.F. Wendel

Allozyme variation in domesticated annual sunflower and its wild relatives

Received: 2 December 1996/Accepted: 4 April 1997

Abstract The annual sunflower (*Helianthus annuus* L.) is a morphologically and genetically variable species composed of wild, weedy, and domesticated forms that are used for ornament, oilseed, and edible seeds. In this study, we evaluated genetic variation in 146 germplasm accessions of wild and domesticated sunflowers using allozyme analysis. Results from this survey showed that wild sunflower exhibits geographically structured genetic variation, as samples from the Great Plains region of the central United States were genetically divergent from accessions from California and the southwestern United States. Sunflower populations from the Great Plains harbored greater allelic diversity than did wild sunflower from the western United States. Comparison of genetic variability in wild and domesticated sunflower by principal coordinate analysis showed these groups to be genetically divergent, in large part due to differences in the frequency of common alleles. Neighbor-Joining analyses of domesticated *H. annuus*, wild *H. annuus* and two closely related wild species (*H. argophyllus* T. & G. and *H. petiolaris* Nutt.) showed that domesticated sunflowers form a genetically coherent group and that wild sunflowers from the Great Plains may include the most likely progenitor of domesticated sunflowers.

Keywords *Helianthus annuus* · Allozymes · Crop evolution · Genetic diversity

Communicated by G.E. Hart

R. Cronn · K. Klier · J.F. Wendel
Department of Botany, Iowa State University, Ames, IA 50011, USA

M. Brothers · P.K. Bretting (✉)
USDA Agricultural Research Service, North Central Regional Plant Introduction Station, Department of Agronomy, Iowa State University, Ames, IA 50011, USA

Introduction

The annual sunflower (*Helianthus annuus* L., Asteraceae) is composed of a diverse assemblage of wild and weedy forms typically characterized by many branches, small capitulae (2–5 cm in diameter), and relatively small achenes (3–7 mm in length). Native to North America, the annual sunflower (herein *H. annuus* or simply "sunflower") covers an expansive range from the Pacific to the Atlantic coast, and from 50° N latitude to northern Mexico, with greatest populational concentrations in the western two-thirds of the United States. Across this range, sunflower displays an amazing variety of morphological variation, ranging in height from less than 1 m to over 4 m, and, having highly branched to unbranched stems, opposite to alternate leaf arrangements, and ray florets ranging in color from yellow to red. This diversity, combined with the ease with which *H. annuus* hybridizes with other *Helianthus* species, has generated considerable taxonomic and nomenclatural confusion; at least 25 specific epithets have been applied to some portion of *H. annuus* sensu lato (Rogers et al. 1982). Although wild sunflowers can be heterogeneous, some of this variation seems to be associated with geographic location. This has led to the description of four geographically coherent subspecies (Heiser 1954): *H. annuus* ssp. *jaegeri* Heiser from the southwestern United States.; *H. a.* ssp. *lenticularis* (Dougl.) Ckll. from the Rocky Mountains and western United States; *H. a.* ssp. *texanus* Heiser from the Gulf coast of Texas; and *H. a.* ssp. *annuus* from the central and eastern United States.

In addition to the wild forms of annual sunflower, domesticated forms also exist which have been treated as *H. a.* var '*macrocarpus*' (DC.) Ckll. These domesticates typically have a single large capitulum (5–30 cm in diameter), large achenes (8–15 mm long), and higher seed oil content (35–50%) than do wild sunflowers. Introduced from the United States into Europe in the

late 1500s, sunflowers initially gained popularity as a garden ornamental (Heiser 1976). The agronomic development of sunflower for oil ("oilseed" types) and edible achenes ("confectionery" types) occurred in eastern Europe and Russia, where by the late 1800s a number of landraces had been developed. Development of modern varieties has relied heavily upon interspecific introgression, initially for improved pathogen resistance (e.g., broomrape and downy mildew resistance from *H. tuberosus* L.; Pustovoit 1976) and, more recently, cytoplasmic male sterility for hybrid seed production (possibly from *H. petiolaris* Nuttall; Dominguez-Gimenez and Fick 1975; Siculella and Palmer 1988). In the United States, sunflower is a major crop in the northern Plains states, where it is grown on 1.1 million hectares (approx. 2.7 million acres) of arable land in North Dakota, Minnesota, and South Dakota (Anonymous 1996). The worldwide popularity of sunflower oil is evident by a global production of sunflower achenes that exceeded 2.3×10^7 metric tons in 1995. Eighty-nine percent of those achenes were processed to yield 8.5×10^6 metric tons of oil, accounting for 10% of the total worldwide vegetable oil production.

At present, the chronology and events leading to domestication of this important crop remain obscure. Heiser (1951, 1976, 1978) suggested that weedy forms of the annual sunflower spread from the southwestern to the central United States, where domestication occurred. A similar hypothesis (Lathrap, as cited in Heiser 1978) stated that sunflower was domesticated in the southwestern United States and later disseminated eastward. Although details of sunflower domestication remain unclear, achenes similar to modern domesticates have been recovered from archeological sites in midwestern states which date to 4,000–3,000 BP (Smith 1992; Crites 1993), providing evidence for an association with humans that extends far into the pre-Columbian past. Recently, Rieseberg and Seiler (1990) and Arias and Rieseberg (1995) investigated patterns of genetic diversity in wild, landrace, and modern cultivars via isozyme and randomly amplified polymorphic DNA (RAPD) markers. Although many loci were sampled in these studies (13 isozyme loci and 68 RAPD bands), the relationships between wild and domesticated sunflowers were sufficiently complex, and in some cases unexpected, that key details regarding sunflower domestication remain unanswered. Nevertheless, the high genetic identities reported for wild and domesticated sunflowers (> 0.90) lend support for a progenitor-derivative relationship. Additionally, early landraces (particularly Native American varieties) show a high degree of genetic similarity inter se, indicative of their common, shared origin.

Sunflower is unusual because it is one of the few crop species to have originated in temperate North America (Heiser 1978). Reflecting the uniqueness and importance of crop sunflowers, the U.S. National Plant Germplasm System (NPGS) collects, characterizes, maintains, evaluates, and distributes more than 3,000 germplasm accessions of cultivated, wild annual, and perennial *Helianthus* species. To characterize sunflower accessions from the NPGS collection, we used the amount and apportionment of isozyme variability in wild and domesticated *H. annuus* to quantify the range of genetic diversity, divergence, and redundancy in the collection. In this report, we describe genetic diversity at 20 allozyme loci for 146 sunflower accessions, including 32 wild populations, 3 ornamental forms, 3 native American landraces, 50 oilseed cultivars, 45 confectionery cultivars and 13 elite cultivars from commercial sources. Our results elucidate patterns of ecogeographic variation in wild sunflowers, quantify the severity of the genetic bottlenecks accompanying the domestication of sunflowers, and clarify interrelationships between wild and domesticated forms of sunflower. Finally, by comparing patterns of genetic diversity in *H. annuus* with two related species (*H. argophyllus* Torrey & Gray and *H. petiolaris* Nuttall), we find support for the hypothesis that domesticated *H. annuus* arose from sunflowers resembling those growing in the Great Plains.

Materials and methods

Plant materials

For this study, allozyme diversity was assayed from 32 wild and 101 domesticated accessions from the NPGS collection as well as from 13 accessions of "elite" (modern hybrid) cultivars from various sunflower seed producers (Table 1). Samples consisted of either original seed or seed derived from one increase cycle of controlled pollination subsequent to incorporation into the NPGS collection. Accessions were selected so that most of the geographical range of wild *H. annuus* was sampled, as were all stages of sunflower domestication (Native American landraces to elite cultivars) and agronomic specialization (oilseed and confectionery types).

Isozyme electrophoresis

Starch gel electrophoresis was performed on crude protein extracts of cotyledon tissue or achenes that had been imbibed for 24 h. A minimum of 5 to a maximum of 20 plants were analyzed per accession. Approximately 40 mg of tissue was homogenized in a 1.5-ml centrifuge tube with a power-driven acetal pestle (on ice) in 75 µl of solubilization buffer [75 m*M* sodium phosphate (pH 7.5), 100 m*M* sodium ascorbate, 10 m*M* sodium diethyldithiocarbamate, 10 m*M* dithioerythritol, 10% (w/v) PVP-40, 5% sucrose, 0.5% bovine serum albumin and 10 m*M* β-mercaptoethanol]. Extracts were frozen at -70°C until use.

Enzymes were separated on 12% starch gels and were visualized with methods detailed in Wendel and Weeden (1989). Twenty enzyme systems revealing a minimum of 30 loci were screened: acid phosphatase (*Acp1*: E.C. 3.1.3.2), aconitate hydratase (*Aco1*, *Aco2*; E.C. 4.3.1.3), alcohol dehydrogenase (*Adh1*, *Adh2*; E.C. 1.1.1.1), aspartate aminotransferase (*Aat1*, *Aat2*, *Aat3*; E.C. 2.6.1.1), catalase (*Cat1*; E.C. 1.11.1.6), glutamate dehydrogenase (*Gdh1*; E.C. 1.4.1.2), glyceraldehyde-3-phosphate dehydrogenase (*G3pdh1*, *G3pdh2*; E.C. 1.2.1.9), β-glucosidase (*Glu1*; E.C. 3.2.1.21), fluorescent esterase (*Est1*; E.C. 3.1.1.-), isocitrate dehydrogenase (*Idh1*, *Idh2*; E.C. 1.1.1.41),

leucine aminopeptidase (*Leu1*, *Leu2*; E.C. 3.4.11.1), NAD^+ malate dehydrogenase (*Mdh1*; E.C. 1.1.1.37), $NADP^+$ malate enzyme (*Me2*; E.C. 1.1.1.37), menadione reductase (*Mr2*; E.C. 1.6.99.-), phosphoglucose isomerase (*Pgi1*, *Pgi2*; E.C. 5.3.1.9), phosphoglucomutase (*Pgm1*, *Pgm2*; E.C. 5.4.2.2), phosphogluconate dehydrogenase (*Pgd2*; E.C. 1.1.1.44), ribulose-bisphosphate carboxylase (*Rbc1*; E.C. 4.1.1.39), shikimate dehydrogenase (*Skd1*; E.C. 1.1.1.25) and triose phosphate isomerase (*Tpi1*, *Tpi2*; E.C. 5.3.1.1). To resolve these loci, we used five different electrophoretic buffer systems, four of which were identical to those described in Wendel and Weeden (1989). Isozymes of ACP, AAT, CAT, GDH, and TPI were resolved with the lithium-borate (pH 8.3) electrode buffer/TRIS-citrate (pH 8.3) gel buffer system; MDH, ME, PGM, and SDK were resolved by using a histidine-citrate (pH 5.7) buffer system; IDH, MR, PGI, and 6-PGD were resolved with a TRIS-citrate (pH 7.0) buffer system; and EST, GLU and RBC were separated by using a TRIS-borate-EDTA (pH 8.6) buffer system. The final buffer system, used to resolve ACO, ADH, G3PDH, GDH, and LEU, consisted of an electrode buffer of 65 m*M* L-histidine–7 m*M* citrate (adjusted to pH 6.5 with citric acid), and a gel buffer consisting of 1 part electrode buffer to three parts water (= 16.5 m*M* histidine–1.9 m*M* citrate, pH 6.5).

Table 1 Accessions of wild and domesticated sunflowers surveyed in this study

PI[a]	Provenance[b]	Identity[c]	*N*[d]	PI	Provenance	Identity	*N*
Wild accessions							
413014	Woonsocket, S.D.	Wild-Plains	5.9	413102	Antioch, Calif.	Wild-California	7.9
413015	Spencer, S.D.	Wild-Plains	4.7	413107	Liberty Isl., Calif.	Wild-California	11.1
413016	Cody, Neb.	Wild-Plains	5.7	413110	Liberty Isl., Calif.	Wild-California	4.9
413021	Wyo.	Wild-Plains	4.4	413117	Rio Vista, Calif.	Wild-California	9.0
413023	Last Chance, Colo.	Wild-Plains	4.8	413127	Woodland, Calif.	Wild-California	5.9
413024	Limon, Colo.	Wild-Plains	4.8	413128	Woodland, Calif.	Wild-California	9.6
413033	Montrose, Kan.	Wild-Plains	7.3	413129	Woodland, Calif.	Wild-California	8.9
413034	Hastings, Neb.	Wild-Plains	5.0	413134	Woodland, Calif.	Wild-California	4.6
413035	Kearney, Neb.	Wild-Plains	4.9	413066	Obregon, Mexico	Wild-Southwest	7.1
413039	Gettysburg, S.D.	Wild-Plains	4.5	413067	Espana, Mexico	Wild-Southwest	6.1
413040	Hagrie, N.D.	Wild-Plains	4.8	413123	Mayo, Mexico	Wild-Southwest	9.1
413052	Route 58, Calif.	Wild-California	7.4	413153	Benson, Ariz.	Wild-Southwest	7.4
413070	Davis, Calif.	Wild-California	4.0	413155	Wilcox, Ariz.	Wild-Southwest	8.8
413078	Gustine, Calif.	Wild-California	4.4	413157	Lordsbourg, N.M.	Wild-Southwest	4.9
413079	Holtville, Calif.	Wild-California	2.9	413159	Mesilla, N.M.	Wild-Southwest	8.1
413084	Barstow, Calif.	Wild-California	4.5	413168	Sonora, Tex.	Wild-Southwest	3.9
Ornamental accessions							
A4297	Europe	Ornamental	10.9	A4309	Europe	Ornamental	6.6
A4302	Europe	Ornamental	8.8				
Native American accessions							
369357	United States	Arikara	4.4	369360	United States	Seneca	13.0
369358	United States	Havasupai	5.0				
Oilseed accessions							
257642	F.S.U.[e]	Vniimk 1646	4.6	291401	Hungary	Lovaspatanoi	5.5
262517	F.S.U.	Vniimk 8931	5.4	291407	Hungary	Hybride Laragues	4.7
287184	Chile	Vniimk 6540	6.8	296288	S. Africa	Franslever	5.0
287232	F.S.U.	Vniimk 6540	8.9	296289	S. Africa	Jupiter	4.5
287233	F.S.U.	Vniimk 8931	5.5	296292	S. Africa	Short Russian	13.6
291411	F.S.U.	Vniimk 8883	4.8	307936	F.S.U.	Yugovostok	4.9
307941	F.S.U.	Vniimk 6540	6.7	340789	F.S.U.	Krasnodarets	9.2
340780	F.S.U.	Vniimk 6540	4.8	371936	F.S.U.	Voshod	4.7
345612	F.S.U.	Vniimk 6540	10.8	377526	Kenya	Black	6.9
372259	F.S.U.	Vniimk 6540	1.9	378896	Argentina	Pehuen	3.8
265500	Colombia	Chermianka II	6.7	386320	F.S.U.	L2600	6.7
343786	Iran	Tchernianka W5	4.9	406646	Australia	Stepniak	5.0
343789	Iran	Tchernianka W8	4.8	408726	France	Relaxed Germpl.	4.9
343790	Iran	Tchernianka W9	7.8	430540	F.S.U.	Tambovskij	4.9
343791	Iran	Tchernianka W10	9.6	430541	F.S.U.	Progress	4.4
343794	Iran	Tchernianka W13	4.6	431507	Poland	T6556 1–2	5.0
287182	Chile	Peredovik	5.0	431508	Poland	T6558 1–1	5.6
287231	F.S.U.	Peredovik	9.4	431513	Romania	Romsun AD946	3.7
289622	France	Peredovik	5.5	431514	Romania	Romsun C5357	4.7
294659	Texas, USA	Peredovik	5.5	431516	Romania	Romsun N2004	4.9
372173	F.S.U.	Peredovik 304	7.6	431517	Romania	Romsun 09573	4.9
372178	F.S.U.	Peredovik 473	5.6	431518	Romania	Romsun P1384	4.8
262520	Poland	Jdanowskii	3.3	431519	Romania	Romsun V337	5.0
265101	F.S.U.	Armavirsky	5.6	431520	Romania	Romsun V1324	5.0
307934	F.S.U.	Armavirsky	5.1	431523	F.S.U.	46–86	9.7

Table 1 (continued)

PI[a]	Provenance[b]	Identity[c]	N[d]	PI	Provenance	Identity	N
Confectionery accessions							
162454	Uruguay	Sunrise	4.9	232905	Hungary	Mezohegyesi	9.4
162675	Argentina	Klein	9.6	232906	Hungary	Szabolesi	9.6
165087	Turkey	Ay Cicegi	8.9	240660	Germany	Spanners	5.8
170386	Turkey	Ay Cicegi	5.7	251901	F.S.U.	Jdanovsky	15.2
174217	Turkey	8149	6.8	251902	F.S.U..	Kruglik	2.7
175723	Turkey	Gonondu	11.0	265499	Colombia	Cinza 42	4.9
175727	Turkey	Alaca	5.9	265501	Chile	Saratov	5.0
175732	Turkey	Lodging Resist	7.1	284862	Poland	Oleifera	5.0
176974	Turkey	Lodging Resist	12.9	291402	Hungary	Szaratovskij	5.5
176975	Turkey	Katmer	5.0	291404	Hungary	Szaratovskij	4.9
181769	Lebanon	Tournesol	8.3	291406	Hungary	Rudorf	4.9
184048	Yugoslavia	Novosadski	4.4	291408	Hungary	Zeljonka 61	4.9
184049	Yugoslavia	Novosadski	7.6	296286	S. Africa	Comet	5.5
192943	China	Cultivated Type	6.5	296290	S. Africa	Laan	4.7
201812	Canada	Rust Resist	5.2	296291	S. Africa	Polestar	5.4
204578	Turkey	Orange Ray	7.8	296293	S. Africa	Universal	4.4
219649	Austria	Lodging Resist	7.8	307834	S. Africa	Beacon	5.0
219650	Austria	Lodging Resist	5.1	377527	Kenya	Comet	7.3
221441	Afghanistan	Aftab Parast	7.7	377528	Kenya	Dark Stripe	7.8
222991	Iran	Aftab Gartan	3.2	377529	Kenya	Gray Stripe	8.6
223671	Iran	Aftab Gardan	3.8	377530	Kenya	Kenya White	4.9
232903	Hungary	Iregszemcsei	5.2	406645	Australia	Polestar	5.5
232904	Hungary	Lovaspatonai	8.9				
Elite cultivars							
—	USA	ST316	7.6	—	USA	KeltGen KO66	6.3
—	USA	ST317	7.3	—	USA	KeltGen KO70	7.7
—	USA	ST330	8.0	—	USA	KeltGen K704.5	4.9
—	USA	Sunbred 254	10.2	—	USA	McCurdy 38–24	6.2
—	USA	Sunbred 262	7.8	—	USA	Groagri Indians	6.2
—	USA	Sunwheat 101	8.1	—	USA	Gjnerik VNK	7.9
—	USA	Sunwheat 102	7.8				
Wild Helianthus species							
413171	Texas	*H. argophyllus*	4.6	413175	Spencer, Neb.	*H. petiolaris*	7.0
494572	Padre Island, Tex.	*H. argophyllus*	4.8	451977	Richland Co., N.D.	*H. petiolaris*	4.6
494573	Port Aransas, Tex.	*H. argophyllus*	8.5	451978	Ellsworth Co., Kan.	*H. petiolaris*	3.8

[a] U.S. NPGS Plant Introduction accession number (if available). Accessions are maintained at the USDA/ARS North Central Regional Plant Introduction Station, Ames, Iowa
[b] Provenance for accession
[c] Group or cultivar name
[d] Average number of plants examined per locus
[e] F.S.U, Former Soviet Union

Genetic interpretations of isozyme and allozyme variation patterns of ADH (Torres 1975, 1983), MDH, 6-PGD, PGI, PGM, and IDH (Kahler and Lay 1985) were based on previously published reports. For the remaining polymorphic loci, banding patterns were interpreted according to three lines of evidence: (1) differences in isozyme patterns between predominantly homozygous breeding materials (kindly provided by Pioneer Hi-Bred Int, Johnston, Iowa) and heterozygous wild accessions; (2) typical patterns of subcellular localization and gene expression from other plants (Weeden and Wendel 1989); and (3) knowledge of quaternary structure of other homologous proteins (reviewed in Weeden and Wendel 1989). Loci encoding the most anodally migrating isozyme for each enzyme system were designated "1", with additional loci numbered sequentially in order of decreasing electrophoretic mobility. Similarly, alleles at each locus were numbered in order of decreasing mobility. This nomenclature is concordant with recent locus/allele designations for isozymes of *H. annuus* (e.g., Rieseberg et al. 1991; Rieseberg and Seiler 1990; Rieseberg and Soltis 1989) but is inverted relative to locus/allele designations by earlier authors (Torres 1975; Kahler and Lay 1985).

Data analysis

Standard statistics for measuring genetic polymorphism and divergence were computed for individual accessions and various groups of accessions. Measures include the proportion of polymorphic loci (P), the mean number of alleles among all loci (A) and among polymorphic loci (A_P), estimated heterozygosity ($H = 1 - \sum [p_i]^2$, where p_i are allelic frequencies), and estimated heterozygosity adjusted for small samples (H_u). Multivariate relationships among all accessions were revealed through principal coordinate (PCO) analysis of average taxonomic distances based upon product

moment correlation and principal component analysis of variance-covariance matrices, each of which were derived from allele frequencies (Sneath and Sokal 1973). Genetic identities (*I*) and distances (*D*) were computed according to Nei (1978) and Rogers (1972), respectively. Distance phenograms were constructed with the Neighbor-Joining method of Saitou and Nei (1987). Computations were facilitated by PC-based programs GENESTAT-PC version 3.3 (Lewis 1993), NTSYS version 1.7 (Rohlf 1992) and MEGA version 1.0 (Kumar et al. 1993). Because of the large size of the data set, allele frequencies and summary statistics (*A*, *P*, *H*) for individual accessions are not reported herein. Tables may be obtained from the North Central Regional Plant Introduction Station homepage (http://www.ars-grin.gov/ars/MidWest/Ames/crops/sunflowr.html) or from the corresponding author.

Results

Genetic variability in *Helianthus annuus*

The 20 enzyme systems we screened revealed a minimum of 30 loci and 72 alleles (A = approx. 2.4) for the 146 accessions. Four of the loci (*Est1*, *Gdh1*, *Glu1*, and *Rbc1*) revealed single genotypes in a preliminary survey of wild and domesticated *H. annuus* and hence were not analyzed in all accessions. Additionally, six polymorphic loci (*Acp1*, *Cat1*, *G3pdh1*, *G3pdh2*, *Idh2*, and *Leu1*) were poorly resolved or stained unreliably under our assay conditions. Because these loci were uninformative with respect to characterizing diversity in *H. annuus*, they were excluded from further analyses. A consequence of excluding these loci is that *all* estimates of diversity (*P*, *A*, *H*; summarized in Table 2) are expected to be inflated relative to more inclusive values.

For the remaining 13 enzymes, an estimated 20 polymorphic loci (approx. 1.5 loci per enzyme system) and 62 alleles (approx. 3.0 alleles per locus) were resolved; hence, no locus was fixed for the same allele in the wild and domesticated accessions. A summary of the loci and alleles resolved in each group of accessions is provided in Table 3. Allelic variation among these 20 loci was not partitioned equally, inasmuch as 6 (*Adh2*, *Aat1*, *Me2*, *Mr2*, *Pgi1*, *and Tpi1*) were minimally variable with only 2 alleles per locus. An additional 9 loci were tri-allelic and the remaining 5 (*Adh1*, *Aco1*, *Aco2*, *Pgi2*, and *Skd1*) were multi-allelic, displaying between 4 and 7 alleles per locus. Average allele frequencies for *H. annuus* (Table 3) indicated that some polymorphic loci were weakly polymorphic, because 9 of 20 loci showed allele frequencies ≥ 0.9 for the most common allele. Consequently, as a species, our sample of *H. annuus* (wild + domesticated) has a moderately high estimated heterozygosity ($H = 0.27$). Averaged across loci, the mean estimated heterozygosity for individual groups of *H. annuus* accessions ranged from a high of approximately 0.21 for wild *H. annuus* to a low of 0.13 for all domesticated sunflower groups (Table 2).

As expected, wild sunflower is more polymorphic than domesticated sunflower, although individual wild accessions may possess levels of allelic diversity and heterozygosity resembling modern cultivars. This trend was evident in the summary statistics (Table 2) and allelic frequencies for groups of sunflower (Table 3). As a group, wild sunflower included 59 alleles at 20 loci, yielding an average of 1.65 alleles per locus (2.18 alleles per polymorphic locus). Eight of these alleles (*Aat3-3*, *Adh2-2*, *Idh1-7*, *Pgd1-3*, *Pgi2-1*, *Pgm1-6*, *Pgm2-6*, and *Tpi2-6*) were unique to wild sunflower (Table 3). The 11 accessions that comprise the Great Plains group showed the highest overall diversity within the wild group with 53 total alleles ($A = 1.75$, $A_P = 2.18$) and an

Table 2 Genetic diversity statistics for wild and domesticated accessions of *Helianthus annuus*

Group	N[a]	A	Ap	P	H	Hu	U
Wild-Great Plains accessions	5.2	1.75	2.18	0.61	0.24	0.27	4
Wild-California accessions	6.6	1.59	2.15	0.52	0.19	0.21	1
Wild-Southwest accessions	6.9	1.66	2.22	0.54	0.20	0.22	0
Ornamental accessions	8.7	1.74	2.23	0.60	0.20	0.22	1
Native American accessions	7.5	1.53	2.19	0.44	0.17	0.19	0
Oilseed accessions	5.9	1.34	2.02	0.33	0.12	0.13	0
Confectionery accessions	6.6	1.40	2.05	0.37	0.14	0.15	0
Elite cultivars	7.4	1.41	2.06	0.39	0.15	0.16	0
Mean, wild accessions	6.2	1.65	2.18	0.56	0.21	0.23	
Mean, domesticated accessions	6.5	1.39	2.05	0.37	0.13	0.15	
Helianthus species							
H. annuus	836.4	3.10	3.10	1.00	0.27	0.27	
H. argophyllus	14.3	1.93	2.27	0.73	0.19	0.20	
H. petiolaris	16.9	1.61	2.09	0.61	0.25	0.27	

[a] Abbreviations for gene diversity statistics include N (average number of plants sampled per accession), A (average number of alleles per locus), Ap (average number of alleles per polymorphic locus), P (proportion of loci polymorphic), H (estimated heterozygosity), Hu (estimated heterozygosity unbiased for sample size), and U (number of unique alleles per group within *H. annuus*)

Table 3 Mean allele frequencies for major groups of *Helianthus annuus* and related species examined in this study. Locus/allele nomenclature is described in the Materials and methods

	Wild-plains	Wild-California	Wild-southwest	Ornamental accessions	Native American	Oilseed accessions	Confection accessions	Elite cultivars	*H. petiolaris*	*H. argophyllus*
Aat1-2	0.907	0.976	0.990	1.000	1.000	0.841	0.925	0.916	1.000	0.700
-4	0.093	0.024	0.010	0.000	0.000	0.159	0.075	0.084	0.000	0.300
Aat2-2	0.983	1.000	0.946	0.985	0.975	1.000	0.974	1.000	1.000	1.000
-3	0.017	0.000	0.036	0.000	0.000	0.000	0.003	0.000	0.000	0.000
-4	0.000	0.000	0.018	0.015	0.025	0.000	0.023	0.000	0.000	0.000
Aat3-2	0.736	0.877	0.886	1.000	—[a]	1.000	0.962	1.000	1.000	—
-3	0.038	0.000	0.000	0.000	—	0.000	0.000	0.000	0.000	—
-4	0.226	0.123	0.114	0.000	—	0.000	0.038	0.000	0.000	—
Adh1-2	0.213	0.100	0.100	0.000	0.050	0.000	0.008	0.000	—	—
-3	0.043	0.000	0.400	0.044	0.050	0.000	0.004	0.000	—	—
-4	0.681	0.900	0.500	0.783	0.900	1.000	0.953	0.965	—	—
-6	0.064	0.000	0.000	0.174	0.000	0.000	0.035	0.034	—	—
Adh2-2	0.044	0.000	0.000	0.000	0.000	0.000	0.000	0.000	—	—
-4	0.956	1.000	1.000	1.000	1.000	1.000	1.000	1.000	—	—
Aco1-2	0.118	0.122	0.088	0.222	0.061	0.138	0.068	0.228	0.000	0.000
-3	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.000	0.000	0.000
-4	0.882	0.878	0.882	0.764	0.939	0.842	0.930	0.772	1.000	1.000
-6	0.000	0.000	0.029	0.000	0.000	0.020	0.002	0.000	0.000	0.000
Aco2-2	0.000	0.084	0.071	0.028	0.000	0.005	0.000	0.000	0.000	0.023
-4	0.410	0.740	0.893	0.806	1.000	0.919	0.850	0.901	0.875	0.977
-5	0.515	0.052	0.018	0.167	0.000	0.000	0.033	0.000	0.000	0.000
-6	0.075	0.123	0.018	0.000	0.000	0.076	0.116	0.099	0.125	0.000
Idh1-4	0.061	0.123	0.137	0.800	0.125	0.122	0.300	0.208	1.000	0.200
-6	0.918	0.877	0.863	0.200	0.875	0.878	0.700	0.792	0.000	0.700
-7	0.021	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.100
Leu2-2	0.000	0.000	0.000	0.000	0.033	0.000	0.005	0.000	0.000	0.000
-4	0.711	0.845	1.000	0.833	0.600	0.576	0.599	0.462	0.000	0.000
-6	0.289	0.155	0.000	0.167	0.367	0.424	0.396	0.538	0.000	0.000
Mdh1-1	*0.706*	*0.697*	*0.888*	*0.456*	*0.460*	*0.198*	*0.253*	*0.368*	*0.636*	*0.620*
-2	0.294	0.224	0.112	0.544	0.360	0.754	0.747	0.632	0.364	0.380
-4	0.000	0.079	0.000	0.000	0.180	0.048	0.000	0.000	0.000	0.000
Me2-2	0.400	0.475	0.679	0.265	0.425	0.295	0.228	0.180	0.786	0.750
-4	0.600	0.525	0.321	0.735	0.575	0.705	0.772	0.820	0.214	0.250
Mr2-2	0.488	0.779	0.932	0.813	0.321	0.555	0.552	0.339	0.667	1.000
-4	0.512	0.221	0.062	0.188	0.679	0.445	0.448	0.661	0.333	0.000
Pgd1-3	0.044	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-4	0.402	0.633	0.634	0.952	0.340	0.348	0.264	0.756	0.881	0.000
-6	0.554	0.367	0.366	0.048	0.660	0.652	0.736	0.244	0.119	0.000
Pgi1-1	0.000	0.000	0.000	0.048	0.000	0.035	0.125	0.000	0.000	0.000
-2	1.000	1.000	1.000	0.952	1.000	0.965	0.875	1.000	1.000	1.000
Pgi2-1	0.103	0.000	0.112	0.000	0.000	0.000	0.000	0.000	0.000	0.000
-2	0.706	0.888	0.858	0.500	0.780	0.746	0.645	0.705	0.500	0.700
-4	0.183	0.106	0.030	0.500	0.180	0.254	0.351	0.295	0.500	0.300
-6	0.008	0.006	0.000	0.000	0.040	0.000	0.004	0.000	0.000	0.000
Pgm1-2	0.221	0.138	0.081	0.319	0.000	0.097	0.171	0.159	0.022	0.022
-4	0.743	0.862	0.871	0.681	1.000	0.903	0.829	0.841	0.978	0.978
-6	0.036	0.000	0.048	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Pgm2-2	0.199	0.087	0.068	0.222	0.017	0.020	0.000	0.000	0.217	0.050
-4	0.780	0.913	0.839	0.778	0.983	0.980	1.000	1.000	0.783	0.950
-6	0.021	0.000	0.093	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Skd1-2	0.258	0.196	0.365	0.357	0.517	0.745	0.667	0.000	0.020	0.905
-3	0.113	0.082	0.083	0.057	0.207	0.029	0.048	0.814	0.000	0.024
-4	0.468	0.392	0.229	0.314	0.000	0.110	0.129	0.062	0.798	0.047
-5	0.008	0.000	0.000	0.043	0.207	0.000	0.000	0.053	0.000	0.000
-6	0.145	0.299	0.229	0.029	0.000	0.116	0.156	0.000	0.182	0.024
-7	0.008	0.010	0.000	0.000	0.000	0.000	0.000	0.071	0.000	0.000
-8	0.000	0.021	0.094	0.200	0.000	0.000	0.000	0.000	0.000	0.000
Tpi1-2	0.196	0.029	0.045	0.045	0.000	0.000	0.000	0.000	0.045	0.088
-4	0.804	0.971	0.955	0.955	1.000	1.000	1.000	1.000	0.955	0.912
Tpi2-2	0.182	0.049	0.030	0.166	0.000	0.000	0.000	0.000	0.091	0.000
-4	0.818	0.930	0.970	0.834	1.000	1.000	1.000	1.000	0.909	0.917
-6	0.000	0.021	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.083

[a] Not determined due to poor separation or staining

average estimated heterozygosity of 0.24. The wild Southwestern and Californian groups had a lower (yet nearly equivalent) degree of polymorphism with 49 alleles/20 loci ($A = 1.66$, $H = 0.20$) and 45 alleles/19 loci ($A = 1.59$, $H = 0.19$), respectively. Among groups of wild sunflowers, Great Plains accessions showed 4 unique alleles (*Aat3-3*, *Adh2-2*, *Idh*1-7, *Pgd*1-3), while the California group showed 1 (*Tpi2-6*).

Our sample of approximately 700 plants from domesticated sunflower revealed a total of 53 alleles, with an average of 1.39 alleles per locus and 2.05 alleles per polymorphic locus (Table 2). Three alleles from this sample (*Aco1-3*, *Leu2-2*, *Pgi1-1*) were unique to domesticated sunflower; of these, only 1 was unique to a single domesticated group (*Aco1-3* from ornamentals). Because the domesticated sunflower has been bred for a variety of uses (e.g., ornament, seed oil, edible seeds), we divided the total sample of domesticated material into five subgroups according to their use or improvement status: an ornamental group ($n = 3$), a Native American group ($n = 3$), an oilseed group ($n = 50$), a confectionery group ($n = 45$), and an elite cultivar group ($n = 13$). Among these five groups, the three ornamental accessions included a high amount of allelic variation and heterozygosity, with average values of 1.74 for A (a value that *exceeds* most wild groups) and 0.20 for H. The remaining domesticated accessions included considerably less polymorphism, with values ranging from 1.53 for A and 0.17 for H in the Native American cultivars, and even less polymorphism in the oilseed cultivars ($A = 1.34$, $H = 0.12$). The sample of 3 Native American cultivars included intermediate levels of allelic diversity and heterozygosity relative to wild accessions and modern cultivars, with diversity estimates nearly identical to those reported by earlier authors (e.g., mean A is 1.53 in this study, 1.36 in Rieseberg and Seiler 1990).

Table 4 Average gene diversity statistics across all loci for all groups of wild and domesticated *H. annuus*. Means were calculated from 30 randomly selected accessions for the wild, oilseed, and confectionery groups, and from all accessions for the ornamental, Native American, and elite groups

Group	H_S	H_T	G_{ST}
Wild	0.209	0.310	0.327
Ornamental	0.202	0.278	0.274
Native American	0.175	0.213	0.181
Oilseed	0.105	0.217	0.514
Confectionery	0.110	0.196	0.438
Elite	0.146	0.206	0.289

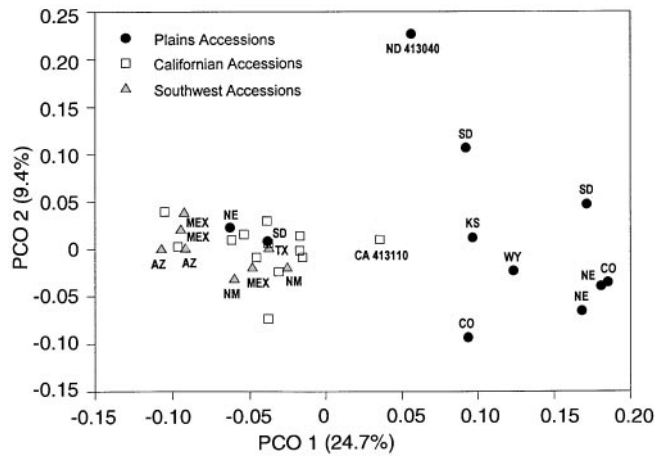

Fig. 1 Principal coordinate analysis of wild *Helianthus annuus* accessions from three geographic regions based on a taxonomic distance matrix of allele frequencies at 20 polymorphic loci. The first two axes account for 24.7% and 9.4% of the total variance, respectively

Interpopulational relationships in wild *Helianthus annuus*

Mean allele frequencies for wild *H. annuus* across 20 polymorphic loci are presented in Table 3 for each geographic region. Inspection of this table reveals that allele frequencies at several loci differ among regions. For example, Californian accessions are nearly fixed for *Adh1-4* (frequency = 0.90), whereas sunflower from the Great Plains (frequency = 0.68) and Southwest (frequency = 0.50) are more diverse at this locus. The effects of regional divergence in allele frequencies on apportionment of genetic diversity were quantified by the gene diversity statistics of Nei (1978) shown in Table 4. Averaged across 20 polymorphic loci, the proportion of total variation associated with interpopulational differentiation was relatively high ($G_{ST} = 0.33$), indicating that a substantial proportion of genetic diversity lies among (rather than within) individual accessions.

To test for an association of allelic variation in wild *H. annuus* with geographic origin, we reduced the dimensionality of the data set by using principal coordinate analysis on a matrix of taxonomic distances derived from allele frequencies (Rohlf 1992). Accessions were plotted by their coordinates along the first two axes (which accounted for 34.1% of the total variance). Two contiguous clusters of accessions were projected along the first principal coordinate axis. This axis (PCO 1, Fig. 1) separates nearly all accessions from the Great Plains (CO, KS, ND, NE, SD, WY) from western United States accessions, although there are three exceptions to this pattern. First, the Great Plains accessions Neb. 413016 and S.D. 413015 lie within the cluster defined by Californian/Southwestern accessions. Although these accessions resembled western sunflowers in the PCO analysis, they did not contain alleles diagnostic for Californian/Southwestern

sunflowers but did show alleles diagnostic for Great Plains accessions (e.g., *Adh2-2* in 413016 and *Pgd1-3* in 413015; data not shown). The remaining two exceptions include Great Plains accession N.D. 413040 and California accession 413110, each of which failed to cluster with their respective regional groups. These 2 accessions lacked diagnostic regional alleles, and occupy unique positions on this PCO plot because of their covariance patterns for common alleles.

For further analyses, we divided wild accessions into three geographic groups. The Great Plains group includes all accessions of the Great Plains cluster and the 3 outliers from Nebraska, North Dakota and South Dakota ($n = 11$ accessions total). The remaining two groups compose a cluster of accessions from the western United States, which were subdivided into two overlapping groups: Californian accessions ($n = 13$) and Southwestern accessions ($n = 8$). Interrelationships between these accessions were also investigated with principal component analysis of variance/covariance matrices derived from allele frequencies. Results from these analyses were nearly identical to those obtained by PCO, with principal components 1 and 2 accounting for 32.8% of the variance (not shown). We chose to illustrate multivariate relationships with PCO because allele information was missing from some accessions, and PCO has been reported to be less sensitive to missing values than is principal component analysis (Rohlf 1992).

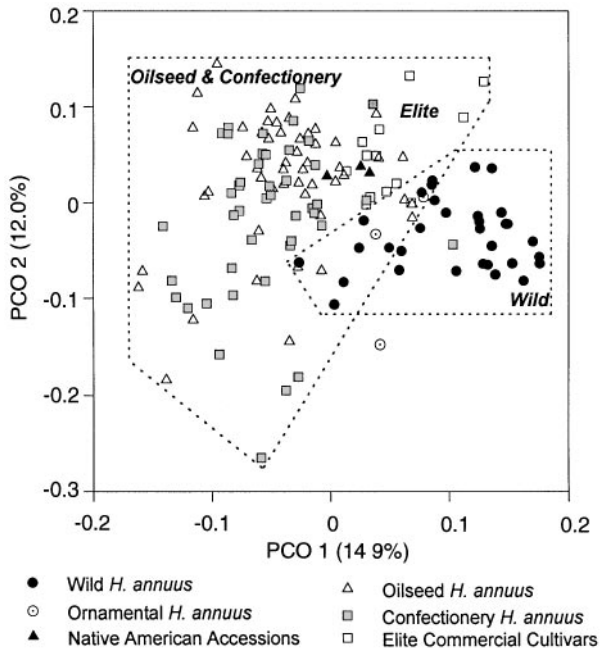

Fig. 2 Principal coordinate analysis of wild and domesticated *Helianthus annuus* accessions based on a taxonomic distance matrix of allele frequencies at 20 polymorphic loci. The first two axes account for 14.9% and 12.0% of the total variance, respectively

Interrelationships between domesticated and wild *Helianthus annuus*

To investigate relationships between domesticated and wild accessions of *H. annuus*, PCO analyses were performed on a matrix of taxonomic distances derived from allele frequencies of all 146 accessions. Accessions were plotted by their first two principal coordinates, which accounted for 26.9% of the total variance in the distance matrix (Fig. 2). This analysis shows that domesticated sunflower is a heterogeneous and diverse assemblage, and although its genetic profile generally does not overlap with wild accessions, its constituent accessions cannot be easily categorized isozymatically according to their agronomic use. Of the five end use/improvement groups, only two – Native American accessions and elite hybrid cultivars – form relatively cohesive clusters. Although we have subdivided the remaining domesticated sunflowers into groups of ornamental, oilseed and confectionery types, these groups lack genetic homogeneity and intergrade with each other. This result is in agreement with measures of genetic partitioning (Table 4), because the high G_{ST} values for oilseed and confectionery groups (0.51 and 0.44, respectively) indicate that much of the observed heterozygosity within these groups is partitioned among (rather than within) individual

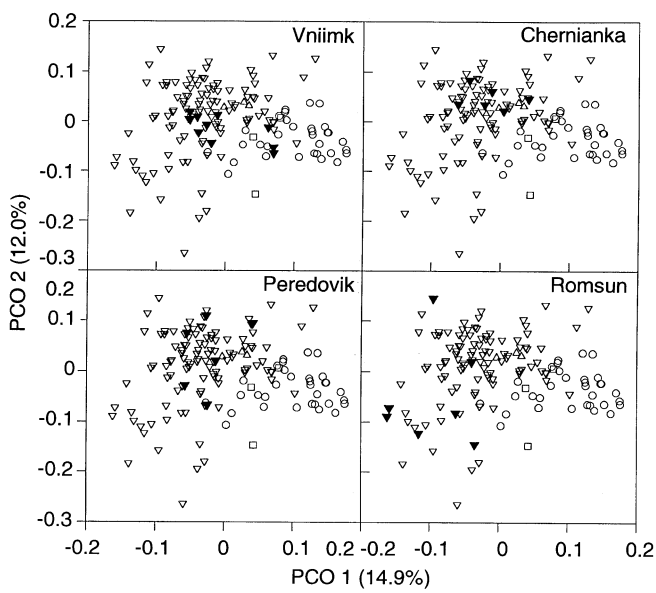

Fig. 3 Genetic variation among domesticated accessions with similar or identical common names, as shown by principal coordinate analysis. Accessions of Vniimk ($n = 10$), Chernianka ($n = 6$), Peredovik ($n = 6$) and Romsun ($n = 7$) are highlighted (*dark inverted triangles*) in the PCO graph shown in Fig. 2.

accessions. The genetic heterogeneity of oilseed sunflowers extends to individual accessions that share a similar provenance or name. By highlighting 10 "Vniimk" (Fig. 3), 6 "Chernianka", 6 "Peredovik", and

7 "Romsun" accessions, it is clear that named cultivars sharing a similar provenance or name may be genetically quite divergent.

While the interrelationships among wild and domesticated accessions are complex, two general trends are evident in Fig. 2. First, wild accessions are generally distinct from most cultivated accessions, although alleles of the latter largely represent a subset of those detected in the former. Second, the positive PCO1 and PCO2 values for elite hybrid accessions place them at coordinates which are unique relative to other domesticated accessions. As was observed with Fig. 1, principal component analyses (data not shown) of the combined "wild + domesticated" data set were concordant with the results derived by PCO, and the first two principal component axes accounted for 25.0% of the total variance.

The systematic relationships among the major groups of wild and domesticated sunflowers and two closely related outgroup species (*Helianthus argophyllus* and *H. petiolaris*) were evaluated by calculating the genetic identity (*I*) coefficient of Nei (1978; Table 5). Within *H. annuus*, the highest genetic identities were observed in pairwise comparisons between confectionery and oilseed groups ($I = 0.99$), and the lowest were between elite cultivars and wild sunflowers from the Southwest ($I = 0.88$). As was reflected by PCO analysis, wild sunflower accessions generally are very similar to other wild sunflowers ($I = 0.94$ to 0.98), although wild sunflowers from the Great Plains are slightly more similar to Native American cultivars than they are to wild sunflowers from the Southwest.

The 3 accessions of ornamental sunflowers were moderately divergent from all other groups of *H. annuus* ($I = 0.89$ to 0.93), and they actually resembled *H. petiolaris* more closely ($I = 0.95$). Although the history of these plants is not well-documented, they originated from gardens in Europe and may have been selected for inclusion into the NPGS because they demonstrated pollen sterility (H. Shands, personal communication). It is possible that the close affinity of the ornamentals to *H. petiolaris* reflects a history of interspecific introgression. Nevertheless, the evidence for introgression remains equivocal since the similarity between these ornamentals and *H. petiolaris* results from similar frequencies of common, rather than unique, alleles.

To evaluate systematic relationships among groups of accessions, we constructed Neighbor-Joining phenograms (Saitou and Nei 1987) derived from Rogers' genetic distances (Rogers 1972), and rooted the trees using either *H. argophyllus* (Fig. 4A) or *H. petiolaris* (Fig. 4B). Although the topologies of the trees change somewhat depending on the outgroup used (particularly with regard to placement of ornamental accessions), domesticated sunflowers consistently represent a discrete lineage which is sister to the wild Great Plains accessions, and the remaining groups are placed more distant to the "domesticated + Great Plains"

Table 5 Nei's genetic identity (*I*; above diagonal) and Rogers's genetic distances (*D*: below diagonal) for major groups of *Helianthus annuus* and related wild species

	Wild plains	Wild California	Wild-southwest	Ornamental	Native	Oilseed	Confectionery	Elite	*H. petiolaris*	*H. argophyllus*
Wild-Great plains	—	0.972	0.942	0.905	0.950	0.937	0.938	0.919	0.865	0.905
Wild-California	0.125	—	0.981	0.935	0.958	0.950	0.946	0.933	0.908	0.949
Wild-Southwest	0.175	0.090	—	0.913	0.930	0.913	0.907	0.883	0.900	0.969
Ornamental	0.199	0.163	0.201	—	0.893	0.912	0.925	0.915	0.952	0.911
Native American	0.156	0.129	0.172	0.224	—	0.981	0.976	0.954	0.849	0.939
Oilseed	0.170	0.142	0.190	0.196	0.089	—	0.995	0.945	0.845	0.951
Confectionery	0.178	0.156	0.204	0.180	0.108	0.054	—	0.942	0.862	0.941
Elite	0.200	0.161	0.218	0.186	0.130	0.112	0.116	—	0.858	0.865
H. petiolaris	0.238	0.181	0.194	0.141	0.243	0.246	0.238	0.232	—	0.883
H. argophyllus	0.223	0.162	0.136	0.231	0.160	0.154	0.175	0.222	0.192	—

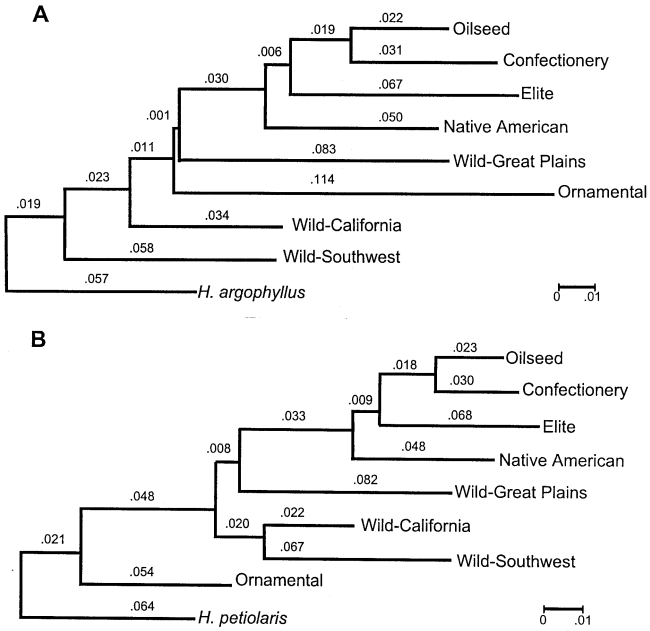

Fig. 4A, B Neighbor-joining cluster analysis of systematic relationships among eight major groups of wild and domesticated *Helianthus annuus* based on Rogers's genetic distance derived from allele frequencies at 20 polymorphic allozyme loci. Trees are rooted with either *H. argophyllus* (**A**) or *H. petiolaris* (**B**)

lineage. In both trees, oilseed and confectionery accessions clustered most closely with elite cultivars. Native American cultivars were the most divergent form of domesticated sunflower and (concordant with gene diversity statistics) occupy an intermediate position between modern cultivars and wild sunflowers from the Great Plains. By incorporating two different outgroup species, the unique nature of the 3 ornamental accessions is clearly shown. Rooting with *H. argophyllus* places the ornamentals basal to the cultivated sunflowers and adjacent to wild sunflowers from the Great Plains (necessitating a long terminal branch), but rooting with *H. petiolaris* places ornamentals basal to all *H. annuus*. Changes in tree topology relative to different outgroups are commonly seen with analyses of hybrids (McDade 1992) that incorporate characteristics of each parent. Our results therefore suggest that these ornamentals may have been introgressed with *H. petiolaris* or an allied wild species.

Discussion

Genetic variation in wild and domesticated *Helianthus annuus*

During the initial phases of domestication, it is thought that only a fraction of the total genetic variation present in an ancestral taxon will be incorporated into a newly evolved domesticate (Doebley 1989). As new domesticates are propagated, additional genetic diversity may be lost through selection and genetic drift. Unless this process is ameliorated by gene flow, perhaps from related cultivars or wild progenitors, loss of genetic variation may be severe. Hence, although genetic variation is expected to be lost during the domestication process, the degree and severity ultimately depend on factors such as the intensity of selection, the prevalence of drift, and the frequency with which new variation is introduced.

The historical development of modern commercial sunflowers has been well-documented (Pustovoit 1976; Fick 1978; Heiser 1976), and it is known that: (1) sunflower was originally domesticated in North America, (2) sunflower was imported to Europe where it was bred for a variety of uses, and (3) European cultivars were imported back to North America during the latter half of the 19th century. Each of these events may have imposed a genetic "bottleneck" on domesticated sunflowers so that the amount of genetic diversity in domesticates might be expected to be strongly associated with the degree of agronomic selection.

Estimates of genetic variability from this study indicate that wild *H. annuus* is more diverse genetically than are most forms of domesticated *H. annuus*, the latter showing reductions in genetic diversity with increasing agronomic selection (Tables 2, 3). When wild sunflowers are compared with domesticated sunflowers (omitting the putatively introgressed ornamentals), the reduction in genetic diversity is manifested in three ways: as a reduction in allelic diversity (A and A_P), as a reduction in the proportion of polymorphic loci (P), and as a reduction in estimated heterozygosity (H). For example, our sample of 3 Native American cultivars, each of which originated in different regions of the United States ('Arikara' from North Dakota, 'Havasupai' from Arizona, and 'Seneca' from New York), collectively show mean reductions of 7% for A (1.53 vs. 1.65), 21% for P (0.44 vs. 0.56), and 19% for H (0.17 vs. 0.21) relative to wild sunflowers. Although our sample is limited to only 3 Native American landraces, the reduced genetic diversity of these accessions likely reflects of the magnitude of the first genetic bottleneck associated with domestication.

The decrease in genetic diversity associated with the second bottleneck (further agronomic selection in Europe) is best exemplified by the oilseed and confectionery accessions, each of which are less polymorphic than are wild sunflowers or Native American cultivars. The oilseed and confectionery accessions show reductions of 17% for A (1.37 vs. 1.65), 37% for P (0.35 vs. 0.56), and 38% for H (0.13 vs. 0.21) relative to wild sunflowers. Interestingly, our data provide little evidence for a third bottleneck during the breeding of elite oilseed cultivars because these accessions include equivalent levels of allelic diversity (A) and *greater* heterozygosity (P, H) than do either the oilseed or confectionery accessions. The genetic diversity in the

elite accessions may be due either to introgression of genes from wild *H. annuus* and/or other wild species (Pustovoit 1976; Fick 1978) or to the increasing use of fertility/restorer lines in making hybrid sunflowers. This latter factor may be more important because the elite domesticates show an increase in P and H (e.g., increased heterozygosity) without a measurable gain in the mean number of alleles per locus.

In addition to reducing allelic diversity and heterozygosity, the process of domestication has also changed the apportionment of genetic variation in domesticates as compared to their wild relatives. Genetic variation is increasingly apportioned among accessions subjected to more intense agronomic selection; wild sunflowers have G_{ST} values of 0.33, whereas oilseed and confectionery sunflowers have values of 0.51 and 0.44, respectively, indicating that genetic differentiation among domesticated accessions is stronger than that observed among wild accessions. Similarly, different oilseed accessions with the same cultivar name (and, presumably, similar pedigrees) may be highly divergent (Fig. 3). Although the increase in G_{ST} and in genetic heterogeneity seems related, they may be due to different phenomena, such as differential survivorship of desirable genotypes or lineages (which would increase G_{ST}) and/or introgression of new genotypes during breeding (which would increase inter-accession divergence). These results indicate that allozyme markers have sufficient power to resolve relationships among cultivars, even those that may have a similar domestication history. These results also have important implications for plant genetic resource management (Bretting and Widrlechner 1995), because it is clear that accessions with similar or identical common names may not be genetically redundant.

Although the "average" diversity estimates reported herein illustrate the prevailing trend of a reduction in genetic diversity during domestication, exceptions exist and averages alone do not fully describe genetic diversity in these sunflowers. Wild sunflowers from California, in particular, include little allelic diversity and low heterozygosity as a group ($A = 1.6$, $H = 0.19$), with individual accessions commonly showing A less than or equal to 1.5 and H below 0.11 (data not shown). These values are equivalent to those of many of the domesticated accessions examined in this study; 10 of 50 oilseed (20%), 14 of 45 confectionery (31%), and 3 of 13 elite (23%) accessions have values of A equal to or greater than 1.5. Hence, the use of average values alone masks the fact that wild sunflowers may also exist in populations that are as genetically homogeneous as are those of domesticated sunflowers. Because of the weedy nature of wild *H. annuus* and its ability to colonize disturbed soils, the genetic homogeneity in wild populations may reflect founder events or recent ecological or geographic expansion. In light of this information, it is interesting to note that Heiser (1949) considered *H. annuus* to be a relatively recent introduction to California, based upon historical accounts and collection information.

Variability of wild *Helianthus annuus* across its geographic range

Sunflower is highly variable genetically across its native range (Heiser 1951, 1954; Rogers et al. 1982) and in Europe (Putt 1978). Although morphological variation intergrades extensively across geographic regions, certain morphological variants are sufficiently discrete that formal taxa have been proposed for them (Heiser 1954; Heiser et al. 1969). Wild sunflowers include several geographic variants: *H. annuus* ssp. *jaegeri* from the southwest; the widely distributed western variant, *H. a.* ssp. *lenticularis*; and *H. a.* ssp. *annuus*, the weed sunflower of the central and eastern United States. These sunflowers typically have small capitulae (2–5 cm diameter), narrow phyllaries (3–5 mm wide) that range from glabrous to hirsute, and have highly branched stems. These characteristics clearly separate the "wild and weedy" sunflower from "domesticated" sunflower, which displays larger capitulae (≥ 5 cm diameter), wide phyllaries (> 6 mm wide), unbranched (or minimally branched) stems, and occassional unique floral characteristics such as "double" (chrysanthemum-like) capitulae or light-yellow to red ray florets. Despite the evident morphological divergence among ecogeographical variants of wild sunflower, prior investigations of this species have uncovered little geographically structured genetic variation. Rieseberg and Seiler (1990) analyzed 12 populations of wild *H. annuus* with isozyme markers and found that wild sunflowers with similar geographic provenances were not necessarily genetically allied. More recently, Arias and Rieseberg (1995) examined variation within and between 10 wild sunflower populations and 20 accessions of domesticated sunflower with RAPD markers. Despite the number of polymorphic RAPD bands detected ($n = 68$), relationships among these groups were not elucidated, and genetic polymorphism among wild sunflower populations was not tangibly associated with geographic provenance. In both of these cases, the lack of association between genetic variation and geographic origin was attributed to repeated episodes of human-mediated long-distance dispersal.

In contrast to previous studies, we found evidence that genetic variation in wild sunflower from different regions of the United States is geographically structured. Specific evidence includes "region-specific" alleles (e.g., *Adh2-2* for Plains accessions, *Tpi2-6* for California accessions; Table 3), moderate geographic partitioning of genetic variation as measured by G_{ST} (Table 4), and patterns of covariance of allele frequencies across loci (Fig. 1). Interestingly, the relatively high degree of taxonomic resolution found in this study is not the result of greater infra-populational

sampling because we examined considerably fewer achenes per accession (between 5 and 20) than were used in previous studies (e.g., 30 achenes per accession in Rieseberg and Seiler 1990 and Arias and Rieseberg 1995). Furthermore, the different conclusions drawn by these studies do not seem to be the result of different electrophoretic or staining techniques. For example, both *Gal* and *Gdh* were invariant in the 12 populations of wild *H. annuus* surveyed by Rieseberg and Seiler (1990), a result that agrees with our findings. Rieseberg and Seiler (1990) found the same number of alleles at most loci (e.g., *Aat3*, *Pgi2*, *Tpi1*, and *Tpi2*) that were found in the present study, and all remaining loci (except *Skd1*) differed at most by 1 allele. These observations lead us to believe that sunflowers from different geographic regions *differ primarily with respect to the frequency of common alleles* rather than by the occurrence of unique and rare alleles that would be detected by intensive sampling. Consequently, although the strategy employed in this study – sampling a few achenes from many accessions – may not be ideal for characterizing the *total* allelic diversity in sunflower, it seems to be adequate for discriminating differences in frequencies of common alleles that define groups of wild sunflower.

The data presented in this paper provide initial genetic evidence for the taxonomic subdivision of wild *H. annuus* along geographic lines, as Californian/Southwest and Great Plains sunflowers form two largely discrete clusters in PCO (Fig. 1) and principal component analyses (data not shown). The traditional subspecific categories of *H. annuus* proposed by Heiser (1954) have not been universally applicable to all wild sunflower populations. Across their range, populations of wild sunflower intergrade morphologically, which can confound their subspecific assignment. In addition, wild sunflower readily hybridizes with domesticated *H. annuus* and other closely related species, giving rise to offspring with intermediate and segregating characteristics. Despite these phenomena, the present study indicates that there may be a natural ecogeographic division in *H. annuus* which coincides with the separation of Heiser's two western subspecies *lenticularis* + *jaegeri* (represented by Californian and Southwestern accessions) from subspecies *annuus* (Great Plains accessions). Although the accessions we studied isozymatically have not been analyzed for the morphological characteristics used in subspecific classification (e.g., ray floret number and length, stamen color, achene length, capitulum diameter; Heiser 1954), the geographic clustering we observed provides some support for the subspecific categories proposed by Heiser.

Insights into the domestication of sunflower

As stated in the Introduction, several hypotheses have been advanced concerning the domestication of annual sunflower. In the first hypothesis, Heiser (1951, 1976, 1978) suggested that wild annual sunflower was originally restricted to the Southwest. From this location, a weedy, "camp following" form developed and spread eastward, with some populations subsequently cultivated and domesticated in the central United States. Through human migration and trade, this domesticated sunflower would subsequently have spread back to the Southwest, in addition to being introduced to the eastern United States. The second hypothesis, by Lathrap (cited in Heiser 1978), suggests that sunflower was both cultivated and domesticated in the southwestern United States, and then the domesticates were disseminated eastward through trading among Native American populations. While these hypotheses differ with respect to the location of the initial domestication event, they do not differ significantly with respect to the ancestral wild germplasm that was domesticated, which would have been "western" in nature, resembling present-day *H. a.* ssp. *jaegeri* or ssp. *lenticularis*.

In contrast to previous studies (Arias and Rieseberg 1995; Rieseberg and Seiler 1990), our analysis of 32 wild populations across the natural range of *H. annuus* indicates that wild sunflower exhibits geographically partitioned genetic variability (Fig. 1), an important clue for identifying the center of origin for domesticated sunflower. Principal component and Neighbor-Joining analyses separate wild sunflower into two groups, one composed primarily of accessions from the Great Plains and a second composed of accessions from the western United States and Mexico. When allelic frequencies are computed for these groups and are compared with allelic frequencies of various domesticated groups (Fig. 4), wild sunflower from the Great Plains is allied most closely to domesticated sunflower, with sunflower from California and the Southwest more distantly related. In addition, the Great Plains contains the highest genetic diversity for annual sunflower, exceeding all others in allelic richness, percent polymorphic loci, and mean panmictic heterozygosity (Table 2). Moreover, Great Plains accessions include the most unique alleles (Table 2) and the highest frequencies for alleles that are rare among all wild accessions (Table 3). These results provide evidence that wild *H. annuus* from the Great Plains is most similar to domesticated sunflower, which leads us to suggest that germplasm from this region has been integral to sunflower domestication and that sunflower may have been originally domesticated in the Midwest. An important assumption underlying this inference is that we have sampled sufficiently to detect the allelic variation that exists in the various sunflower gene pools.

The proposal that domesticated sunflower has a midwestern origin appears inconsistent with the earlier hypotheses of Heiser and Lathrap, since their ancestral domesticate would have contained southwestern germplasm. However, given the potential for gene flow and

human-mediated migration, the two scenarios – domestication of southwestern germplasm followed by introduction to the Great Plains versus domestication of Great Plains germplasm – might be difficult to distinguish if early domesticates were repeatedly introgressed with wild material from the Great Plains. In this context, it is important to note that sunflower domestication in the southwest is contraindicated by archaeological evidence, since achenes from archaeological sites in the western United States are consistently small (< 8 mm long) and resemble wild *H. annuus* as compared to the "improved" achenes (> 8 mm long) found in archaeological sites in the central United States dating to 4,000–3,000 BP (Heiser 1978; Smith 1992).

Additional hypotheses have been proposed that implicate species other than *H. annuus* in sunflower domestication. In particular, Edgar Anderson (cf. Heiser 1978) suggested that the originally domesticated annual sunflower may have been introgressed with *H. petiolaris*. Our data show that the domesticates grown by the Arikara, Havasupai and Seneca Native Americans are genetically dissimilar to *H. petiolaris*. In addition, interspecific hybridization with *H. petiolaris* is commonly thought to have occurred with modern domesticates because *H. petiolaris* is one of several potential sources of the cytoplasmic male sterility (CMS) used in the development of elite cultivar hybrids (Dominguez-Gimenez and Fick 1975; but also see Rieseberg et al. 1994). Despite the anecdotal evidence for introgression from *H. petiolaris*, the current study revealed no evidence for *H. petiolaris* genes in any of the oilseed, confectionery, or elite sunflowers. Given that as few as five backcrosses can regenerate 98.5% of the original nuclear complement in CMS lines of *H. annuus* (Fick 1978), a survey of 20 loci is probably insufficient to reveal any introgression. This suggests that the precise genetic contribution of other wild species to the genome of the ancestral domesticate will probably remain a mystery unless a large number of nuclear markers such as restriction fragment length polymorphisms (RFLPs) (Berry et al. 1995) or RAPDs (Rieseberg et al. 1993) are assayed from the genome of *H. annuus* and close relatives such as *H. petiolaris*.

As highlighted by Stebbins (1947), human activity has played an important role in the evolution of many plants, including sunflower, whether they have been domesticated or appear truly "wild". The relatively recent (ca. 400 years), large-scale human disturbance across the natural range of *H. annuus* has likely changed the evolutionary dynamics of this species, perhaps most dramatically by increasing and expanding the range of available habitats and by aiding long-distance dispersal via intentional (e.g., cultivation) and/or accidental (e.g., bird feeders, rail, truck, barge) transport of novel genotypes into new regions. Given the number of factors that could alter the original ecogeographical distribution of annual sunflower genotypes, it seems remarkable that patterns of variation associated with ecogeographical distribution (such as those shown in this paper) can still be discerned in natural populations. Heiser (1954) suggested that these faint but observable patterns of variation in wild *H. annuus* most likely are the result of selective advantages that are present in locally adapted plants but absent in newly introduced immigrants. The extent to which locally advantageous, co-adapted gene complexes maintain the original pattern of *H. annuus* genetic diversity is unclear and beyond the scope of this paper, but it remains an intriguing question for future examination.

Acknowledgements The authors thank M. Kim-Joo for assistance during early phases of this project, J.S.C. Smith, Gerald Seiler, Jerry Miller, and Henry Shands for helpful discussion concerning accession provenances, and Curt Brubaker, Loren Rieseberg, and Randall Small for reviewing drafts of the manuscript. This work was supported by Pioneer Hi-Bred International of Johnston, Iowa, the Iowa State University Interdepartmental Genetics Program, and the USDA/ARS. This is Journal Paper No. J-191960 of the Iowa Agriculture and Home Economics Experiment Station, Ames, Iowa. Project No. 1018, and supported by Hatch Act and State of Iowa funds.

References

Anonymous (1996) 1995 U.S. sunflower crop quality report. National Sunflower Association, Bismarck, N.D.

Arias DM, Rieseberg LH (1995) Genetic relationships among domesticated and wild sunflowers (*Helianthus annuus*, Asteraceae). Econ Bot 49:239–248

Berry ST, Leon JA, Hanfrey CC, Challis P, Burkholz A, Barnes SR, Rufener GK, Lee M, Caligari PDS (1995) Molecular marker analysis of *Helianthus annuus* L. 2. Construction of an RFLP linkage map for cultivated sunflower. Theor Appl Genet 91: 195–199

Bretting PK, Widrlechner MP (1995) Genetic markers and plant genetic resource management. Plant Breed Rev 13:12–86

Crites GD (1993) Domesticated sunflower in fifth millennium b.p. temporal context: new evidence from middle Tennessee. Am Antiq 58:146–148

Doebley J (1989) Isozymic evidence and the evolution of crop plants. In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides Press, Portland, Ore., pp 165–191

Dominguez-Gimenez J, Fick GN (1975) Fertility restoration of male-sterile cytoplasm in wild sunflowers. Crop Sci 15:724–726

Fick GN (1978) Breeding and genetics. In: Carter JF (ed) Sunflower science and technology. American Society of Agronomy, Madison, Wis., pp 279–338

Hamrick JL (1989) Isozymes and the analysis of genetic structure in plant populations. In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides Press, Portland, Ore., pp 87–105

Heiser CB (1949) Study in the evolution of the sunflower species *Helianthus annuus* and *H. bolanderi*. Univ Calif Publ Bot 23:157–208

Heiser CB (1951) The sunflower among the North American Indians. Proc Am Philos Soc 95:432–448

Heiser CB (1954) Variation and subspeciation in the common sunflower, *Helianthus annuus*. Am Midl Nat 51:287–305

Heiser C, Smith DM, Clevenger S, Martin WC (1969) The North American Sunflowers (*Helianthus*). Mem Torrey Bot Club 22:1–218

Heiser CB Jr (1976) The sunflower. University Oklahoma Press, Norman, Okla

Heiser CB Jr (1978) Taxonomy of *Helianthus* and origin of domesticated sunflower. In: Carter JF (ed) Sunflower science and technology. American Society of Agronomy, Madison, WI, pp 31–53

Kahler AL, Lay CL (1985) Genetics of electrophoretic variants in the annual sunflower. J Hered 76:335–340

Kumar S, Tamura K, Nei M (1993) MEGA, molecular evolutionary genetics analysis, version 1.0. Pennsylvania State University, University Park, Pa.

Lewis P (1993) GENESTAT-PC, version 3.3. North Carolina State University, Raleigh, N.C.

McDade L (1992) Hybrids and phylogenetic systematics. II. The impact of hybrids on cladistic analysis. Evolution 46:1329–1346

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89:583–590

Pustovoit VS (1976) Selected seed culture and some agrotechnical problems of sunflower (English translation). USDA publ TT70-57248, Washington, D.C.

Putt ED (1978) History and present world status. In: Carter JF (ed) Sunflower science and technology. American Society of Agronomy, Madison, Wis., pp 1–29

Rieseberg LH, Seiler GJ (1990) Molecular evidence and the origin and development of the domesticated sunflower (*Helianthus annuus*, Asteraceae). Econ Bot [suppl] 44:79–91

Rieseberg LH, Soltis DE (1989) Assessing the utility of isozyme number for determining ploidal level: evidence from *Helianthus* and *Heliomeris* (Asteraceae). Aliso 12:277–286

Rieseberg LH, Beckstrom-Sternberg SM, Liston A, Arias DM (1991) Phylogenetic and systematic inferences from chloroplast DNA and isozyme variation in *Helianthus* sect. *Helianthus* (Asteraceae). Syst Bot 16:50–76

Rieseberg LH, Choi H, Chan R, Spore C (1993) Genomic map of a diploid hybrid species. Heredity 70:285–293

Rieseberg LH, VanFossen C, Arias D, Carter RL (1994) Cytoplasmic male sterility in sunflower: origin, inheritance and frequency in natural populations. J Hered 85:233–238

Rogers JS (1972) Measures of genetic similarity and genetic distance. Studies in Genetics VII. University of Texas Publ 7213, Austin, Tex., pp 145–153

Rogers CE, Thompson TE, Seiler GJ (1982) Sunflower species of the United States. The National Sunflower Association, Fargo, N.D.

Rohlf FJ (1992) NTSYS-PC: numerical taxonomy and multivariate analysis system, version 1.7. Applied Biostatistics, Setauket, N.Y.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Siculella L, Palmer JD (1988) Physical and gene organization of mitochondrial DNA in fertile and male sterile sunflower. CMS-associated alterations in structure and transcription of the *Atp*A gene. Nucleic Acids Res 16:3787–3799

Smith BD (1992) Rivers of change: essays on early agriculture in eastern North America. Smithsonian Institute Press, Washington, D.C.

Sneath PHA, Sokal RR (1973) Numerical taxonomy. Freeman, San Francisco

Stebbins GL (1947) Evidence on rates of evolution from the distribution of existing and fossil plant species. Ecol Monogr 17:149–158

Torres AM (1975) Genetics of sunflower alcohol dehydrogenase: *Adh-2*, non-linkage to *Adh-1*, and *Adh-1* early alleles. Biochem Genet 12:385–392

Torres AM (1983) Sunflowers. In: Tanksley SD, Orton TJ (eds) Isozymes in plant genetics and breeding. Elsevier, Amsterdam, pp 329–338

Weeden NF, Wendel JF (1989) Genetics of plant isozymes. In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides Press, Portland, Ore., pp 46–72

Wendel JF, Weeden NF (1989) Visualization and interpretation of plant isozymes.In: Soltis DE, Soltis PS (eds) Isozymes in plant biology. Dioscorides Press, Portland, Ore., pp 5–45