Theor Appl Genet (2003) 106:990–1003
DOI 10.1007/s00122-002-1127-6

Shunxue Tang · Steven J. Knapp

Microsatellites uncover extraordinary diversity in native American land races and wild populations of cultivated sunflower

Received: 3 June 2002 / Accepted: 12 August 2002 / Published online: 11 October 2002

Abstract The contemporary oilseed sunflower (*Helianthus annuus* L.) gene pool is a product of multiple breeding and domestication bottlenecks. Despite substantial phenotypic diversity, modest differences in molecular genetic diversity have been uncovered in anciently and recently domesticated sunflowers. The paucity of molecular marker polymorphisms in early analyses led to the hypothesis of a single domestication origin. Phylogenetic analyses were performed on 47 domesticated and wild germplasm accessions using 122 microsatellite loci distributed throughout the sunflower genome. Extraordinary allelic diversity was found in the Native American land races and wild populations, and progressively less allelic diversity was found in germplasm produced by successive cycles of domestication and breeding. Of 1,341 microsatellite alleles, 489 were unique to land races, exotic domesticates and wild populations, whereas only 15 were unique to elite inbred lines. The number of taxon-specific alleles was 35-fold greater among wild populations (26.27) than elite inbred lines (0.75). Microsatellite genotyping uncovered the possibility of multiple domestication origins. Land races domesticated by Native Americans of the southwestern US (Hopi and Havasupai) formed a clade independent of land races domesticated by Native Americans of the Great Plains and eastern US (Arikara and Seneca). Predictably, domestication and breeding have ratcheted genetic diversity down in sunflower. The contemporary oilseed sunflower gene pool, while not imperiled, could profit from an infusion of novel alleles from the reservoir of latent genetic diversity present in wild populations and Native American land races.

Communicated by H.H. Geiger

S. Tang · S.J. Knapp (✉)
Department of Crop and Soil Science, Oregon State University, Corvallis, OR 97331-3002, USA
e-mail: steven.j.knapp@orst.edu
Tel.: (541)-737-5842, Fax: (541)-737-1334

Keywords *Helianthus* · Simple sequence repeats · Native American land races · Hopi · Crop domestication

Introduction

The elite parents of contemporary single-cross hybrids of cultivated sunflower (*Helianthus annuus* L.) are the product of multiple domestication and breeding bottlenecks. The domesticated gene pool was shaped by complex forces: domestication by Native Americans, dispersion of wild sunflower achenes in the Great Plains by bison (*Bison bison* L.) and other mammals, dispersion of wild, weedy and domesticated forms by Native Americans, intercontinental migration and trade of seeds transported by early European explorers, outcrossing between wild and cultivated forms, artificial selection from intraspecific and interspecific hybrids, greatly intensified artificial selection for increased seed oil in the mid-twentieth century in eastern Europe and, finally, the remigration of seeds of founders of present-day 'oilseed' sunflowers back to North America from Europe (Pustovoit 1964; Semelczi-Kovacs 1975; Heiser 1976; Arias and Rieseberg 1994; Linder et al. 1996; Putt 1997; Seiler and Rieseberg 1997; Heiser 2001; Lentz et al. 2001a). Collectively, the archaeological, historical, and breeding records show that sunflower domestication from wild ancestors to present-day oilseed hybrids has not progressed in a linear way.

Sunflowers played an important role in the diet and culture of Native Americans and were cultivated as early as 3,000 BC in North America (Heiser 1976, 2001). Despite an anecdotal claim by Whiting (1939) (recounted by Putt 1997) that sunflowers were domesticated before maize (*Zea mays* L.), thereby implying that sunflowers were cultivated earlier than 5,000 years before the present (BP), archaeological specimens to support such a claim have not been discovered (Heiser 1976, 2001; Lentz et al. 2001a, b). Lentz et al. (2001a) unearthed achenes of domesticated sunflowers from an archeological site in San Andrés, Mexico, and, using ac-

celerated mass spectroscopy, estimated the age of deposit to be 4,085 to 4,130 BP. The San Andrés achenes were larger than achenes from wild populations (a hallmark of domestication) and thus seem to be the most ancient domesticated sunflower achenes discovered to-date (Lentz et al. 2001a). Because the San Andrés site is far south of the estimated native range of sunflowers in 4,000 BP, the achenes almost certainly had to have been imported by Native Americans, implying that sunflowers must have been domesticated earlier than 4,130 BP (Heiser 1976; Rogers et al. 1982; Lentz et al. 2001a).

Wild and domesticated sunflowers were introduced to Europe by Spanish explorers in the early sixteenth century and rapidly gained in popularity (Semelczi-Kovacs 1975). The Russians were the first to grow sunflowers for oil on a significant scale (Heiser 1976; Putt 1997). By the turn of the twentieth century, sunflower seed oil concentrations ranged from about 100 to 300 g/kg (Putt 1997). By selecting for thin-hulled, oil-rich achenes, Pustovoit (1964) and other Russian plant breeders produced the key milestone in the development of modern-day oilseed sunflowers (Putt 1997). Seed oil concentrations were increased from an upper limit of 330 g/kg in 1940 to as much as 550 g/kg by 1964 (Pustovoit 1964). The other key milestones in the development of contemporary oilseed sunflowers were the discovery of cytoplasmic-genic male-sterility (CMS) (Leclercq 1969) and, soon thereafter, genes for restoring the fertility of CMS lines (Kinman 1970). Thus, in the 30 year period spanning 1940 to 1970, two dramatic phenotypes substantially transformed sunflower as a crop and spawned two genetic bottlenecks leading up to present-day oilseed sunflower hybrids.

The ancestral history of oilseed germplasm originating in the post-CMS era has been reconstructed using pedigree records (Korell et al. 1992), coancestry analysis (Cheres and Knapp 1998) and phylogenetic analyses based on a variety of molecular markers (Gentzbittel et al. 1992, 1994; Quillet et al. 1992; Berry et al. 1994; Lawson et al. 1994; Tersac et al. 1994; Hongtrakul et al. 1997; Paniego et al. 2002; Yu et al. 2002a). Typically, elite oilseed fertility restorer (R) and sterility maintainer (B) lines coalesce into distinct clades. The oilseed B and R groupings found in sunflower mirror genetic bottlenecks created by the necessity of maintaining heterotic groups and hybrid seed production traits (branching and fertility restoration). Moreover, dominant downy mildew [*Plasmopara halstedii* (Farl.) Berlese et de Toni] and rust (*Puccinia helianthi* Schw.) resistance genes have historically been incorporated into fertility restorer lines (male heterotic groups) in sunflower, thereby perpetuating the fertility restorer bottleneck.

Oilseed sunflowers are the most recent domesticates in sunflower, with a history spanning slightly more than half a century (Pustovoit 1964; Putt 1997). The other important domesticates are the Native American land races (Heiser 1945, 1946, 1951, 1955, 1976; Nabhan 1982; Rieseberg and Seiler 1990) and confectionery sunflowers. Presumably, Native American land races are living genetic records of the first monocephalic (unbranched) sunflowers and the most ancient extant domesticates in sunflower. Monocephaly pleiotropically increases achene dimension and concentrates achene yield, and was the key breakthrough in sunflower domestication, fostering cultivation in prehistoric times (Heiser 1976). Confectionery sunflowers, while morphologically similar to the Native American land races, are chronologically recent domesticates, originating in the last two or three centuries (Putt 1997). Confectionery sunflowers are typified by Mennonite and other large, easily dehulled, black and white striped achenes. Confectionery inbred lines developed over the last quarter century originated from a narrow genetic base (Cheres and Knapp 1998). Key founders of the group are Commander, Mennonite, Bonita Giant Manchurian, VNIIMK1646 and Sundak.

Rieseberg and Seiler (1990) reviewed hypotheses on the origin of domesticated sunflowers and assessed molecular genetic diversity among elite open-pollinated (OP) oilseed and confectionery germplasm, Native American land races, and wild populations using allozyme and chloroplast DNA polymorphisms. Domesticated and wild sunflowers were found to be extraordinarily similar. The mean genetic identity between wild and domesticated germplasm accessions was 0.93 (Rieseberg and Seiler 1990). Moreover, "mean genetic identities among wild populations of *H. annuus* and among accessions of domesticated *H. annuus* were 0.96 and 0.95, respectively", and of 30 allozyme alleles found in elite germplasm, only one was not found in the wild gene pool (Rieseberg and Seiler 1990). Using random amplified polymorphic DNA (RAPD) markers, Arias and Rieseberg (1995) found domesticated and wild sunflowers to be extraordinarily similar. Genetic similarities among 20 domesticated and 11 wild sunflower accessions ranged from 0.976 to 0.997. Hence, allozyme and RAPD polymorphisms were insufficient for gaining insights into the origins of domesticated sunflowers or distinguishing between closely or distantly related germplasm accessions (Rieseberg and Seiler 1990; Arias and Rieseberg 1995). Heiser (1985) postulated independent origins of domesticated sunflowers in the central US and Mexico and the southwestern US, whereas Seiler and Rieseberg (1997) concluded that "the lack of domesticated achenes in archaeological records outside of the central USA, the morphological similarity among Native American varieties of the domesticated sunflower, and the virtual monomorphism for isozymes and chloroplast DNA in cultivated lines do not support the multiple origin hypothesis (Rieseberg and Seiler 1990)". Lentz et al. (2001a) subsequently discovered achenes of domesticated sunflowers in Mexico and argued for a single domestication in Mexico, a supposition questioned by Heiser (2001). Using allozymes, Cronn et al. (1997) found that domesticated and wild sunflowers formed two nearly independent groups. They speculated that sunflowers may have been domesticated from wild populations of the Great Plains.

Table 1 Common names, plant introduction numbers, germplasm groups, and origins for 47 sunflower (*H. annuus* L.) germplasm accessions

Common name	Number	Germplasm group	Origin
HA287	PI552933	Confectionery B line	USDA-ARS
HA292	PI552937	Confectionery B line	USDA-ARS
RHA280	PI552943	Confectionery R line	USDA-ARS
RHA282	PI552944	Confectionery R line	USDA-ARS
HA89	PI599773	Oilseed B line	USDA-ARS
HA369	PI534655	Oilseed B line	USDA-ARS
HA370	PI534656	Oilseed B line	USDA-ARS
HA371	PI534657	Oilseed B line	USDA-ARS
HA372	PI534658	Oilseed B line	USDA-ARS
HA383	PI578872	Oilseed B line	USDA-ARS
HA407	PI597371	Oilseed B line	USDA-ARS
HA821	PI599984	Oilseed B line	USDA-ARS
RHA274	PI599759	Oilseed R line	USDA-ARS
RHA373	PI560141	Oilseed R line	USDA-ARS
RHA377	PI560145	Oilseed R line	USDA-ARS
RHA392	PI603988	Oilseed R line	USDA-ARS
RHA409	PI603990	Oilseed R line	USDA-ARS
RHA417	PI600000	Oilseed R line	USDA-ARS
RHA801	PI599768	Oilseed R line	USDA-ARS
Peredovik	Ames 1838	OP Oilseed cultivar	Russia
VNIIMK8931	PI340790	OP Oilseed cultivar	Russia
Pervenets	PI483077	OP Oilseed cultivar	Russia
Tchernianka Select W13	PI343794	OP Oilseed cultivar	Russia
Arikara	PI369357	Native American land race	North Dakota
Havasupai	PI369358	Native American land race	Arizona
Hopi	PI369359	Native American land race	Arizona
Seneca	PI369360	Native American land race	New York
Tarahumara	–	OP Confectionery cultivar	Mexico
Mennonite	Ames 7574	OP Confectionery cultivar	Canada
Jilin	Ames 10106	OP Confectionery cultivar	China
Zambian	PI500689	OP Confectionery cultivar	Zambia
Abendsonne Red	PI490316	Ornamental	Germany
ANN1811-TX	PI494567	Wild population	Texas
ANN1238-NE	–	Wild population	Nebraska
PI-CO	PI468625	Wild population	Colorado
PI-NV	PI468596	Wild population	Nevada
PI-WA	PI531018	Wild population	Washington
PI-WY	PI413019	Wild population	Wyoming
PI-ND	PI468439	Wild population	North Dakota
PI-AZ	PI468575	Wild population	Arizona
PI-OK	PI435619	Wild population	Oklahoma
PI-MX	PI413123	Wild population	Mexico
PI-OR	PI531015	Wild population	Oregon
PI-CA	PI435593	Wild population	California
PI-MT	PI531022	Wild population	Montana
PI-SD	PI413039	Wild population	South Dakota
PI-UT	PI468619	Wild population	Utah

The recent development of several hundred microsatellite markers for sunflower (Tang et al. 2002; Yu et al. 2002a), coupled with the unparalleled power of microsatellites for discriminating between closely related taxa and individuals within taxa (Bowcock et al. 1994; Goldstein et al. 1995a, b; Powell et al. 1996; Matsuoka et al. 2002b), has opened the way to a reanalysis of molecular genetic diversity in cultivated sunflower. We present an analysis of the allelic diversity of 122 microsatellite loci among anciently and recently domesticated lineages and geographically diverse wild populations of sunflower. The goals of the present study were to test the hypothesis of a single domestication origin (Rieseberg and Seiler 1990; Arias and Rieseberg 1995; Seiler and Rieseberg 1997; Heiser 2001; Lentz et al. 2001a, b) and gain insights into the structure and magnitude of molecular genetic diversity in the domesticated and wild gene pools of cultivated sunflower.

Materials and methods

Microsatellite genotyping

Microsatellite genotypes were produced for 19 elite inbred lines and 28 domesticated and wild germplasm accessions using 122 microsatellite markers selected from the public collection (Tang et al. 2002; Yu et al. 2002a) (Table 1). Seed samples for the analysis were acquired from Jerry Miller (USDA-ARS, Fargo, N.D.), Mary Brothers (USDA-ARS National Plant Germplasm System, North Central Plant Introduction Station, Ames, Iowa) and Seeds of Change (Albuquerque, N.M.) (Table 1). Genomic DNA was isolated from one individual per germplasm accession using fresh leaf tissue harvested from 6-week-old plants and a modified

CTAB method (Webb and Knapp 1990). Microsatellite genotyping assays were performed on an ABI377 (Applied Biosystems, Foster City, Calif.) using fluorescently labelled amplicons, filter set C, and 10 to 20 ng of template DNA per sample as described by Tang et al. (2002). Microsatellite fragment lengths were recorded using GeneScan 2.1 and genotypes were ascertained using Genotyper 2.1. Genotyping assays were performed by separately amplifying individual microsatellite markers, pooling six or more amplicons by color and length, and diluting the pooled samples 15- to 25-fold.

Statistical analyses

Heterozygosities (Ott 1991), allele numbers, taxon-specific allele numbers, and null allele numbers and frequencies were estimated for each microsatellite marker locus using MicroSat (http://hpgl.stanford.edu/MicroSat). Pairwise genetic distances among the 47 germplasm accessions were estimated using the "proportion of shared alleles" estimator (D_{PS}) of Bowcock et al. (1994). MicroSat was used to estimate the mean D_{PS} matrix from 1,000 bootstrap samples drawn from the complete set of polymorphic microsatellite loci (112 out of 122). MEGA (Kumar et al. 2001) was used to construct and draw a minimum evolution tree (Rzhetsky and Nei 1992a, b; Nei and Kumar 2000; Takahashi and Nei 2000) from the bootstrap mean D_{PS} matrix. Principal component analysis (PCA) was performed on the bootstrap mean D_{PS} matrix using the SAS program PROC PRINCOMP (Statistical Analysis System, Cary, N.C.).

MicroSat was used to estimate 1,000 D_{PS} matrices by bootstrapping the complete set of polymorphic microsatellite loci. The PHYLIP program NEIGHBOR (http://evolution.genetics.washington.edu/phylip.html) was used to estimate 1,000 UPGMA trees from the bootstrapped D_{PS} matrices. The PHYLIP program CONSENSUS was used to construct consensus trees (Margush and McMorris 1981) from the 1,000 bootstrapped UPGMA trees. The criteria for setting branches (identifying monophyletic groups) in the consensus tree was the presence of a particular branch in 60% or more of the bootstrapped trees. We produced a 'fully resolved' consensus tree to ascertain branch frequencies for every germplasm accession. Both trees were drawn using TreeView (http://taxonomy.zoology.gla.ac.uk/rod/rod.html).

MicroSat was used to estimate pairwise $(\delta\mu)^2$ (Goldstein et al. 1995b) genetic distances among the 47 germplasm accessions from the complete set of polymorphic microsatellite loci and a subset of 56 dinucleotide repeat loci with allele distributions fitting the stepwise mutation model (SMM) (Ohta and Kimura 1973; Schlötterer and Tautz 1992; Shriver et al. 1993; Valdes et al. 1993). Principal component analyses were performed on both $(\delta\mu)^2$ matrices using PROC PRINCOMP.

Results

Allelic diversity of microsatellites in domesticated and wild sunflowers

Nineteen elite inbred lines and 28 domesticated and wild outbred populations (Table 1) were genotyped using 105 mapped and 17 unmapped microsatellite markers (Table 2). The former are dispersed throughout the sunflower genome (Tang et al. 2002; Yu et al. 2002b). The selected microsatellite markers amplified a single locus each across the 47 germplasm accessions. Heterozygosities, allele numbers, and other summary statistics are reported for each microsatellite locus in Table 2. The number of universally codominant microsatellite markers was greater for inbred lines (116) than Native American land races and open-pollinated germplasm and cultivars (exotic domesticates) (111) and wild populations (92). Of 122 microsatellites, 88 were codominant (produced no null alleles) across the 47 germplasm accessions (Table 2). Null allele frequencies were 0.0142 among inbred lines, 0.0170 among exotic domesticates and 0.0508 among wild populations (Table 3). ORS534, an anomaly, produced 11 null alleles among elite inbred lines, but only one among exotic domesticates and three among wild populations.

The polymorphic SSRs (112) produced 1,341 alleles among the 47 germplasm accessions (12.0 alleles per locus) (Fig. 1). The mean number of alleles per locus was nearly three-fold greater among wild populations (9.7) than elite inbred lines (3.5) (Table 3). The maximum number of alleles per locus was eight for elite inbred lines, 11 for exotic domesticates, 17 for wild populations, and 21 across germplasm accessions (Fig. 1). Eighteen microsatellite markers (14.8%) were monomorphic among elite inbred lines, whereas 11 (9.0%) were monomorphic among wild populations (Table 2 and Fig. 1). Of the latter, ten were monomorphic across the 47 germplasm accessions.

Mean heterozygosities estimated from polymorphic microsatellite marker loci only, were 0.515 among inbred lines, 0.638 among exotic domesticates and 0.817 among wild populations (Table 3). Heterozygosities for individual microsatellite markers ranged from 0.10 to 0.81 among inbred lines, 0.15 to 0.89 among exotic domesticates, and 0.36 to 0.92 among wild populations (Table 2). The shapes of the heterozygosity distributions were markedly different across germplasm groups (Fig. 2). The histogram for elite inbred lines was platykurtic and nearly uniform, whereas the histogram for wild populations was right-skewed and approximately exponential. The histogram for exotic domesticates was intermediate between the two. The histograms shapes were characteristic of an intensely bred species. Predictably, heterozygosity decreased as the germplasm became progressively more domesticated (Table 3).

The number of taxon-specific alleles was significantly greater among wild populations (394) than exotic domesticates (95) and elite inbred lines (15) (Fig. 3). The mean number of taxon-specific alleles per germplasm accession was 10-fold greater among exotic domesticates (7.31) and 35-fold greater among wild populations (26.27) than among elite inbred lines (0.75). The number of taxon-specific alleles per germplasm accession ranged from zero to three among elite inbred lines and from 13 (PI468439-ND) to 47 (PI468619-OK) among wild populations (Fig. 3). The Native American land races (Arikara, Hopi, Havasupai and Seneca) had six to 15 taxon-specific alleles each, less than any of the wild populations other than PI468439-ND. Three oilseed R-lines (RHA392, RHA409 and RHA417) and one oilseed B-line (HA370) had one taxon-specific allele each, and two oilseed B-lines (HA369 and HA372) had three taxon-specific alleles each. No other taxon-specific alleles were identified among oilseed B or R lines. Of the four confectionery inbred lines,

Table 2 Heterozygosities among elite inbred lines (H_E), exotic domesticates (H_D), and wild populations (H_W) and across groups (H_T), allele numbers among elite inbred lines (A_E), exotic domesticates (A_D), and wild populations (A_W) and across groups (A_T), and null allele numbers among elite inbred lines (N_E), exotic domesticates (N_D), and wild populations (N_W) and across groups (N_T) for 122 sunflower microsatellite marker loci

Marker	H_E	H_D	H_W	H_T	A_E	A_D	A_W	A_T	N_E	N_D	N_W	N_T
ORS7	0.545	0.726	0.780	0.771	3	5	6	11	0	0	0	0
ORS8	0.635	0.554	0.671	0.657	3	3	4	4	0	0	0	0
ORS53	0.755	0.747	0.892	0.864	5	7	11	16	0	0	0	0
ORS57	0.455	0.684	0.792	0.714	2	4	7	9	0	6	4	10
ORS64	0.455	0.765	0.818	0.772	2	6	7	9	0	0	0	0
ORS70	0.595	0.789	0.922	0.833	5	7	14	18	0	0	0	0
ORS78	0.655	0.681	0.360	0.713	4	4	4	5	0	0	0	0
ORS134	0.480	0.684	0.830	0.788	2	8	11	18	0	0	6	6
ORS148	0.265	0.789	0.813	0.718	3	8	7	12	0	0	0	0
ORS158	0.590	0.719	0.840	0.787	5	6	7	8	0	1	4	5
ORS169	0.445	0.695	0.895	0.773	3	7	13	17	0	0	1	1
ORS176	0.725	0.704	0.859	0.850	4	5	10	14	0	1	4	5
ORS187	0.375	0.486	0.864	0.694	2	3	11	12	0	9	3	12
ORS229	0.620	0.775	0.853	0.836	5	6	10	10	0	0	0	0
ORS243	0.445	0.498	0.735	0.604	3	3	8	9	0	0	1	1
ORS296	0.000	0.000	0.469	0.251	1	1	2	2	0	0	10	10
ORS297	0.688	0.744	0.868	0.841	5	5	11	11	0	0	0	0
ORS299	0.460	0.578	0.816	0.682	3	6	7	12	0	0	0	0
ORS301	0.460	0.398	0.796	0.653	3	3	7	7	0	0	0	0
ORS303	0.420	0.375	0.000	0.350	2	2	1	2	0	0	0	0
ORS304	0.000	0.000	0.000	0.000	1	1	1	1	0	0	0	0
ORS305	0.000	0.000	0.000	0.000	1	1	1	1	0	0	0	0
ORS307	0.545	0.641	0.781	0.789	3	5	11	13	0	0	0	0
ORS309	0.480	0.543	0.485	0.568	2	3	5	5	0	0	0	0
ORS310	0.700	0.859	0.893	0.883	4	9	12	18	0	0	0	0
ORS311	0.320	0.633	0.866	0.710	2	6	11	16	0	1	0	1
ORS312	0.465	0.611	0.865	0.724	4	5	9	13	0	0	2	2
ORS313	0.000	0.000	0.000	0.000	1	1	1	1	0	0	0	0
ORS314	0.000	0.000	0.000	0.000	1	1	1	1	0	0	0	0
ORS316	0.515	0.665	0.885	0.827	3	5	12	12	0	0	0	0
ORS317	0.740	0.649	0.893	0.866	5	4	12	16	0	0	0	0
ORS318	0.000	0.255	0.810	0.488	1	3	6	7	0	0	2	2
ORS319	0.000	0.000	0.000	0.000	1	1	1	1	0	0	0	0
ORS320	0.000	0.000	0.000	0.000	1	1	1	1	0	0	0	0
ORS322	0.000	0.000	0.000	0.000	1	1	1	1	0	0	0	0
ORS326	0.420	0.785	0.916	0.828	2	7	14	18	0	0	0	0
ORS329	0.515	0.667	0.780	0.737	3	3	7	8	0	0	0	0
ORS331	0.725	0.766	0.868	0.838	4	7	9	10	0	0	0	0
ORS337	0.375	0.464	0.830	0.726	2	4	9	9	0	0	0	0
ORS338	0.265	0.659	0.917	0.780	3	5	14	17	0	0	0	0
ORS339	0.000	0.338	0.867	0.627	1	3	11	13	0	0	0	0
ORS342	0.465	0.790	0.908	0.826	3	9	14	17	0	0	0	0
ORS343	0.255	0.357	0.757	0.633	2	3	8	8	0	0	0	0
ORS344	0.265	0.789	0.846	0.737	3	8	8	12	0	0	0	0
ORS345	0.000	0.000	0.000	0.000	1	1	1	1	0	0	0	0
ORS346	0.000	0.000	0.000	0.000	1	1	1	1	0	0	0	0
ORS349	0.000	0.593	0.851	0.674	1	4	9	9	0	0	2	2
ORS351	0.415	0.711	0.824	0.723	4	5	8	9	0	0	0	0
ORS352	0.375	0.486	0.750	0.625	2	3	7	7	5	3	7	15
ORS354	0.345	0.676	0.891	0.710	4	7	11	16	0	0	1	1
ORS355	0.000	0.475	0.890	0.658	1	3	13	13	0	0	0	0
ORS356	0.000	0.338	0.696	0.462	1	3	6	7	0	0	0	0
ORS358	0.410	0.679	0.890	0.771	4	6	11	13	0	0	4	4
ORS361	0.180	0.645	0.781	0.667	2	4	8	9	0	0	0	0
ORS363	0.690	0.560	0.830	0.823	5	4	8	10	5	0	1	6
ORS365	0.255	0.450	0.719	0.625	2	3	6	8	0	0	4	4
ORS366	0.625	0.737	0.825	0.826	6	8	9	14	0	0	0	0
ORS371	0.420	0.748	0.837	0.779	2	5	8	8	0	0	0	0
ORS375	0.000	0.292	0.512	0.317	1	3	4	5	0	0	0	0
ORS380	0.555	0.630	0.837	0.752	3	4	9	14	0	0	0	0
ORS381	0.670	0.695	0.870	0.764	3	4	11	11	0	0	1	1
ORS388	0.600	0.730	0.923	0.829	5	7	16	18	0	0	0	0
ORS391	0.273	0.667	0.484	0.509	3	5	6	7	0	0	0	0
ORS393	0.000	0.000	0.000	0.000	1	1	1	1	0	0	0	0
ORS395	0.000	0.418	0.698	0.459	1	3	6	6	0	0	0	0

Table 2 (continued)

Marker	H_E	H_D	H_W	H_T	A_E	A_D	A_W	A_T	N_E	N_D	N_W	N_T
ORS397	0.500	0.526	0.781	0.693	2	4	5	5	0	0	0	0
ORS398	0.480	0.715	0.885	0.812	2	6	11	14	0	0	0	0
ORS400	0.495	0.150	0.798	0.633	2	2	8	9	0	0	0	0
ORS407	0.675	0.728	0.874	0.846	4	5	11	12	0	0	0	0
ORS409	0.720	0.720	0.687	0.823	6	6	6	10	0	0	0	0
ORS423	0.720	0.770	0.913	0.862	4	8	15	16	0	0	0	0
ORS426	0.500	0.000	0.716	0.666	2	1	5	6	0	0	0	0
ORS428	0.515	0.720	0.776	0.756	3	5	6	7	0	0	0	0
ORS432	0.595	0.635	0.806	0.757	3	3	8	8	0	0	0	0
ORS437	0.685	0.704	0.867	0.833	4	7	10	14	0	0	0	0
ORS442	0.545	0.588	0.857	0.783	4	4	10	12	0	0	0	0
ORS447	0.750	0.859	0.859	0.881	6	10	12	21	3	1	5	9
ORS457	0.590	0.680	0.910	0.819	4	4	13	14	0	0	0	0
ORS468	0.485	0.480	0.562	0.580	3	3	4	6	0	0	0	0
ORS471	0.095	0.642	0.858	0.689	2	6	11	15	0	0	5	5
ORS485	0.500	0.245	0.835	0.676	2	3	9	13	0	0	1	1
ORS503	0.180	0.305	0.888	0.662	2	2	12	13	0	0	1	1
ORS509	0.645	0.720	0.853	0.842	6	5	11	15	0	0	0	0
ORS510	0.410	0.595	0.870	0.749	4	3	9	9	0	0	0	0
ORS513	0.715	0.756	0.749	0.808	5	5	7	10	0	0	0	0
ORS533	0.645	0.826	0.853	0.887	5	7	9	15	0	1	0	1
ORS534	0.630	0.814	0.898	0.868	5	7	12	15	11	1	3	15
ORS546	0.735	0.809	0.900	0.875	5	8	13	15	0	0	0	0
ORS547	0.645	0.740	0.912	0.826	4	6	14	16	0	0	1	1
ORS561	0.710	0.810	0.915	0.904	5	7	14	19	0	0	0	0
ORS578	0.515	0.357	0.845	0.595	3	4	10	12	0	0	3	3
ORS595	0.810	0.866	0.875	0.910	6	9	12	17	0	0	0	0
ORS596	0.255	0.582	0.864	0.675	2	4	9	9	0	0	0	0
ORS599	0.800	0.891	0.864	0.925	8	11	11	21	0	0	0	0
RS605	0.685	0.685	0.832	0.838	4	4	9	11	0	0	1	1
ORS607	0.685	0.637	0.796	0.795	4	4	7	10	0	0	6	6
ORS608	0.335	0.688	0.870	0.737	3	5	12	15	3	2	0	5
ORS609	0.320	0.611	0.901	0.763	2	5	13	15	0	0	1	1
ORS610	0.635	0.654	0.910	0.818	5	5	13	14	0	0	0	0
ORS612	0.095	0.255	0.898	0.651	2	3	12	14	0	0	0	0
ORS613	0.665	0.741	0.882	0.880	4	7	13	15	0	0	0	0
ORS617	0.445	0.337	0.581	0.673	3	3	6	7	0	0	0	0
ORS620	0.410	0.640	0.870	0.733	4	4	10	13	0	0	0	0
ORS621	0.725	0.735	0.857	0.847	5	6	10	14	0	0	0	0
ORS662	0.785	0.820	0.923	0.900	5	6	17	20	0	0	0	0
ORS665	0.710	0.810	0.911	0.868	5	8	15	17	0	0	0	0
ORS674	0.760	0.671	0.820	0.819	5	5	8	10	0	0	0	0
ORS677	0.525	0.398	0.805	0.688	4	3	7	9	0	0	0	0
ORS678	0.180	0.798	0.813	0.744	2	8	12	16	0	0	1	1
ORS683	0.540	0.676	0.858	0.750	3	5	9	11	0	0	3	3
ORS687	0.600	0.680	0.848	0.807	4	4	10	12	0	0	0	0
ORS691	0.535	0.851	0.902	0.864	3	9	14	17	0	0	0	0
ORS697	0.180	0.691	0.851	0.688	2	6	11	11	0	0	0	0
ORS703	0.180	0.531	0.813	0.611	2	4	7	7	0	0	0	0
ORS767	0.445	0.622	0.907	0.752	3	4	14	14	0	0	0	0
ORS774	0.545	0.750	0.884	0.774	3	5	12	14	0	0	0	0
ORS779	0.255	0.674	0.899	0.655	2	5	14	14	0	0	0	0
ORS782	0.480	0.858	0.910	0.859	5	9	13	17	0	0	2	2
ORS799	0.705	0.759	0.695	0.820	4	6	7	11	6	1	3	10
ORS807	0.590	0.735	0.903	0.834	6	9	13	19	0	0	0	0
ORS810	0.495	0.561	0.451	0.613	2	3	4	6	0	0	0	0
ORS811	0.615	0.681	0.632	0.763	4	4	7	8	0	0	0	0

RHA292 had one, and HA287 and RHA282 had two, taxon-specific alleles each (Fig. 3).

Domesticated and wild sunflowers form distinct groups

Genetic distances (D_{PS}) among the 47 germplasm accessions ranged from 0.252 (HA371–HA372) to 0.945 (RHA280–ANN1811). Wild populations clustered independent of elite inbred lines and other domesticated sunflowers (Figs. 4–6). The tree constructed from the D_{PS} matrix has several prominent clades and depicted the narrowing of genetic diversity from undomesticated sunflowers (lowermost clade) to recent and prehistoric domesticates (middle clades) to most recent domesticates (uppermost clades) (Fig. 5).

Table 3 Mean heterozygosities for monomorphic and polymorphic microsatellite marker loci ($\bar{H}_T$) and polymorphic microsatellite marker loci only $\bar{H}$, mean number of alleles per microsatellite marker locus ($\bar{n}_A$), and mean number of null alleles per SSR marker locus ($\bar{n}_N$) for 122 microsatellite marker loci genotyped on 19 elite confectionery and oilseed inbred lines, four Native American land races and nine elite and exotic open-pollinated populations (exotic domesticates), and 15 wild populations

Group	$\bar{H}_T$	$\bar{H}$	($\bar{n}_A$)	($\bar{n}_N$)
Elite inbred lines	0.430	0.515	3.5	0.3
Exotic domesticates	0.568	0.638	5.2	0.2
Wild populations	0.740	0.817	9.7	0.8
Total	0.674	0.740	12.0	1.3

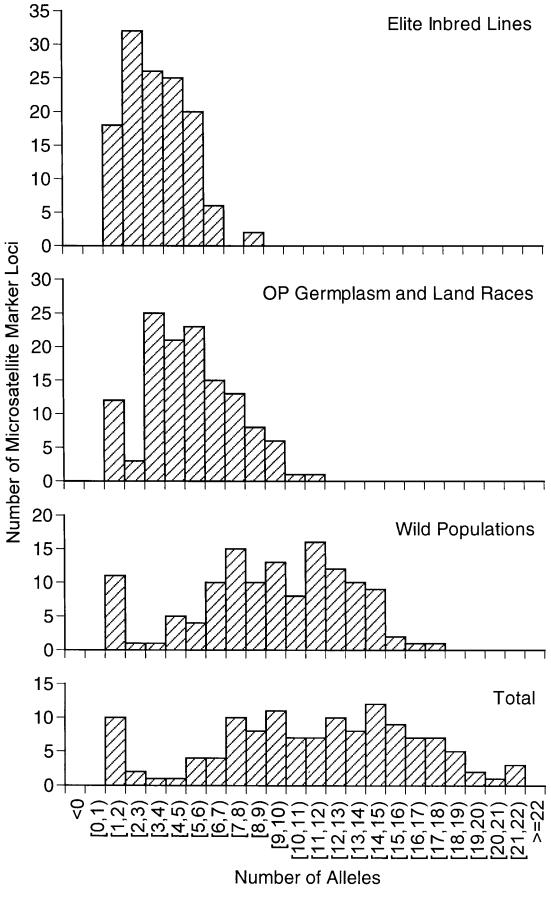

Fig. 1 The number of alleles per locus for 122 microsatellite markers genotyped on 47 germplasm accessions

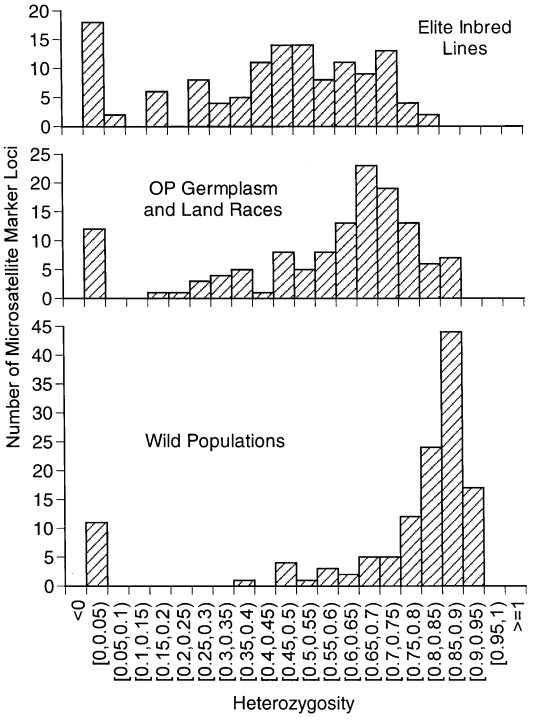

Fig. 2 Heterozygosities for 122 microsatellite markers genotyped on 47 germplasm accessions

Hopi and Havasupai (D_{PS} = 0.416) formed a clade distinct from other domesticated and wild sunflower clades (Fig. 5). The Hopi-Havasupai branch was present in 100% of the bootstrapped trees (Fig. 6). Genetic distances between Hopi-Havasupai and the other Native American land races ranged from 0.714 to 0.798. The distinctness of the Hopi-Havasupai group was particularly prominent in the principal score plot (Fig. 4), where both land races were well separated from other domesticated and wild sunflowers (the first two principal components accounted for 23.4 and 6.7% of the variance in the genetic distance matrix).

PI-ND and Abendsonne Red, the only ornamental sunflower genotyped, clustered independent of other germplasm accessions (Figs. 4–6). The PI-ND branch was present in 98.4% and the Abendsonne Red branch was present in 80.2% of the bootstrapped trees (Fig. 6). PI-ND, a wild population from North Dakota (PI468439), was the only outlier among the wild populations. Genetic distances among wild populations ranged from 0.714 between PI-MX and PI-AZ, to 0.878 between PI-OK and PI-AZ, and PI-WY and ANN1238. Genetic distances between PI-ND and oilseed B- and R-lines ranged from 0.521 to 0.756, whereas genetic distances between the other wild populations and oilseed B- and R-lines ranged from 0.794 to 0.849 on the low end (PI-OK) to 0.866 to 0.929 on the high end (PI-CO); thus, PI-ND had more alleles in common with oilseed inbred lines than the other 14 wild populations and grouped between the domesticated and wild sunflower clades (Figs. 4–6). We speculate that PI-ND is the product of the introgression of genes from modern-day oilseed cultivars (single-cross hybrids) into a native North Dakotan population.

Recently domesticated sunflowers clustered into predictable groups. Oilseed inbred lines and OP cultivars formed groups separate from confectionery inbred lines and OP cultivars (Figs. 4–5). Oilseed B- and R-lines

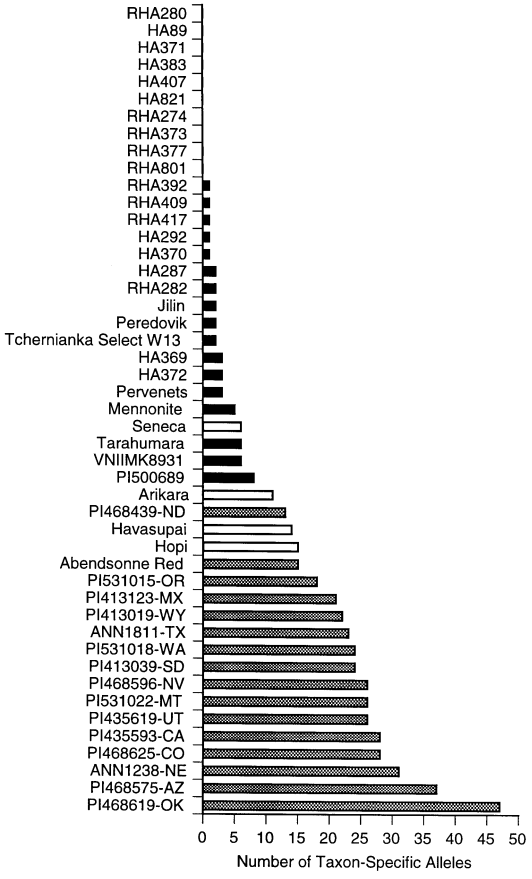

Fig. 3 The number of taxon-specific alleles per germplasm accession for 122 microsatellite markers genotyped on elite confectionery and oilseed sunflower germplasm (*solid bars*), Native American land races (*open bars*), and wild populations (*stippled bars*)

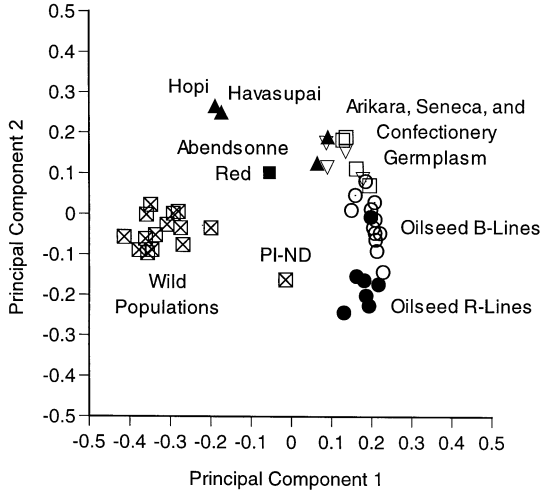

Fig. 4 Principal scores for the first two principal components produced by principal component analysis of the mean genetic distance matrix (D_{PS}) estimated from 1,000 bootstrap samples among oilseed fertility restorer (*R*) lines (●), oilseed sterility maintainer (*B*) lines (○), open-pollinated oilseed germplasm (⊙), confectionery B and R lines (□), open-pollinated confectionery germplasm (▽), ornamental germplasm (■), Native American land races (▲) and wild populations (⊠) of cultivated sunflower

formed separated groups. Open-pollinated oilseed cultivars and populations (Pervenets, VNIIMK8931, Peredovik and Tchernianka Select W13) clustered close to oilseed B- and R-lines. RHA392, an oilseed R-line, grouped closest to confectionery B- and R-lines and was the only oilseed R-line to fall outside the oilseed R-line cluster (Fig. 5). Similarly, HA369, an oilseed B-line, clustered with OP oilseed cultivars and was the only oilseed B-line to fall outside the oilseed B-cluster.

Elite confectionery inbred lines (identified by -C suffixes in Figs. 5–6) formed a single group, but did not separate into B and R subgroups, partly because we only sampled four inbred lines reported to belong to different heterotic groups (Cheres and Knapp 1998). On the basis of pedigree records, elite confectionery inbred lines were found to be one of the most genetically narrow germplasm groups in sunflower (Cheres and Knapp 1998). However, genetic distances among confectionery inbred lines ranged from 0.408 to 0.529 and the confectionery gene pool, as a whole, seems to be as diverse as the oilseed gene pool (Figs. 4–5; Cronn et al. 1997).

Two Native American land races (Arikara and Seneca) and four open-pollinated confectionery cultivars (Mennonite, Tarahumara, Jilin and PI500689) formed a clade proximal to the oilseed and confectionery inbred line clades (Fig. 5). Similarly, the two land races grouped close to HA287, RHA280, Jilin and PI500689 in the principal score plot (Fig. 4). Thus, land races domesticated by Native Americans of the Great Plains and northeastern US seem to be closely related to historically important confectionery germplasm and probably played an important role in founding the contemporary confectionery sunflower gene pool. The whole group is anchored by Mennonite, a prototypical black and white striped confectionery sunflower and one of the principal founders of contemporary confectionery inbred lines (Cheres and Knapp 1998). Tarahumara, a white hulled confectionery sunflower, grouped with Arikara, Seneca and other confectionery sunflowers (Figs. 4–6). Despite the Native American name, Tarahumara is a descendent of a Canadian Mennonite land race spread by Chihuahuan Mennonites and has only been cultivated by the Tarahumarans for the last 40 years (Seeds of Change, Albuquerque, New Mexico). Hence, Tarahumara may be a recent domesticate (founded in the last 200 to 300 years) along with Mennonite, Jilin, PI500689, and other elite open-pollinated populations and inbred lines descended from ancient land races or crosses to the latter.

Phylogeographic diversity among wild populations

Using D_{PS} as a measure of genetic distance, phylogeographic patterns of diversity were not found among the 15 wild populations (Figs. 4–6). By contrast, using allozyme markers, Cronn et al. (1997) found that wild populations from the Great Plains and the western US formed two distinct groups, although two outliers from the Great Plains were present in the western group. The

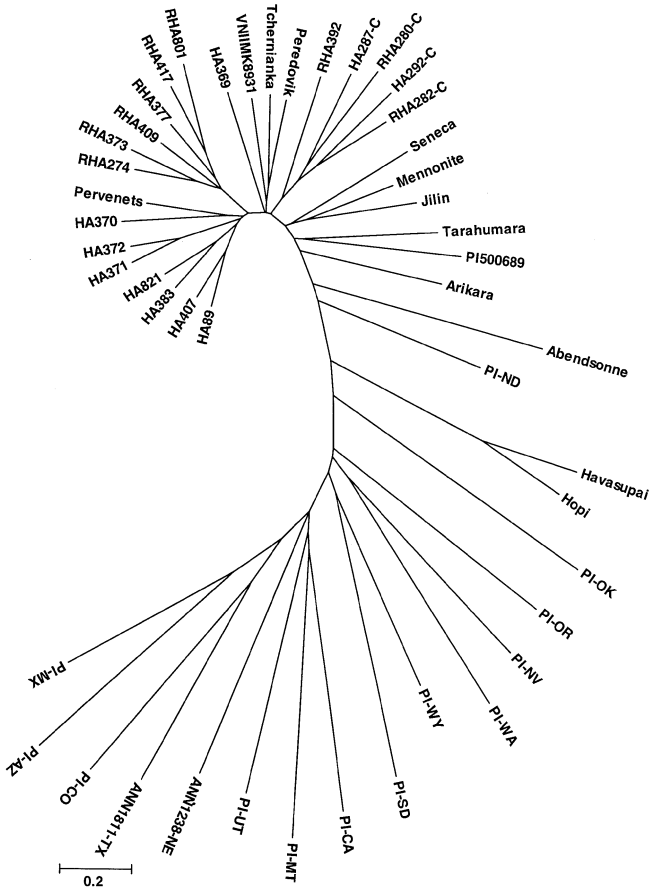

Fig. 5 Minimum evolution tree produced from the mean genetic distance matrix (D_{PS}) estimated from 1,000 bootstrap samples among 47 germplasm accessions

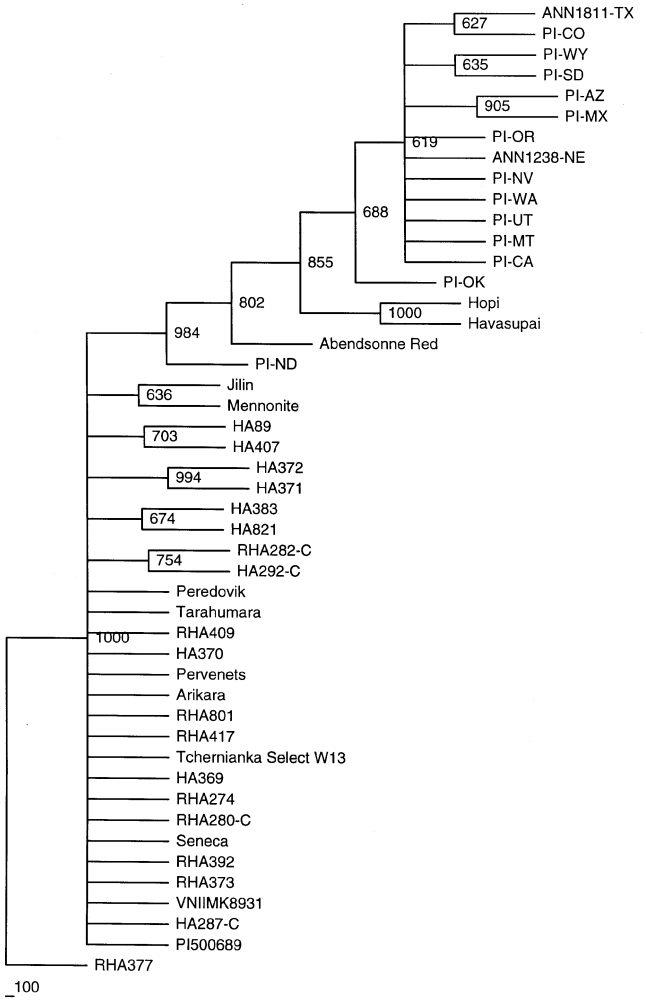

Fig. 6 Consensus tree produced from 1,000 UPGMA trees estimated from bootstrap estimates of the genetic distance matrix (D_{PS}) among 47 germplasm accessions

absence of a correlation between geographic origin and genetic distance (D_{PS}) in the present study might have been a consequence of the hypervariable nature of microsatellite loci, lack of intrapopulation sampling, the specific sample of wild populations drawn, statistical properties of D_{PS}, or a combination thereof (Goldstein et al. 1995a, b; Takezaki and Nei 1996).

We performed additional analyses of the wild populations using $(\delta\mu)^2$, a genetic distance estimator described by Goldstein et al. (1995b) for genetic dating and reconstructing phylogenies among distantly related taxa using microsatellite loci with allele distributions fitting the stepwise mutation model (SMM) (Ohta and Kimura 1973; Schlötterer and Tautz 1992; Shriver et al. 1993; Valdes et al. 1993; Goldstein et al. 1995a, b; Takezaki and Nei 1996). D_{PS}, in contrast to $(\delta\mu)^2$, is not weighted for allelic variants produced by sequentially different (stepwise) mutations (Goldstein et al. 1995a, b; Takezaki and Nei 1996). While D_{PS} accurately estimates tree topologies and reconstructs phylogenies among closely related taxa, $(\delta\mu)^2$ seems to more-accurately estimate genetic distances (branch lengths) among distantly related taxa (Goldstein et al. 1995a, b; Takezaki and Nei 1996). Neither estimator performs optimally for every circumstance or estimation problem.

The allele distributions of many microsatellite loci do not fit the SMM (Valdes et al. 1993; Matsuoka et al. 2002a). More than half of the microsatellite loci in the present analysis were found to have allele distributions uncharacteristic of a stepwise mutation process. We identified 56 dinucleotide repeats with SMM allele distributions. Principal component analysis was performed on the $(\delta\mu)^2$ matrix estimated from the 56 locus subset (Fig. 7). The first two principal components accounted for 73.4 and 14.2% of the variance of the genetic distance matrix and uncovered a deep split between four wild populations originating west of the Continental Divide (PI-CA, PI-OR, PI-MX and PI-NV) and the other 11 wild populations, eight originating in the Great Plains and three originating in the western US (PI-AZ, PI-UT and PI-WA). While our analysis lent support to the hypothesis of an east-west split among wild populations

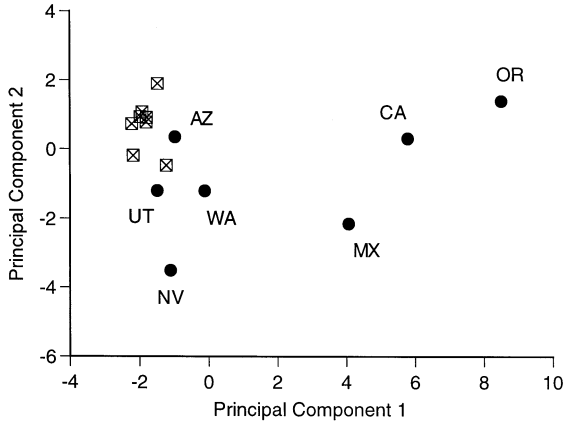

Fig. 7 Principal scores for the first two principal components estimated from the mean $(\delta\mu)^2$ genetic distance matrix estimated from 1,000 bootstrap samples among 15 wild populations of cultivated sunflower originating east (⊠) or west (●) of the Continental Divide in North America

(Cronn et al. 1997), much deeper genetic sampling is needed to assess patterns of genetic diversity in wild sunflowers.

Discussion

Patterns of genetic diversity in domesticated and wild sunflowers

The patterns of genetic diversity uncovered by microsatellites, in hindsight, were predictable but were not uncovered in earlier analyses because of a peculiar scarcity of RAPD and allozyme polymorphisms (Rieseberg and Seiler 1990; Arias and Rieseberg 1995; Cronn et al. 1997). We found a gradual but dramatic narrowing of allelic diversity from one end of the germplasm spectrum (wild populations) to the other (elite oilseed inbred lines). The narrowing was most prominently displayed in the plot of taxon-specific allele numbers (Fig. 3) and the minimum evolution tree (Fig. 5). The genetic distances separating individuals within and between germplasm groups followed a logical pattern, and the germplasm groupings were concordant with historical and pedigree records (where known), e.g., individuals sampled from outbred oilseed culitvars (Peredovik, Pervenets, Tchernianka Select 13 and VNIIMK8931) grouped with oilseed inbred lines, and individuals sampled from outbred confectionery cultivars (Mennonite and Jilin) grouped with, or proximal to, confectionery inbred lines (Figs. 4–5). Because of the lack of intrapopulation sampling, the local spatial arrangements of outbred germplasm accessions in the tree (Fig. 5) and PCA plot (Fig. 4) were only approximate, and the scope of inference within clades was restricted to the individuals sampled (from the outbred germplasm accessions). However, the trends were clearcut and unequivocally showed that wild populations are a wellspring of genetic diversity (Figs. 3–5).

Some of the alleles identified as unique (taxon-specific) (Fig. 3) might be shared by one or more of the outbred populations sampled, and thus not be unique (when compared across the 47 germplasm accessions). Conversely, the unique allele counts were produced by comparing alleles found in one germplasm accession against alleles found in the other 46, and hence were conservative estimates of the number of alleles found in exotic and wild germplasm accessions that are not found in elite inbred lines (intrapopulation sampling was not needed to identify most of the alleles found in inbred lines). The present study only scratched the surface of the allelic diversity present in wild populations. Sampling within populations might identify additional alleles not found in elite inbred lines, in addition to identifying alleles shared by other outbred populations. The number of unique alleles identified in wild populations was extraordinary (Fig. 3) and should increase when additional individuals and wild populations are sampled. The impetus for such an analysis was minimal prior to the microsatellite analysis presented here.

Microsatellites uncover the possibility of multiple domestication origins

Genome-wide microsatellite genotyping has uncovered for the first time the possibility of multiple domestication origins in sunflower. Native American land races of the southwestern US (e.g., Hopi and Havasupai) were found to be distantly related to land races of the Great Plains and eastern US (e.g., Seneca and Arikara). The land races formed two distinct clades (Figs. 4–5) and do not seem to be members of a single monophyletic group (Fig. 6), as had been previously reported (Rieseberg and Seiler 1990; Arias and Rieseberg 1995; Cronn et al. 1997; Seiler and Rieseberg 1997). We cannot rule out the possibility that the eastern and western land races originated from an as yet undiscovered common ancestor and have since diverged through outcrossing, selection, migration, and other domestication forces. The single ancestor hypothesis, however, seems improbable because land races from the two groups were separated by genetic distances ranging from 0.714 to 0.798. Hopi and Havasupai may have originated from hybrids between the first domesticated (monocephalic) sunflowers (the common ancestor) and wild sunflowers indigenous to the desert southwest (e.g., recurring outcrosses to wild sunflowers over hundreds of years coupled with constant selection for monocephaly).

Because human exploration and colonization of North America by Europeans progressed east to west, land races originating east of the Continental Divide were logical candidates for founders of ‘recently domesticated’ sunflowers, e.g., Mennonite, Tarahumara and other ‘confectionery’ cultivars originating in the last 300 years. Seneca and Arikara (ancient domesticates) were

found to be most closely related to open-pollinated confectionery cultivars (Mennonite, Jilin, Tarahumara and PI5000689) and present-day confectionery inbred lines developed over the last quarter century (Cheres and Knapp 1998). Native American land races originating east of the Continental Divide (e.g., Seneca and Arikara) were probably key founders of early European confectionery cultivars and important components of early to mid-twentieth century oilseed sunflower breeding programs.

The uniqueness of the Hopi and Havasupai gene pool was postulated by Heiser (1976) who stressed the importance of preserving the genetic wealth of Native American land races: "Today a few Indian groups still grow their original strains of sunflower – the Hopis and Havasupais in Arizona, the Mandans and Arikaras of North Dakota, and a few Iroquoian survivors in New York and Canada. These original Indian varieties are on the verge of extinction. Not only do they have considerable historical value but, equally important, they are quite different from our modern varieties of sunflower and hence might contain valuable germ plasm for incorporation into our varieties to develop superior plants." Several Native American land races have since been collected by conservationists in the US (e.g., Charles Heiser, Gary Nabhan, Gerald Seiler and others) and carefully preserved by the United States Department of Agriculture National Plant Germplasm System (http://www.ars-grin.usda.gov). While the Great Plains and eastern US land races we sampled seem to be close in heritage to confectionery sunflowers, Hopi and Havasupai do indeed seem to be sources of novel alleles for enhancing present-day sunflowers (Figs. 3–5). Wild sunflowers, however, seem to be a much richer source of novel alleles than the Native American land races (Fig. 3).

Lentz et al. (2001a, b) argued for a single domestication of sunflower and speculated that the probable founders of domesticated sunflowers originated in Mexico, not the eastern US, as had been proposed earlier (Heiser 1985). Heiser (2001) disputed the hypothesis of a single domestication in Mexico (Lentz et al. 2001a, b) and, on the basis of archaeological data from Middle Tennessee (Crites 1993), suggested the possibility of multiple domestication origins, a possibility supported by the analysis presented here (Figs. 4–6). Moreover, Heiser (2001) pointed out that there are "no early historical references to the sunflower in Mexico", an improbability if sunflowers had only been domesticated in Mexico.

Regardless of the number of independent domestications, probable wild progenitors to domesticated sunflowers have not yet been discovered. Microsatellites might be the tool needed for such a discovery. Our data shed no light on the hypothesis of a domestication origin in Mexico (Heiser 1985). We only sampled one wild population from Mexico (PI413123), originally collected from a site far north of the San Andrés archaeological site (Lentz et al. 2001a), and found strong discontinuities between wild and domesticated sunflowers spanning different eras and geographies. One potential pitfall in the search for progenitors is distinguishing between true wild progenitors and hybrids between modern-day cultivars and non-progenitor wild populations (Arias and Rieseberg 1994; Linder et al. 1996), a problem restricted to wild populations exposed to commercial and garden sunflowers. PI-ND is an excellent example (Figs. 4–5). Much deeper genetic sampling than that presented here is needed to produce insights into possible progenitors of ancient domesticates and more fully test the hypothesis of multiple domestication origins.

The power of microsatellite genotyping

Because many microsatellite markers amplify null alleles or multiple loci in sunflower (Tang et al. 2002; Yu et al. 2002a), we sought to identify more than 100 robust, polymorphic, codominant, single-locus microsatellite markers for discriminating between genotypes (identifying individuals, inbred lines and populations) and performing analyses of molecular genetic diversity across genetically and geographically diverse sunflowers. The mean heterozygosity of the selected set was 0.74 among the 47 germplasm accessions (Table 2). The selected set was drawn from the first 600 publicly released microsatellite markers (Tang et al. 2002; Yu et al. 2002a). The number of microsatellite markers developed for sunflower has doubled since the present study was undertaken, and several hundred microsatellite marker loci have been mapped (Burke et al. 2002; Tang et al. 2002; Yu et al. 2002b). Most of the microsatellite markers genotyped in the present study belong to a dense framework of 300 single-locus microsatellite markers selected for genome-wide screening and PCR-multiplexing in sunflower (unpublished data).

The power of microsatellite genotyping for discriminating between closely related individuals is unparalleled and has shed new light on long-standing evolutionary and phylogenetic questions in diverse taxa (Bowcock et al. 1994; Matsuoka et al. 2002b). The power comes from two sources, the hypervariability of microsatellites and the capacity to genotype a large number of individuals for a large number of loci by multiplexing. Both factors formed the basis for a re-analysis of maize domestication (Matsuoka et al. 2002a, b) and the re-analysis of molecular genetic diversity in sunflower presented here. High-throughput microsatellite genotyping enabled Matsuoka et al. (2002b) to screen 193 potential progenitors of domesticated maize with 99 microsatellite marker loci and yielded data challenging the long-standing hypothesis of multiple domestication origins. Similarly, where other marker systems either failed or fell short in humans, hypervariable microsatellites produced new insights into genetic and demographic relationships among closely and distantly related populations (Bowcock et al. 1994; Goldstein et al. 1995a, b; Capelli et al. 2001). We presented a parallel in sunflower where molecular markers other than microsatellites had either

failed to uncover the possibility of multiple domestication origins or were never tested, as was true for restriction fragment length polymorphisms (RFLPs).

Because RFLP and microsatellite markers seem to be equally polymorphic among elite inbred lines in sunflower (Berry et al. 1994; Gentzbittel et al. 1994; Yu et al. 2002a), we speculate that the findings reported here (the possibility of multiple domestication origins and the presence of extraordinary allelic diversity in wild populations) possibly could have been discovered by RFLP genotyping. The point is moot because RFLP markers went untested on Native American land races and wild populations.

Predictably, microsatellites were found to be extraordinarily polymorphic in early domesticates and wild populations, especially the latter, and progressively less polymorphic in genes pools produced by successive cycles of domestication and breeding (Figs. 3–5). Based on the number of taxon-specific alleles identified in anciently domesticated and wild sunflowers (489) and recently domesticated sunflowers (confectionery and oilseed inbred lines) (15), domestication has reduced the number of taxon-specific alleles per germplasm accession nearly 33-fold (Fig. 3).

The extraordinary allelic diversity of microsatellites uncovered in wild sunflowers was somewhat unforeseen, although Cronn et al. (1997) had clearly shown that domesticated and wild sunflowers formed two partially intersecting, nearly distinct groups. Based on the microsatellite loci we sampled (Table 2) and the allozyme loci sampled by Cronn et al. (1997), the number of alleles per locus was found to be two- to four-fold greater for microsatellites than allozymes among domesticated germplasm accessions and six-fold greater for microsatellites than allozymes among wild populations. Similarly, Rieseberg and Seiler (1990) and Cronn et al. (1997) reported mean heterozygosities for allozymes ranging from 0.06 to 0.19 for "domesticated accessions" and 0.10 to 0.27 for "wild accessions". Mean heterozygosities for the microsatellites in the present study (Table 2 and Fig. 2) were three- to eight-fold greater than has been reported for allozymes across different germplasm groups, thereby highlighting stark differences in the mutation rates of microsatellites in non-coding regions versus coding regions sampled by allozymes (functionally important loci) (Schlötterer and Tautz 1992; Chakraborty et al. 1997). Finally, genetic distances separating germplasm accessions were phenomenal in the present study, ranging from 0.252 for HA371–HA372 to 0.945 for RHA280–ANN1811 (Figs. 4–6). By contrast, genetic distances ranged from 0.02 to 0.16 in the allozyme analysis of Rieseberg and Seiler (1990), 0.11 to 0.21 in the allozyme analysis of Cronn et al. (1997), and 0.00 to 0.02 in the RAPD analysis of Arias and Rieseberg (1995). Thus, the shortest genetic distance estimated from microsatellites was greater than the longest genetic distance estimated from allozyme or RAPD markers.

Wild populations as a resource for constructing a dense reference map for cultivated sunflower

The development of 'complete' molecular genetic linkage maps for sunflower from individual crosses in the pre-microsatellite era was challenging, as witnessed by the development of only one *individual* RFLP map with 17 linkage groups (Berry et al. 1995; Gentzbittel et al. 1995; Jan et al. 1998). The first microsatellite map developed for sunflower from an individual cross was complete (coalesced into 17 linkage groups) (Tang et al. 2002), and several complete or nearly complete maps have since been developed using elite × elite or elite × wild crosses (Burke et al. 2002; Yu et al. 2002b).

While past genetic mapping has primarily focused on progeny from elite × elite crosses, our data shows that the densest possible *individual* maps can be produced by using progeny from wild × elite or wild × wild crosses. This conclusion may seem intuitively obvious; however, early work in sunflower suggested that a minimum could be gained by using elite × wild as opposed to elite × elite crosses in sunflower. Moreover, elite × wild crosses nearly always necessitate the use of self-incompatible outbred individuals from wild populations, thereby negating the use of F_2 and other inbred progenies, e.g., recombinant inbred lines. Without the prospect of substantial gains in DNA polymorphism rates, wild × elite crosses have not been the focal point of genetic linkage map development (Rieseberg et al. 1993; Berry et al. 1995; Gentzbittel et al. 1995; Jan et al. 1998; Rieseberg 1998; Knapp et al. 2001). More importantly, maps constructed using progeny from elite × elite crosses identify the subset of molecular markers that tend to be most polymorphic in the elite gene pool and thus have the greatest utility for molecular breeding. That aside, our analysis shows that wild × elite crosses have great promise for increasing the density of the molecular genetic linkage map of sunflower.

Acknowledgements This research was funded by a grant from the USDA-NRICGP Plant Genome Program (#98-35300-6166) to S.J.K. The authors are grateful to Loren Rieseberg for many helpful suggestions and insights into the domestication of sunflower. Oregon Agric. Exp. Stat. Technical Paper No. 11,918.

References

Arias DM, Rieseberg LH (1994) Gene flow between cultivated and wild sunflowers. Theor Appl Genet 89:655–660

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

Berry ST, Allen RJ, Barnes SR, Caligari PDS (1994) Molecular marker analysis of *Helianthus annuus* L. 1. Restriction fragment length polymorphism between inbred lines of cultivated sunflower. Theor Appl Genet 89:435–441

Berry ST, Leon AJ, Hanfrey CC, Challis P, Burkholz A, Barnes S, 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

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorhic microsatellites. Nature 368:455–457
Burke J, Tang S, Knapp SJ, Rieseberg LH (2002) Genetic analysis of sunflower domestication. Genetics 161:1257–1267
Capelli C, Wilson JF, Richards M, Stumpf MPH, Gratix F, Oppenheimer S, Underhill P, Pascali VL, Ko TM, Goldstein DB (2001) A predominantly indigenous paternal heritage for the Austronesian-speaking peoples of insular Southeast Asia and Oceania. Am J Hum Genet 68:432–443
Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetra-nucleotide microsatellite loci. Proc Natl Acad Sci USA 94:1041–1046
Cheres MT, Knapp SJ (1998) Ancestral origins and genetic diversity of cultivated sunflower: coancestry analysis of public germplasm. Crop Sci 38:1476–1482
Crites GD (1993) Domesticated sunflower in fifth millenium B.P. termporal context: new evidence from Middle Tennessee. Am Antiquity 58:146–158
Cronn R, Brothers M, Klier K, Bretting PK, Wendel JF (1997) Allozyme variation in domesticated annul sunflower and its wild relatives. Theor Appl Genet 95:532–545
Gentzbittel L, Perrault A, Nicolas P (1992) Molecular phylogeny of the *Helianthus* genus, based on nuclear restriction-fragment-length polymorphism (RFLP). Mol Biol Evol 9:872–892
Gentzbittel L, Zhang YX, Vear F, Griveau B, Nicolas P (1994) RFLP studies of genetic relationships among inbred lines of the cultivated sunflower, *Helianthus annuus* L.: evidence for distinct restorer and maintainer germplasm pools. Theor Appl Genet 89:419–425
Gentzbittel L, Vear F, Zhang YX, Berville A (1995) Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus* L.). Theor Appl Genet 90:1079–1086
Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995a) An evaluation of genetic distances for use with microsatellite loci. Genetics 139:463–471
Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW (1995b) Genetic absolute dating based on microsatellites and the origin of modern humans. Proc Natl Acad Sci USA 92: 6723–6727
Heiser CB (1945) The Hopi sunflower. Miss Bot Garden Bull 3:163–166
Heiser CB (1946) A "new" cultivated sunflower from Mexico. Madrono 8:226–229
Heiser CB (1951) The sunflower among the North American Indians. Proc Am Phil Soc 95:432–448
Heiser CB (1955) Origin and development of the cultivated sunflower. Am Biol Teach 17:161–167
Heiser CB (1976) The sunflower. University Oklahoma Press, Norman
Heiser CB (1985) Some botanical considerations of the early domesticated plants north of Mexico. In: Ford RI (ed) Prehistoric food production in North America. University Michigan Museum Anthropology, Ann Arbor, pp 57–82
Heiser CB (2001) Letter to the editor. Econ Bot 55:470–471
Hongtrakul V, Huestis GM, Knapp SJ (1997) Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasms: genetic diversity among oilseed inbred lines. Theor Appl Genet 95:400–407
Jan CC, Vick BA, Miller JF, Kahler AL, Butler ET (1998) Construction of an RFLP linkage map for cultivated sunflower. Theor Appl Genet 96:15–22
Kinman ML (1970) New developments in the USDA and state experiment station breeding programs. In: Proc 4th Int Sunflower Conf, Int Sunflower Assoc, Paris, pp 181–183
Knapp SJ, Berry ST, Rieseberg LH (2001) Genetic mapping in sunflower. In: Philips RL, Vasil IK (eds) DNA markers in plants. Kluwer, Dordrecht, The Netherlands, pp 379–403
Korell M, Mosges G, Friedt W (1992) Construction of a sunflower pedigree map. Helia 15:7–12
Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245
Lawson WR, Henry RJ, Kochman JK, Kong GA (1994) Genetic diversity in sunflower (*Helianthus annuus* L.) as revealed by random amplified polymorphic DNA analysis. Aust J Agric Res 45:1319–1327
Leclercq P (1969) Une Sterilite male cytoplasmique chez le tournesol. Ann Amélior Plant 19:99–106
Lentz DL, Pohl MED, Pope KO, Wyatt AR (2001a) Prehistoric sunflower (*Helianthus annuus* L.) domestication in Mexico. Econ Bot 55:370–376
Lentz DL, Pohl MED, Pope KO, Wyatt AR (2001b) Letter to the editor. Econ Bot 55:471–472
Linder CR, Taha I, Seiler GJ, Snow AA, Rieseberg LH (1996) Long-term introgression of crop genes into wild sunflower populations. Theor Appl Genet 96:339–347
Margush T, McMorris FR (1981) Consensus *n*-trees. Bull Math Biol 43:239–244
Matsuoka Y, Mitchell SE, Kresovich S, Goodman MM, Doebley J (2002a) Microsatellites in *Zea* – variability, patterns of mutations, and use for evolutionary studies. Theor Appl Genet 104:436–450
Matsuoka Y, Vigouroux Y, Goodman MM, Sanchez GJ, Buckler E, Doebley J (2002b) A single domestication for maize shown by multilocus microsatellite genotyping. Proc Natl Acad Sci USA 99:6080–6084
Nabhan G (1982) Sunflower of Indians of the southwest. Sunflower 8:30–32
Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York
Ohta T, Kimura K (1973) The model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a genetic population. Genet Res 22:201–204
Ott J (1991) Analysis of human genetic linkage. John Hopkins University Press, Baltimore
Paniego N, Echaide M, Munoz M, Fernandez L, Torales S, Faccio P, Fuxan I, Carrera M, Zandomeni R, Suarez EY, Hopp HE (2002) Microsatellite isolation and characterization in sunflower (*Helianthus annuus* L.). Genome 45:34–43
Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparision of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol Breed 2:225–238
Pustovoit VS (1964) Conclusions of work on the selection and seed production of sunflowers (in Russian). Agrobiology 5:662–697
Putt ER (1997) Early history of sunflower. In: Schneiter AA (ed) Sunflower technology and production. Crop Sci Soc Am, Madison, Wisconsin, pp 1–19
Quillet MC, Vear F, Branlard G (1992) The use of isozyme polymorphism for identification of (*Helianthus annuus* L.) inbred lines. J Genet Breed 91:1195–1202
Rieseberg LH (1998) Genetic mapping as a tool for studying speciation. In: Soltis DE, Soltis PS, Doyle JJ (eds) Molecular systematics of plants. Chapman and Hall, New York, pp 459–487
Rieseberg LH, Seiler GJ (1990) Molecular evidence and the origin and development of the domesticated sunflower (*Helianthus annuus*, Asteraceae.). Econ Bot 44:79–91
Rieseberg LH, Choi H, Chan R, Spore C (1993) Genomic map of a diploid hybrid species. Heredity 70:285–293
Rogers CE, Thompson TE, Seiler GJ (1982) Sunflower species of the United States. National Sunflower Association, Bismark, North Dakota
Rzhetsky A, Nei M (1992a) A simple method for estimating and testing minimum evolution trees. Mol Biol Evol 9:945–967
Rzhetsky A, Nei M (1992b) Statistical properties of the ordinary least-squares, generalized least-squares, and minimum-evolution methods of phylogenetic inference. J Mol Evol 35:367–375
Schlötterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. Nucleic Acids Res 20:211–215
Shriver MD, Jin L, Chakraborty R, Boerwinkle E (1993) VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. Genetics 134:983–993
Seiler GJ, Rieseberg LH (1997) Systematics, origin, and germplasm resources of the wild and domesticated sunflower. In:

Schneiter AA (ed) Sunflower technology and production. Crop Sci Soc Am, Madison, Wisconsin, pp 21–65

Semelczi-Kovacs A (1975) Acclimatization and dissemination of the sunflower in Europe (in German). Acta Ethnogr Acad Sci Hung 24:47–88

Takezaki N, Nei M (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite data. Genetics 144:389–399

Takahashi K, Nei M (2000) Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. Mol Biol Evol 17:1251–1258

Tang S, Yu JK, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. Theor Appl Genet (in press)

Tersac M, Blanchard P, Brunel D, Vincourt P (1994) Relations between heterosis and enzymatic polymorphism in populations of cultivated sunflowers (*Helianthus annuus* L.). Theor Appl Genet 88:49–55

Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133:737–749

Webb DM, Knapp SJ (1990) DNA extraction from a previously recalcitrant plant genus. Mol Biol Rep 8:180–185

Whiting AE (1939) Ethnobotany of the Hopi. Mus Northern Ariz Bull 15

Yu JK, Mangor J, Thompson L, Edwards KJ, Slabaugh MB, Knapp SJ (2002a) Allelic diversity of simple sequence repeat markers among elite inbred lines in cultivated sunflower. Genome 45:652–660

Yu JK, Tang S, Slabaugh MB, Heesacker A, Cole G, Herring M, Soper J, Han F, Webb DM, Thompson L, Edwards KJ, Berry S, Leon A, Olungu C, Maes N, Knapp SJ (2002b) Towards a saturated molecular genetic linkage map for sunflower. Crop Sci (in press)