Aust. J. Biol. Sci., 1986, **39**, 255–70

Genetic Structure of Natural Populations of Wild Sunflowers (*Helianthus annuus* L.) in Australia

P. J. Dry and J. J. Burdon

Division of Plant Industry, CSIRO, G.P.O. Box 1600, Canberra, A.C.T. 2601.

Abstract

The genetic structure of 11 wild populations of *H. annuus* occurring in New South Wales and Queensland was determined by isozyme analysis. Considerable isozyme diversity was found among loci within and between populations, with three to five alleles being identified at each of 10 loci. Mean levels of heterozygosity ranged from 0·19 to 0·38 and gene diversity values from 0·29 to 0·52. In all populations Wright's fixation indices were positive (0·09–0·51) suggesting a degree of inbreeding. Differences in the level of genetic differentiation between populations were not correlated with geographic distance. Indeed, notable genetic diversity was detected between six sites occurring within a 2-km radius of Gunnedah, N.S.W., where the genetic distance relationships were $D = 0{\cdot}13 \pm 0{\cdot}08$, the same as those between populations throughout the region.

There was statistically significant heterogeneity in gene frequency differences between alleles among populations at all loci ($115 < \chi^2 < 254$, d.f. = 30, $P < 0{\cdot}001$). The significant diversity found in these populations suggests that there is considerable scope to utilize wild *H. annuus* in the improvement of commercial sunflowers.

Introduction

Helianthus annuus (sunflower), a native of the United States, was probably introduced and grown in Australia as an ornamental before its value as a crop was recognized and it is likely that populations of wild sunflowers in Australia are derived from this source. As a cultivated crop, *Helianthus annuus* L. was introduced to Australia before 1896 when it was grown as a source of poultry feed. It was not until the 1940s that sunflower received interest as an oilseed crop and today, it is annually grown for this use over 300 000 hectares in New South Wales and Queensland alone.

In many areas of eastern Australia cultivated *H. annuus* crops grow close to populations of wild sunflowers. The latter plants although taxonomically the same species as their cultivated relatives are morphologically quite distinct. Typical wild *H. annuus* plants have multiple branching stems, small leaves and multiple, small capitula. Cultivated *H. annuus* plants, on the other hand, are large-leaved, single-stemmed individuals with a single, large capitulum. Sunflowers have a propensity for outcrossing, self-incompatibility and pollination by bees (Cardon 1922; Heiser 1976). The potential usefulness of wild populations of *H. annuus* as sources of useful germplasm for crop improvement has been assessed in North America for rust resistance (Putt and Sackston 1957), for fertility restoration, and variability for plant

0004-9417/86/030255$02.00

and seed characteristics (Fick *et al.* 1974) and, in Australia, for Rutherglen bug resistance (Downes and Tonnet 1982).

The present investigation was concerned with obtaining a measure of the overall genetic diversity present in wild sunflower populations and how this is distributed

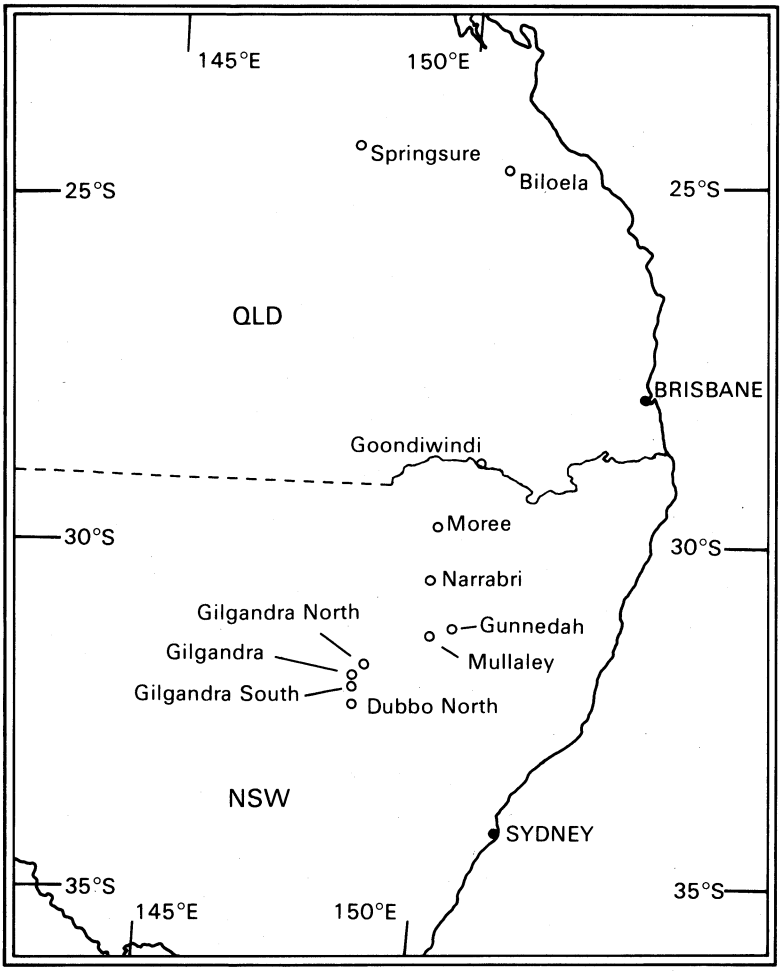

Fig. 1. Geographic distribution of wild *H. annuus* populations of eastern Australia examined in the electrophoretic survey.

within and between populations growing in eastern Australia. In addition the genetic variation between sites of *H. annuus* at a single geographic location was assessed. These questions were addressed through an analysis of electrophoretic variation which

enabled comparison between the geographic distance between populations and their genetic relationships.

Materials and Methods

Population Sampling

Collections of seeds of *H. annuus* were made in the autumn of 1984 from 11 wild populations distributed from central New South Wales to central Queensland. These populations were at Biloela, Dubbo North, Gilgandra, Gilgandra North, Gilgandra South, Goondiwindi, Gunnedah, Mullaley, Moree, Narrabri and Springsure (Fig. 1). Population sizes ranged from 30 to >500 individuals. At Gunnedah, N.S.W., six separate collections of seed were made. Each of these collections was 1 km or more apart. These populations were considered separately in the local geographic analysis but bulked to form a composite population in the regional geographic study.

At each population, a random sample of mature capitula were collected from 12 to 70 individual plants and bagged separately. All individuals sampled had a wild *H. annuus* morphology. Occasional individuals with a cultivated *H. annuus* morphology were not sampled.

Table 1. Details of the electrophoretic systems used in this survey of isozymic variation in wild *H. annuus*

Symbol	EC. No.	Enzyme	Buffer system[A]	Migration distance of all bands (cm)	Heterozygote (bands)
ACP	3.1.3.2	Acid phosphatase	B	4·8,4·4,3·8,3·2	3+conformer
ACO	4.2.1.3	Aconitate hydratase	A	5·7,5·4,5·0,4·5	2
ADH2	1.1.1.1	Alcohol dehydrogenase	B	6·0,5·6,5·1,4·6	3+intergenic
LAP1	3.4.11.1	Cytosol aminopeptidase	B	6·5,6·4,6·2,6·0,5·8	2
LAP2	3.4.11.1	Microsomal aminopeptidase	B	5·4,5·2,5·0,4·7	2
EST	3.1.1.2	Arylesterase	B	7·8,7·6,7·5,7·2,6·8	2+2 conformers
GOT	2.6.1.1	Aspartate aminotransferase	C	3·0,2·5,2·0	3
PGI2	5.3.1.9	Glucose-6-phosphate isomerase	C	4·6,4·0,3·3	3
GDH	1.4.1.2	Glutamate dehydrogenase	B	2·0,1·8,1·4,1·0	5
PGM2	5.4.2.2	Phosphoglucomutase	A	6·8,6·0,5·9,5·4	2+1 conformer

[A] System A: electrode buffer—0·4 M sodium citrate, pH 8·0; gel buffer—5·0 mM histidine, pH 8·0.
System B: electrode buffer—0·3 M borate, 0·1 M sodium hydroxide; gel buffer—3·0 mM citrate, 15·2 mM Tris.HCl.
System C: electrode buffer—75·0 mM lithium hydroxide, titrated to pH 8·5 with boric acid; gel buffer—6% electrode buffer, 94% 9·1 mM citrate, 65 mM Tris.HCl.

Electrophoretic Analysis

One seed taken from each head was pricked and placed in the dark for 20 h on moistened filter paper. The pericarp was removed and the contents were crushed using a Perspex rod in 0·5 ml of 0·05 M phosphate buffer, pH 7·0, containing 1 mg ml^{-1} dithiothreitol. The extract was absorbed on paper chromatography wicks (6 by 5 mm). The wicks were inserted in slots in a horizontal 11% (w/v) starch gel, and electrophoresis was carried out in one continuous and two discontinuous systems (see Table 1 for details). In each gel two controls were included to ensure consistent scoring of alleles. In the continuous system, electrophoresis was conducted for 5 h and in the discontinuous systems electrophoresis was carried out until the front had migrated 10 cm from the sample slot. Each gel was then cut horizontally into three slices and the anodal portion of the gel was assayed for the range of enzymes described in Table 1. The staining procedures were similar to those described by Brewer and Sing (1970) and Brown *et al.* (1978).

Data Collection and Analysis

Nine enzyme systems detected in *H. annuus* (Table 1) were selected from a survey of systems on the basis of the clarity, reproducibility and resolution of zones of enzymatic activity. In scoring isozyme

Table 2. Allele frequencies and sample size, *N*, for loci surveyed in populations of *H. annuus*

Enzyme locus	Allele	Populations[A]											$\bar{p}$	±s.d.
		DN	GlS	Gl	GlN	M	Gu[B]	Na	Mo	Go	Sp	Bi		
		N: 13	12	39	22	16	69	18	23	25	28	14		
Acp	a	0·29	0·08	0·28	0·12	–	0·09	0·28	0·22	0·12	0·13	0·07	0·15	0·05
	b	0·33	0·08	0·46	0·53	0·16	0·52	0·47	0·63	0·42	0·64	0·75	0·45	0·07
	c	0·25	0·67	0·19	0·15	0·53	0·24	0·22	0·11	0·03	0·14	0·18	0·27	0·07
	d	0·13	0·17	0·07	0·20	0·31	0·13	0·03	0·04	0·16	0·09	–	0·12	0·05
Aco	a	–	–	–	0·12	–	–	–	–	–	0·02	–	0·01	0·01
	b	0·50	0·92	0·52	0·60	0·97	0·73	0·61	0·61	0·84	0·73	0·43	0·67	0·07
	c	–	–	0·04	0·02	–	–	–	–	0·02	–	0·11	0·02	0·02
	d	0.50	0·08	0·44	0·26	0·03	0·27	0·39	0·39	0·14	0·25	0·46	0·29	0·07
Adh2	a	0·15	0·63	0·40	0·39	0·22	0·53	0·34	0·33	0·34	0·29	0·75	0·40	0·07
	b	0·69	0·37	0·50	0·61	0·78	0·36	0·63	0·67	0·66	0·64	0·25	0·56	0·07
	c	0·04	–	–	–	–	0·10	0·03	–	–	0·07	–	0·02	0·02
	d	0·11	–	0·10	–	–	–	–	–	–	–	–	0·02	0·02
Lap1	a	–	–	–	–	0·09	–	–	–	0·12	–	–	0·02	0·02
	b	0·15	0·08	0·08	0·59	0·53	0·18	0·26	0·09	0·40	0·05	0·38	0·25	0·02
	c	0·54	0·42	0·38	0·32	0·29	0·71	0·21	0·52	0·40	0·67	0·62	0·46	0·07
	d	0·31	0·42	0·54	0·09	0·09	0·11	0·38	0·33	0·16	0·28	–	0·25	0·06
	e	–	0·08	–	–	–	–	0·15	0·06	–	–	–	0·02	0·02
Lap2	a	0·42	0·13	0·03	0·14	0·18	0·08	0·72	0·59	–	–	–	0·21	0·05
	b	0·58	0·58	0·86	0·64	0·41	0·70	0·25	0·30	1·00	0·76	0·82	0·63	0·07
	c	–	0·29	0·11	0·20	0·41	0·22	0·03	0·11	–	0·24	0·18	0·16	0·05

	d	–	–	–	0·02	–	–	–	–	–	–	–	0·00	0·00
Est	a	0·08	–	0·10	0·36	–	0·19	0·08	0·25	0·10	0·07	0·12	0·12	0·04
	b	0·12	–	–	–	–	–	–	–	0·02	–	–	0·01	0·02
	c	0·23	0·79	0·23	0·31	–	0·19	0·47	0·39	0·24	0·11	0·38	0·30	0·07
	d	0·50	0·21	0·64	0·31	0·78	0·57	0·45	0·36	0·58	0·78	0·50	0·51	0·07
	e	–	–	0·03	0·02	0·22	0·04	–	–	0·06	0·04	–	0·04	0·03
Got	a	0·61	0·79	0·74	0·64	1·00	0·91	n.s.[C]	0·67	0·68	0·91	0·93	0·79	0·06
	b	0·04	0·21	0·04	0·08	–	0·02	n.s.	0·20	0·04	0·09	–	0·07	0·04
	c	0·35	–	0·22	0·28	–	0·07	n.s.	0·13	0·28	–	0·07	0·14	0·05
Pgi2	a	–	0·08	–	0·07	–	0·01	0·22	–	–	–	–	0·03	0·03
	b	0·85	0·87	0·67	0·91	1·00	0·73	0·67	0·72	0·96	0·82	0·93	0·83	0·05
	c	0·15	0·04	0·33	0·02	–	0·26	0·11	0·28	0·04	0·18	0·07	0·13	0·04
Gdh	a	–	–	0·05	–	–	–	–	–	–	0·04	–	0·01	0·01
	b	0·15	0·17	0·32	–	–	0·02	0·31	0·09	0·08	0·02	–	0·10	0·04
	c	–	–	0·02	0·09	–	–	–	–	–	–	–	0·01	0·01
	d	0·85	0·83	0·60	0·91	1·00	0·98	0·69	0·91	0·92	0·94	1·00	0·87	0·05
Pgm2	a	0·15	0·25	0·13	0·68	0·78	0·33	0·26	0·07	0·24	0·27	0·25	0·32	0·06
	b	0·04	0·04	0·03	0·05	0·03	0·18	0·03	0·06	–	–	–	0·04	0·03
	c	–	–	0·03	–	–	–	–	–	–	–	–	0·00	0·00
	d	0·81	0·71	0·82	0·27	0·19	0·49	0·71	0·87	0·76	0·73	0·75	0·64	0·07

[A] DN, Dubbo North; GlS, Gilgandra South; Gl, Gilgandra; GlN, Gilgandra North; M, Mullaley; Gu, Gunnedah; Na, Narrabri; Mo, Moree; Go, Goondwindi; Sp, Springsure; Bi, Biloela.

[B] A composite of six local subpopulations.

[C] Not scored.

phenotypes each zone was assumed to represent a single enzyme locus. Loci were designated by arabic numerals beginning with the most anodal (fastest migrating) zone of activity except for *Adh* (Torres 1983) where *Adh2* was most anodal. Alleles at each locus were distinguished from one another by an alphabetic coding. The most anodal allele was designated allele 'a'. All others were lettered in order of their decreasing electrophoretic mobility. In addition to the inclusion of two reference lines of cultivated sunflower on each gel, photographs were taken of stained gels. This enabled comparison between populations and reconfirmation of allelic mobilities. Allele frequencies were determined by counting of genes from the isozyme genotypes.

To provide a measure of the degree of within and between population differentiation, the following population genetic statistics were computed: (i) the observed frequency of heterozygotes per locus (H); (ii) the gene diversity index (h); (iii) genetic contingency χ^2 analysis; (iv) Wright's fixation index (F); and (v) the genetic distance between populations (D).

The frequency of observed heterozygotes per locus, H, was determined in each population and the average heterozygosity for a population, $\bar{H}$, was the average of the H values for all loci in that population.

The gene diversity index, h, which is a measure of evenness of allele frequencies and effectively determines the heterozygosity in alleles at a locus under random mating, was determined for each locus as

$$h = 1 - \Sigma p_i^2,$$

where p_i is the gene frequency of the ith allele at a locus (Nei 1973). The average gene diversity ($\bar{h}$) for a population was calculated as the average of the diversity values for all loci tested in the population. To avoid the systematic bias introduced to the estimate of the diversity index when the sample size is small, an unbiased estimate of the diversity index, $\bar{h}_c$, is (cf. Nei and Roychoudhury 1974):

$$\bar{h}_c = [(1 - \Sigma \bar{p}_i^2)2N]/(2N - 1).$$

Wright's fixation index, F, was estimated for each population (Brown and Weir 1983) as

$$F = 1 - (\bar{H}/\bar{h}),$$

where $\bar{H}$ and $\bar{h}$ are, respectively, the average observed and expected heterozygosities for all loci.

The genic contingency, χ^2, for each locus measures the heterogeneity in gene frequencies between panmictic populations (Workman and Niswander 1970). In general, if there are k alleles at a locus, r populations, and N is the total sample size, then the χ^2 value for the corresponding $r \times k$ contingency table was determined by

$$\chi^2 = 2N(\sum_{i=1}^{k} \sigma_{\bar{p}_i}^2/\bar{p}_i),$$

where $\bar{p}_i$ and $\sigma_{\bar{p}_i}^2$ are the mean and variance of the frequencies of the ith allele. The degrees of freedom (d.f.) for an $r \times k$ contingency table are $(r-1).(k-1)$. The genic contingency χ^2 was computed for each locus and for all loci.

F_{ST} values (Wright 1965, 1978; Nei 1978) measure the amount of differentiation among subpopulations relative to the limiting amount under complete fixation. Each population is treated as a subdivision of the total set of populations. For each allele at each locus and for the locus as a whole the variance components of sampling variance (s.e. q^2), and actual variance ($\sigma^2{}_q$), limiting variance $q_T(1-q_T)$ and the F statistic, F_{ST}, of each population relative to the total was computed. F_{ST} is the ratio of the actual variance to the limiting variance. The among-populations mean square was corrected for sampling error by subtracting the sampling variance, $q(1-q)/2n$ (Wright 1978).

The genetic distance, D, relates to the accumulated number of gene differences per locus between the populations (Nei 1972) and may be defined as $D = -\log_e I$, where I is the normalized identity, or the likelihood of drawing in a sampling event the same allele from two different populations.

Values for F_{ST} for individual loci and genetic distance were determined by using the computer program BIOSYS-1 (Swofford and Selander 1981).

Results

Regional Geographic Population Analysis

Considerable isozyme variation was encountered between individuals both within and between wild populations of *H. annuus*. A single locus was assessed in the

nine enzyme systems scored except for LAP with two loci. Table 2 shows the allele frequencies for each of the 10 loci assessed in each population. Populations have been listed according to their relative geographic proximity. The average gene frequencies for each allele of each locus, $\bar{p}$, and standard deviation (s.d.) for all populations analysed is included. Given the small sample sizes of some populations assessed, alleles with a gene frequency of less than 0·1 represent tentative estimates.

Within a population some loci showed greater genetic variation than others. The *Acp* and *Est* loci had a greater number of alleles with $\bar{x} = 3{\cdot}8 \pm 0{\cdot}4$ and $\bar{x} = 3{\cdot}4 \pm 0{\cdot}9$ respectively compared with *Gdh* present in each population with $\bar{x} = 2{\cdot}1 \pm 0{\cdot}8$ alleles, where $t = 6{\cdot}1$, d.f. = 20, $P < 0{\cdot}001$, and $t = 3{\cdot}7$, d.f. = 20, $P < 0{\cdot}01$ respectively (Table 3). The genetic diversity within a population when measured by allelic richness was greatest in the Gilgandra population, with an average of $3{\cdot}3 \pm 0{\cdot}2$ alleles per locus and lowest in the Biloela and Mullaley populations with an average of $2{\cdot}2 \pm 0{\cdot}3$ alleles per locus (Table 4*b*). This difference was statistically significant ($t = 9{\cdot}8$, d.f. = 10, $P < 0{\cdot}001$).

Table 3. Genetic variation among enzyme loci and geographic distribution of alleles in *H. annuus*

Enzyme locus	Mean No. of alleles per locus ± s.d.	Allele frequency			
		Widespread and common	Widespread and rare	Locally common	Locally rare
Acp	3·8±0·4	b<c, a	d		
Aco	2·6±0·7	b>d		a	c
Adh2	2·3±0·5	b>a		d	c
Lap1	3·2±0·7	c>d			a, e
Lap2	2·5±0·8	b>c		a	d
Est	3·4±0·9	d>c	a, e		
Got	2·3±0·8	a>c	b		
Pgi2	2·2±0·6	b>c		a	
Gdh	2·1±0·8	d>b			a, c
Pgm2	2·8±0·6	d>a	b		c

Geographic distribution of alleles

There was a considerable range in gene frequency (p) for some alleles between populations. Many of these differences were statistically significant. For instance, allele $Lap1^b$ showed a range from 0·08–0·09 in three populations to $p = 0{\cdot}59$ in Gilgandra North ($\chi^2_c = 16{\cdot}6$, $P < 0{\cdot}001$) and $Pgm2^d$ varied from $p = 0{\cdot}19$ in Mullaley, to $p = 0{\cdot}87$ in Moree ($\chi^2_c = 15{\cdot}4$, $P < 0{\cdot}001$).

Overall four basic geographic patterns of distribution of alleles were found (Table 3). For most loci two alleles were widespread and common to most populations; however, one of these was always at a considerably greater frequency. Widespread but rare alleles like $Pgm2^b$, Est^a, Est^e and Acp^d alleles were found in a few individuals in most populations. Alleles detected in many individuals of a few populations (locally common alleles), helped to characterize that population such as the Aco^a and $Pgi2^a$ alleles in Gilgandra and Narrabri respectively.

Some alleles were found in a few individuals in a few populations (locally rare alleles, Table 3). Of these, four, $Lap1^a$, $Lap1^e$, Gdh^a and Gdh^c, were only found

in two populations whilst two alleles *Lap2*^{d} and *Pgm2*^{c} were detected in a single population.

The level of heterozygosity varied amongst loci and populations (Table 4*a*). The maximum heterozygosity was found for the *Est* locus in Gilgandra North where $H = 0{\cdot}59$. Table 4*b* shows that when the evenness of the allele frequencies was assessed by the mean observed heterozygosity, $\bar{H}$, in each population, the maximum value of 0·38 occurred in the Mullaley and Narrabri populations. Reduced heterozygosity was observed in Goondiwindi, $\bar{H} = 0{\cdot}24$, and Gilgandra South, $\bar{H} = 0{\cdot}19$.

Table 4. (*a*) Observed heterozygosity (H) for each enzyme system in wild populations of *H. annuus*. (*b*) Average number of alleles per locus ($\bar{x}$), average observed heterozygosity ($\bar{H}$), corrected diversity index ($\bar{h}_c$) and Wright's fixation index (F) for each population of *H. annuus*

Enzyme locus	Populations[A] DN	GlS	Gl	GlN	M	Gu	Na	Mo	Go	Sp	Bi
					(*a*) Individual locus observed heterozygosity, *H*						
Acp	0·33	0·17	0·44	0·25	0·44	0·65	0·56	0·57	0·40	0·53	0·50
Aco	0·38	0·00	0·51	0·52	0·06	0·19	0·56	0·50	0·16	0·21	0·43
Adh2	0·31	0·42	0·38	0·41	0·31	0·37	0·29	0·30	0·28	0·43	0·36
Lap1	0·23	0·00	0·36	0·27	0·44	0·19	0·39	0·43	0·32	0·22	0·15
Lap2	0·17	0·17	0·11	0·18	0·38	0·17	0·28	0·26	0·00	0·26	0·07
Est	0·31	0·25	0·33	0·59	0·44	0·41	0·44	0·55	0·44	0·15	0·77
Got	0·46	0·25	0·47	0·56	0·00	0·16	—[B]	0·30	0·32	0·11	0·14
Pgi2	0·31	0·25	0·21	0·14	0·00	0·01	0·50	0·30	0·08	0·29	0·14
Gdh	0·00	0·00	0·03	0·00	0·00	0·01	0·06	0·00	0·00	0·04	0·00
Pgm2	0·38	0·42	0·26	0·55	0·31	0·38	0·35	0·17	0·40	0·39	0·50
					(*b*) Individual population, allelic richness, heterozygosity diversity and fixation indices						
$\bar{x}$	2·9	2·7	3·3	3·2	2·2	3·0	3·0	2·8	2·8	2·8	2·2
(±s.d.)	(0·3)	(0·3)	(0·2)	(0·2)	(0·3)	(0·2)	(0·3)	(0·2)	(0·4)	(0·2)	(0·2)
$\bar{H}$	0·29	0·19	0·31	0·35	0·38	0·25	0·38	0·34	0·24	0·25	0·31
$\bar{h}_c$	0·50	0·41	0·49	0·48	0·30	0·44	0·48	0·47	0·38	0·37	0·35
(±s.e.)	(0·05)	(0·05)	(0·04)	(0·06)	(0·09)	(0·06)	(0·06)	(0·05)	(0·08)	(0·04)	(0·06)
F	0·40	0·51	0·34	0·26	0·19	0·43	0·40	0·26	0·35	0·32	0·09

[A] Abbreviations as for Table 2. [B] No data collected.

Allelic diversity

High gene diversity index (h) values ($h > 0{\cdot}65$) were found in six instances (*Acp* in Goondiwindi, Gilgandra and Narrabri and *Est* in Dubbo North, Gilgandra North and Moree populations). The Mullaley population showed the greatest range over the different loci with $h < 0{\cdot}06$ for four loci and $h > 0{\cdot}6$ for three loci.

When all loci were considered in each population to obtain an unbiased estimate of the average of the diversity indices ($\bar{h}_c$), the Dubbo North population showed the greatest gene diversity, $\bar{h}_c = 0{\cdot}50 \pm 0{\cdot}05$. The least genetic diversity was evident in the Mullaley population, $\bar{h}_c = 0{\cdot}30 \pm 0{\cdot}09$.

Wright's fixation index

Wright fixation index values, F, were determined for all loci in each population (Table 4*b*) and represent the standard deviation of the observed frequency from the Hardy–Weinberg expected frequency of heterozygotes. F values may range from

-1 to $+1$ and give a general indication of the extent of inbreeding. The value of F expected under panmixia (random mating; no selection) is zero; an increasing deficiency of heterozygotes give F values that approach $+1$, while excesses of heterozygotes give F values that are negative. The consistently positive F values detected in the wild sunflower populations ($+0{\cdot}09 < F < +0{\cdot}51$) are suggestive of a degree of inbreeding in these populations.

When the four measures of within-population genetic variation (Table 4*b*) were compared between populations allelic richness was greatest in Gilgandra which also had a high gene diversity index similar to Dubbo North, Gilgandra North and Narrabri populations. Gilgandra South had the highest fixation index and the lowest level of observed heterozygosity which suggests there was a deficiency in heterozygotes and some inbreeding in this population. Similarly the Dubbo North population had a high fixation index but also a high diversity index. Of interest was the high fixation index in the Gunnedah and Narrabri populations yet Narrabri showed both the maximum observed and expected levels of heterozygosity. In contrast, Biloela and Mullaley showed the least allelic richness the lowest fixation indices and in addition Mullaley had the lowest gene diversity.

Table 5. Contingency χ^2 analysis of gene frequencies at all loci and F statistics (F_{ST}) among wild *H. annuus* populations
***$P < 0{\cdot}001$

Locus	No. of alleles	χ^2	d.f.	F_{ST}
Acp	4	115·7***	30	0·10
Aco	4	122·9***	30	0·09
Adh2	4	135·5***	30	0·09
Lap1	5	264·9***	40	0·11
Lap2	4	254·0***	30	0·21
Est	5	192·7***	40	0·11
Got	3	99·2***	20	0·21
Pgi2	3	118·1***	20	0·08
Gdh	4	143·0***	30	0·11
Pgm2	4	138·4***	30	0·17
Total		1584·4***	300	$\bar{F}_{ST} = 0{\cdot}13$

Population heterogeneity of gene frequencies

To determine whether there was significant heterogeneity in gene frequency differences between alleles a contingency χ^2 analysis of heterogeneity among populations was completed. For the 10 loci, comparison of gene frequencies was based on the formula of Workman and Niswander (1970), where the genic contingency χ^2 is a function of the total sample size and the means and variances of the gene frequencies. Differences in the gene frequencies among the 11 populations were statistically highly significant at all loci with $P < 0{\cdot}001$ (Table 5). These results show there are significant differences among the gene pools of the wild sunflower populations and suggests considerable regional genetic differentiation in *H. annuus*.

Interpopulation genetic heterogeneity was measured, for all alleles, by estimates of an F statistic, F_{ST}. The F_{ST} values for each locus (Table 5) ranged from 0·08 to 0·21. The *Lap2* and *Got* loci showed the greatest deviation from zero.

The matrix of the coefficients of genetic distance (Nei 1972) is shown in Table 6. Small genetic distance values imply close genetic relationship between two populations and it is evident that a close genetic relationship exists between seven of the populations where $D < 0{\cdot}08$: Biloela, Dubbo North, Gilgandra, Goondiwindi, Gunnedah, Moree and Springsure. The tight cluster of populations which show close genetic relationships are not always related by their geographic distance. The smallest genetic distance occurred between Gunnedah and Springsure, $D = 0{\cdot}03$, yet these are 1000 km apart. Similarly Mullaley, which is geographically closest to Gunnedah, is genetically as closely related to populations at Gilgandra North ($D = 0{\cdot}11$), and Goondiwindi ($D = 0{\cdot}13$), as with Gunnedah ($D = 0{\cdot}13$). The largest genetic distance, $D = 0{\cdot}28$, occurred between Mullaley and Gilgandra and between Biloela and Narrabri populations.

Gilgandra South forms an intermediate genetic relationship with the tight cluster of seven populations outlined above where $0{\cdot}13 < D < 0{\cdot}17$ but the data suggests a considerable degree of differentiation in Gilgandra South. Similarly, Narrabri has a minimum genetic distance value of 0·10 with Moree, its closest geographic neighbour, and shows some genetic relationship with Mullaley $D = 0{\cdot}12$, and Dubbo North, $D = 0{\cdot}13$. However, in general, the Narrabri population is genetically distinct from the other populations.

Table 6. Matrix of the coefficients of genetic distance for 11 wild populations of wild *H. annuus*

Population[A]	DN	GlS	Gl	GlN	M	Gu	Na	Mo	Go	Sp
GlS	0·17	–								
Gl	0·07	0·17	–							
GlN	0·15	0·20	0·19	–						
M	0·22	0·21	0·28	0·11	–					
Gu	0·11	0·13	0·10	0·09	0·13	–				
Na	0·13	0·23	0·21	0·24	0·12	0·27	–			
Mo	0·05	0·17	0·11	0·17	0·28	0·11	0·10	–		
Go	0·08	0·13	0·08	0·08	0·13	0·08	0·22	0·14	–	
Sp	0·08	0·16	0·06	0·13	0·15	0·03	0·25	0·10	0·06	–
Bi	0·13	0·17	0·11	0·11	0·24	0·04	0·28	0·13	0·09	0·07

[A] Abbreviations as in Table 2.

Local Geographic Population Analysis

Allele frequencies at the 10 loci studied at the six sites of *H. annuus* growing at Gunnedah, N.S.W., are given in Table 7. The patterns of variation between loci were similar to those observed in the regional geographic population analysis where the *Est* and *Acp* loci were multiallelic while the *Gdh* and *Aco* loci were diallelic. Over the six sites the EST system showed greatest diversity with $\bar{x} = 3{\cdot}8 \pm 0{\cdot}4$ alleles; however, the *Gdh* locus was fixed for a single allele at many sites ($\bar{x} = 1{\cdot}3 \pm 0{\cdot}5$ alleles), and the Got^a allele was fixed at two sites.

Table 7 shows that there was considerable differentiation in isozymes between sites. A few alleles occurred at high frequency in one or two sites and were absent elsewhere, such as $Adh2^c$ and $Lap2^a$. Rare alleles were $Pgi2^a$ and Gdh^b.

In contrast to the most frequent alleles, $Adh2^a$, Acp^b and Aco^b, occurring in the composite Gunnedah population, $Adh2^c$, Acp^a and Aco^d were the most frequent alleles in site 1. These differences were statistically significant where $\chi^2 = 25{\cdot}4$, $P < 0{\cdot}001$, $\chi^2 = 16{\cdot}3$, $P < 0{\cdot}001$ and $\chi^2 = 7{\cdot}8$, $P < 0{\cdot}005$ respectively.

Table 7. Allele frequencies for loci and sample size, *N*, in six sites of *H. annuus* growing around Gunnedah, N.S.W.

Enzyme locus	Allele	Site 1	2	3	4	5	6	$\bar{p}$	±s.d.
		N: 14	8	12	9	15	11		
Acp	a	0·43	–	–	0·11	–	–	0·09	0·05
	b	0·36	0·43	0·50	0·50	0·70	0·64	0·52	0·10
	c	0·07	0·37	0·46	0·11	0·17	0·27	0·24	0·09
	d	0·14	0·19	0·04	0·17	0·13	0·09	0·13	0·07
Aco	b	0·29	0·94	0·87	0·72	0·70	0·86	0·73	0·08
	d	0·71	0·06	0·12	0·28	0·30	0·14	0·27	0·08
Adh2	a	0·21	0·56	0·50	0·61	0·67	0·64	0·53	0·17
	b	0·21	0·44	0·50	0·39	0·33	0·32	0·36	0·10
	c	0·57	–	–	–	–	0·04	0·10	0·04
Lap1	b	–	0·06	–	0·44	0·40	0·18	0·18	0·07
	c	0·89	0·81	0·75	0·50	0·47	0·82	0·71	0·09
	d	0·11	0·12	0·25	0·05	0·13	–	0·11	0·06
Lap2	a	n.s.[A]	–	–	0·50	–	–	0·08	0·05
	b	n.s.	0·50	0·46	0·50	0·93	0·82	0·70	0·09
	c	n.s.	0·50	0·54	–	0·07	0·18	0·22	0·07
Est	a	0·04	0·19	0·25	0·17	0·13	0·36	0·19	0·08
	c	0·14	0·19	0·17	0·05	0·47	0.14	0·19	0·08
	d	0·75	0·56	0·54	0·72	0·37	0·50	0·57	0·10
	e	0·07	0·06	0·04	0·05	0·03	–	0·04	0·04
Got	a	0·89	0·87	0·83	1·00	1·00	0·86	0·91	0·06
	b	0·03	–	0·08	–	–	–	0·02	0·03
	c	0·07	0·12	0·08	–	–	0·14	0·07	0·05
Pgi2	a	–	–	0·04	–	0·03	–	0·01	0·02
	b	0·43	0·75	0·71	0·83	0·73	0·95	0·73	0·09
	c	0·57	0·25	0·25	0·17	0·23	0·05	0·26	0·08
Gdh	b	0·07	–	0·04	–	–	–	0·02	0·02
	d	0·93	1·00	0·96	1·00	1·00	1·00	0·98	0·02
Pgm2	a	0·68	0·50	0·21	0·22	0·40	–	0·33	0·09
	b	0·11	0·06	0·17	0·22	0·13	0·36	0·18	0·08
	d	0·21	0·44	0·62	0·56	0·47	0·64	0·49	0·10

[A] Not scored.

The observed frequency of heterozygotes for each locus tested in each site is shown in Table 8*a* and the average observed heterozygosity, $\bar{H}$ which ranged from 0·17 to 0·35 between sites, in Table 8*b*. The gene diversity, varied between enzyme systems within a site but there were no significant differences in the average gene diversity indices between sites.

The Wright fixation index (F) showed the largest value at site 1, $F = 0{\cdot}53$, and the least in site 6, $F = 0{\cdot}10$ (Table 8*b*), a comparable range in F values to that observed in the F values of the regional population analysis.

The contingency χ^2 analysis of heterogeneity among the six local sites showed that differences in gene frequencies were statistically highly significant in eight of

the loci, where for six loci, $P < 0{\cdot}001$ (Table 9). Local population heterogeneity was also analysed by an F statistic, F_{ST}, and the values for actual variance in gene frequencies between sites were corrected for sampling error. Estimates of

Table 8. (*a*) Observed heterozygosity (H) for each enzyme system of wild populations of *H. annuus* at Gunnedah, N.S.W. (*b*) Average observed heterozygosity ($\bar{H}$), the corrected diversity indices ($\bar{h}_c \pm$ s.e.) and the fixation index (F) for each site

Enzyme locus	Site 1	2	3	4	5	6
	(*a*) Individual locus observed heterozygosity (H)					
Acp	0·36	0·62	0·50	0·67	0·33	0·73
Aco	0·14	0·12	0·25	0·11	0·20	0·27
Adh2	0·43	0·29	0·33	0·11	0·40	0·54
Lap1	0·00	0·25	0·17	0·44	0·20	0·09
Lap2	n.s.[A]	0·25	0·42	0·11	0·00	0·18
Est	0·21	0·37	0·50	0·44	0·47	0·45
Got	0·07	0·25	0·17	0·00	0·00	0·27
Pgi2	0·00	0·50	0·50	0·33	0·33	0·09
Gdh	0·00	0·00	0·08	0·00	0·00	0·00
Pgm2	0·29	0·37	0·58	0·44	0·33	0·27
	(*b*) Site indices					
$\bar{H}$	0·17	0·31	0·35	0·28	0·23	0·29
$\bar{h}_c$	0·37	0·41	0·43	0·42	0·38	0·34
(± s.e.)	(0·07)	(0·07)	(0·05)	(0·08)	(0·08)	(0·06)
F	0·53	0·18	0·15	0·30	0·38	0·10

[A] Not scored.

Table 9. Contingency χ^2 analysis and F statistics (F_{ST}) at all loci among six *H. annuus* populations at Gunnedah, N.S.W.
$*P < 0{\cdot}025$; $**P < 0{\cdot}01$; $***P < 0{\cdot}001$

Locus	No. of alleles	χ^2	d.f.	F_{ST}
Acp	4	56·9***	15	0·06
Aco	2	33·7***	5	0·20
Adh2	3	68·2***	10	0·10
Lap1	3	38·1***	10	0·10
Lap2	3	78·6***	10	0·26
Est	4	29·1*	15	0·03
Got	3	12·8	10	0·01
Pgi2	3	23·1*	10	0·09
Gdh	2	5·6	5	0·01
Pgm2	3	34·8***	10	0·09
		$\Sigma\chi^2 = 381{\cdot}0$***	100	$\bar{F}_{ST} = 0{\cdot}10$

F_{ST} values for all alleles were used to obtain F_{ST} values for each locus (Table 9). Heterogeneity between sites where $0{\cdot}01 < F_{ST} < 0{\cdot}26$ was greatest at the *Lap2* and *Aco* loci.

The matrix of coefficients of genetic distance (Nei 1972) between the six sites sampled at Gunnedah, N.S.W., are given in the following tabulation:

Site	1	2	3	4	5
2	0·22	–			
3	0·26	0·02	–		
4	0·24	0·10	0·10	–	
5	0·19	0·09	0·10	0·07	–
6	0·26	0·06	0·05	0·07	0·06

The small genetic distance values for sites 2–6, $0{\cdot}02 < D < 0{\cdot}10$, reflect the close genetic relationship of populations between those sites and suggests that these five sites formed a tight cluster. Site 1 was genetically distinct, $0{\cdot}19 < D < 0{\cdot}26$, with other sites.

Discussion

In Australia, wild *H. annuus* show a considerable level of genetic variation both within and between different populations. In this study genetic variability within populations was measured using allelic richness, and Nei's gene diversity index. Most loci were polymorphic, the average number of alleles per locus exceeded two in all populations; the observed level of heterozygosity averaged over 10 isozyme loci ranged from 0·19 to 0·38 (mean value, 0·29) and values for the Nei diversity index (h_c) varied between 0·30 and 0·50.

Comparison of these results with a mean value derived from 36 primarily outcrossing species (Hamrick *et al.* 1979) shows that both the number of alleles per locus and the polymorphic index (equivalent to the heterozygote frequency under Hardy–Weinberg proportions) were markedly greater in the Australian populations of *H. annuus* (number of alleles per locus: 2·20–3·30 versus 1·85; polymorphic index 0·34–0·48 versus 0·19). These populations of *H. annuus* were also genetically more variable than 35 populations of the *H. debilis* complex occurring in their native range in the United States (Wain 1983). In that study the average number of alleles per locus was 1·69.

Between populations of *H. annuus* there was considerable variation in the frequency of different alleles of many loci. In general, alleles could be grouped into the four basic patterns of widespread and common, widespread and rare, locally common and locally rare that Marshall and Brown (1975) proposed when considering sampling strategies. Of all alleles detected 70% were widespread, occurring in five or more populations, of which 54% of these alleles were found in all populations. In contrast less than 20% of alleles were locally restricted to one or two populations; 38% were generally rare ($\bar{p} < 0{\cdot}1$) across all populations and only 13 alleles (33%) occurred with a frequency of greater than 0·3. These results confirm the conclusions of Marshall *et al.* (1981) that maximum variation is likely to be achieved if samples are taken from as many sites as possible over a broad geographic and environmental range.

Clearly because of the small sizes of samples taken from some of the populations studied, the distribution of rare alleles and the level of genetic diversity found within wild *H. annuus* in these populations may vary somewhat from the actual values. However, a comparison of populations with approximately 20 individuals indicates only minor differences in the mean diversity index (0·43 versus 0·39) and no difference in the overall observed levels of heterozygosity. Fewer alleles

per locus were found in the six smaller populations but this difference was not significant (2·98 ± 0·20 versus 2·53 ± 0·35).

For all populations mean estimates of Wright's fixation index were positive, ranging from +0·09 to +0·51. In most populations there was clearly a marked deficiency of heterozygotes compared with that expected under random mating. This result was unexpected as outcrossing is encouraged in *H. annuus* by self-incompatibility and bee pollination (Cardon 1922; Heiser 1976). Factors which might contribute to positive F values include ecological traits like asynchrony of flowering and the tendency for seeds to fall close to the parent plant so that adjacent plants are more likely to be related. These factors would increase the degree of inbreeding. In addition self-incompatibility is not complete since Hamilton (1926) reported that many lines when self-pollinated produced 15–50% as much seed compared with cross-pollination.

A comparison of the overall genetic similarity of the different populations of wild sunflower showed that the degree of genetic relatedness was not associated with the geographic distance between individual populations. For example, three populations sampled in the vicinity of Gilgandra (separated by a maximum of 15 km) showed genetic distance values $0{\cdot}17 < D < 0{\cdot}20$, (Table 7) while other geographically more distant populations like Dubbo North and Goondiwindi (400 km apart) had a genetic distance value of only 0·08. For populations with close genetic relationships it is possible that germplasm of a common origin was introduced at locations geographically disparate which now persists in the wild state. In a study of *H. debilis* in the United States, Wain (1983) found very close genetic distance values between populations of the same species ($D = 0{\cdot}01$). Overall, however, the genetic distance values found in this study are quite large. This suggests the possibility of mutation and selection forces operating which have resulted in a degree of evolutionary divergence between populations. If these populations are effectively isolated, gene pools could diverge rapidly through adaptation to different environments.

The study of genetic variation at a local geographic scale gave a picture similar to that obtained from the regional study. The range in the values of observed levels of heterozygosity (0·17–0·35) and in the corrected gene diversity index (0·34–0·43) was marginally smaller in the six sites of *H. annuus* found around Gunnedah than amongst the 11 widely spread populations. Even in the least diverse of the Gunnedah sites (No. 6) over half of all alleles detected were present.

Unexpectedly, the average genetic distance between individual sites at Gunnedah was the same as that obtained in the broader geographic comparison of populations that were up to 2000 km apart. The reasons for this result appear to lie in the occurrence of a single anomalous population (site 1). With the exception of this site, the genetic distance values for the Gunnedah subpopulations show that these sites (Nos 2–6) are more closely related to one another ($0{\cdot}02 < D < 0{\cdot}10$) than are the more widely dispersed populations of the broad geographic survey ($\bar{D} = 0{\cdot}13$). This result is in general agreement with those obtained for a range of other species (Loveless and Hamrick 1984).

Gunnedah site 1 was, however, genetically quite different to others in the area ($0{\cdot}19 < D < 0{\cdot}26$). The inclusion of this population in the overall local comparison gave rise to the unusual similarity between the local and broad geographic genetic distance values. A possible explanation for the anomalous nature of this site

is found in its location. Site 1 was located close to cattle yards where repeated, frequent seed input is always possible as a result of extensive stock movement.

The source of the high levels of diversity encountered in the wild populations is not entirely clear. Sunflowers have been used commercially in Australia for 90 years; however, current cultivars are less polymorphic than wild *H. annuus* populations (Dry, unpublished data). Thus cultivated sunflowers are likely to have made little contribution to the genetic diversity of the wild populations. Wild *H. annuus* were apparently introduced and grown in Australia for ornamental value and it is likely that the derivatives today are escapes from this source.

Important contributions to genetic differentiation among wild sunflower populations may be made by founder effects, differential selection for particular isozyme phenotypes and variations in migration rate, inbreeding and genetic drift. Many of the populations studied were composed of only 100 individuals and some were as small as 30 individuals. Founder effects could operate following seed migration by animal-mediated dispersal and the resultant small population sizes could represent a restricted and different gene pool to the source population. Evidence for non-random mating (deficiency in heterozygotes) has been suggested from the positive F values found here and in wild populations of *H. annuus* in Kansas, United States (Ellstrand *et al.* 1978). Finally genetic drift or random fluctuation in the frequencies of specific genes in small populations may rapidly lead to marked differences between different populations.

The high levels of genetic variation present in these wild populations of *H. annuus* and the significant degree of genetic differentiation occurring between different populations suggests that they represent a diverse source of germplasm which may prove useful in the agronomic improvement of cultivated sunflower. Further studies (Dry, unpublished data) have already demonstrated the existence of considerable diversity in the response of individual plants to infection by the economically important pathogen *Puccinia helianthi*.

Acknowledgments

We would like to thank Dr R. W. Downes for suggestions of locations of wild sunflower populations, Ms S. Carder and Mrs E. Lewis for technical assistance and Dr A. H. D. Brown and Dr R. W. Downes for comments on the manuscript. P.J.D. was funded by Arthur Yates and Co. Pty Ltd.

References

Brewer, A. F., and Sing, C. F. (1970). 'An Introduction to Isozyme Techniques.' (Academic Press: New York.)

Brown, A. H. D., Nevo, E., Zohary, D., and Dagan, O. (1978). Genetic variation in natural populations of wild barley (*Hordeum spontaneum*). *Genetica* **49**, 97–108.

Brown, A. H. D., and Weir, B. S. (1983). Measuring genetic variability in plant populations. In 'Isozymes in Plant Genetics and Breeding, Part A'. (Eds S. D. Tanksley and T. J. Orton.) pp. 219–39. (Elsevier Science Publishers: Amsterdam.)

Cardon, P. V. (1922). Sunflower studies. *J. Am. Soc. Agron.* **14**, 69–72.

Downes, R. W., and Tonnet, M. L. (1982). Selection in sunflowers for resistance to Rutherglen bug (*Mysius vinitor*). Proc. 10th Int. Sunflower Conf. Surfers Paradise, pp. 261–64.

Ellstrand, N. C., Torres, A. M., and Levin, D. A. (1978). Density and the rate of apparent outcrossing in *Helianthus annuus* (Asteraceae). *Syst. Bot.* **3**, 403–7.

Fick, G. N., Zimmer, D. E., Dominguez-Gimenez, J., and Rehder, D. A. (1974). Fertility restoration and variability for plant and seed characteristics in wild sunflowers. Proc. 6th Int. Sunflower Conf. (Bucharest, Romania). pp. 333–8.

Hamilton, R. I. (1926). Improving sunflowers by inbreeding. *Sci. Agric.* **6**, 190–2.

Hamrick, J. L., Linhart, Y. B., and Mitton, J. B. (1979). Relationships between life history characteristics and electrophoretically detectable genetic variation in plants. *Annu. Rev. Ecol. Syst.* **10**, 173–200.

Heiser, C. B. Jr (1976). 'The Sunflower.' (Univ. Oklahoma Press.)

Loveless, M. D., and Hamrick, J. L. (1984). Ecological determinants of genetic structure in plant populations. *Annu. Rev. Ecol. Syst.* **15**, 65–95.

Marshall, D. R., and Brown, A. H. D. (1975). Optimal sampling strategies in genetic conservation. In 'Crop Genetic Resources for Today and Tomorrow'. (Eds O. H. Frankel and J. G. Hawkes.) pp. 53–70. (Cambridge Univ. Press.)

Marshall, D. R., Burdon, J. J., and Brown, A. H. D. (1981). Optimal sampling stategies in the biological control of weeds. Proc. Fifth Int. Symp. Biol. Control Weeds, Brisbane, Australia, 1980. (Ed. E. S. Del Fosse.) pp. 103–11. (CSIRO: Melbourne.)

Nei, M. (1972). Genetic distance between populations *Am. Nat.* **106**, 283–92.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proc. Natl Acad. Sci. U.S.A.* **70**, 3321–3.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89**, 883–90.

Nei, M., and Roychoudhury, A. K. (1974). Sampling variances of heterozygosity and genetic distances. *Genetics* **76**, 379–90.

Putt, E. D., and Sackston, W. E. (1957). Studies on sunflower rust. I. Some sources of rust resistance. *Can. J. Plant Sci.* **37**, 43–54.

Swofford, D. L., and Selander, R. B. (1981). Biosys-1, a computer program for the analysis of allelic variation in genetics. (University of Illinois at Urbana, Champaign, Illinois, U.S.A.)

Torres, A. M. (1983). Sunflowers (*Helianthus annuus* L.). In 'Isozymes in Plant Genetics and Breeding, Part B'. (Eds S. D. Tanksley and T. J. Orton.) pp. 320–38. (Elsevier Science Publishers: Amsterdam.)

Wain, R. P. (1983). Genetic differentiation during speciation in the *Helianthus debilis* complex. *Evolution* **37**, 1119–27.

Workman, P. L., and Niswander, J. D. (1970). Population studies on south western Indian tribes. II. Local genetic differentiation in the Papago. *Am. J. Hum. Genet.* **22**, 24–49.

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* **19**, 395–420.

Wright, S. (1978). Variability within and among natural populations. In 'Evolution and the Genetics of Populations'. Vol. 4. (University of Chicago Press: Chicago.)

Manuscript received 3 February 1986, accepted 22 May 1986