

A Method for the Efficient Management and Utilization of Large Germplasm Collections

D. Z. Skinner,* G. R. Bauchan, G. Auricht, and S. Hughes

ABSTRACT

To make efficient use of large germplasm collections, it is advisable to assemble a representative core collection and to evaluate the relationships among the traits studied. However, the assemblage of a core collection from very large germplasm collections is problematic. The computing resources needed to carry out genetic distance calculations and comparisons with commonly available programs is prohibitively large. The objects of this study were (i) to develop a method which assembles a core collection by maximizing the diversity (measured as mean Euclidean distance) from within groups of accessions defined by species, subspecies, and geographic origin and (ii) to test the effectiveness of the method on a collection of 20 997 annual *Medicago* accessions from the Australian *Medicago* Resource Center in Adelaide, South Australia, that had been evaluated for 27 agronomic characteristics. The method resulted in a core collection of 1705 accessions that represented 74% of the extremes of the 27 characters, indicating that the entire range of the characters was represented in most cases. Accessions representing the extremes easily could be added to the core collection. The method used requires relatively minor computing resources and should be useful to curators of large germplasm collections. To assess the relationships among the 27 measured traits, correlation coefficients of all possible combinations of traits were calculated. The most strongly associated traits were, as expected, such traits as grams of seed per plant and grams of pods per plant and indicated that some traits could be omitted from future evaluations with little loss of information, thereby increasing the efficiency with which germplasm evaluations can be carried out.

VERY LARGE COLLECTIONS of germplasm from around the world are being assembled in efforts to conserve the extant genetic variation of numerous species. To evaluate and utilize best these collections, it is necessary to identify a smaller subset or 'core' collection that likely represents most of the genetic variation in the entire collection.

On the basis of several statistical models, Brown (1989 a,b) suggested that at least 70% of the alleles present in the entire collection will be represented in a core collection comprised of at least 10% of the accessions. Large increases in core collection size have increasingly marginal effects on the levels of diversity retained (Brown, 1989a). For example, under conditions of variable levels of diversity in a population of 10 000, about 70% of alleles were predicted to be retained in a core comprised of 10% of the accessions, but doubling the number of retained accessions to 20% increased the

predicted diversity retention by only about 5% (Brown, 1989a). This result suggests that a core collection comprised of 10% of the accessions is nearly as efficient as much larger core collections in representing allelic diversity, provided that the selection of the accessions retained is carried out in a manner likely to capture most of the diversity.

Where in a collection the most diverse accessions occur is difficult to predict. As discussed by Brown (1989a), diversity is not randomly distributed in any germplasm collection. The primary determinant of how different populations are, one from another, is the characteristic breeding system of the plant species (Brown, 1989a). Self-pollinated species show more intense population differences, and more uneven distribution of genetic diversity among populations, than do outcrossing species (Brown, 1989a). In the present case of annual *Medicago* species, all of the species in the collection were self-pollinating. Hence, it was imperative that in developing a method of defining a core collection, all of the populations sampled be represented in the core, and as much of the diversity as possible from each population be represented, while maintaining the total size of the core collection to about 10%. Because all populations must be represented, each population would not necessarily contribute to the core in the same proportion as it contributed to the entire collection. As a result, it is likely that the variance of each trait would be greater in the core collection than in the entire collection. Brown (1989a) suggested that numerical methods may be useful for guiding the selection of accessions for inclusion in the core.

Diwan et al. (1995) used 11 different numerical methods to assemble core collections from the U.S. National collection of 3159 accessions of *Medicago* species. The data evaluated included passport and/or field evaluation data. The 11 methods differed in the manner by which clusters of accessions were determined, and how accessions within clusters were chosen for inclusion in the core collection (Diwan et al., 1995). A core collection was considered acceptable if 70% or more of the means and ranges of each trait in each cluster were not significantly different from the corresponding means and ranges in a systematically chosen subset of 1240 of the original 3159 accessions (Diwan et al., 1995). By this criterion, six of the 11 methods produced acceptable core collections (Diwan et al., 1995). These six methods all relied on calculation of distance matrices including all possible comparisons of accessions. Very large data sets (10 000 accessions or more) are not amenable to this method except on very large computers, simply because the system memory requirements are too great. For example, a 10 000 by 10 000 matrix each represented by 16 digits, would occupy about 1.6 gigabytes of mem-

D.Z. Skinner, USDA-ARS, Agronomy Dep., Throckmorton Hall, Kansas State Univ., Manhattan, KS 66506-5501; G.R. Bauchan, USDA-ARS, PSI, Soybean & Alfalfa Research Lab., Beltsville, MD 20705-2350; G. Auricht and S. Hughes, South Australian Research and Development Institute, G.P.O. 397, Adelaide, South Australia 5001. Received 28 Sep. 1998. *Corresponding author (dzolek@ksu.edu).

ory. Germplasm collections larger than 10 000 accessions do exist.

Annual *Medicago* species are an important component of the cereal–pasture ley farming system in southern Australia. Annual *Medicago* species are closely related to perennial alfalfa (*Medicago sativa* L.). Annual *Medicago* species are endemic to the Mediterranean region of the world. All of the annual *Medicago* species have yellow flowers, are self-pollinating and the seed pods are very distinct with seeds larger than alfalfa (Lesins and Lesins, 1979). Annual *Medicago* species have several desirable agronomic characteristics (i) they are fast growing and most will complete their life cycle in 65 to 100 d (Diwan et al., 1995), (ii) they appear to be best adapted to soils with a pH of 6 and above (Crawford et al., 1989), (iii) they have been found to produce up to 70 kg ha⁻¹ of nitrogen provided effective inoculant is used (Webber et al., 1976), (iv) they can be cut for hay or used as pasture and will produce between 4202 to 6415 kg ha⁻¹ of forage (Zhu and Sheaffer, 1997), and (v) the seed can remain viable in the soil for long periods of time and are thus able to reseed themselves after a period of time (Quinlivan, 1971; Rumbaugh and Johnson, 1986).

In the early 1960s, it was realized that the *Medicago* species which were being grown at the time were based on a very limited gene pool of introduced and naturalized species. Thus, the Australian *Medicago* Genetics Resource Center (AMGRC) located at Adelaide, South Australia, was established for the purpose of introducing and evaluating germplasm from the Mediterranean region of the world for the improvement of pastures. There are presently 22 234 accessions in the collection representing 34 species collected in 64 different countries. This collection has been designated by the International Board for Plant Genetic Resources (IBPGR) as the base collection for these species. Over the years, the AMGRC has assembled passport and evaluation data on 20 997 accessions of the annual *Medicago* species.

The evaluation data provides a valuable opportunity to assess relationships among traits. Relatively weak associations of traits may not be distinguishable from statistical 'noise' in small data sets, but in data sets of several thousand observations of each trait, it may be possible to detect obscure trends. Such information is useful from two perspectives: plant breeding and additional evaluations. Plant breeders can utilize the knowledge of associations of expression of different traits to be used as a guide for which plant accessions to evaluate for a desired trait. For example, if two traits are shown to be associated and evaluation data is available for one of the traits in a given set of plant accessions, the association may be exploited as a guide in seeking desirable levels of expression of the second trait. While weak associations will not provide absolute indications of desirable plant genotypes for the second trait, the association will provide an increased probability of finding desirable genotypes in the plants examined.

Germplasm curators can utilize the knowledge of associations of trait expression as a guide to determining how time and resources may be best allocated. For ex-

ample, if the expression of two traits is correlated, and evaluating both traits is resource-expensive, it may be desirable to evaluate only one of the two traits. If, at a later date it becomes necessary to identify plant accessions with a given level of expression of the second, unevaluated trait, the levels of expression of the first trait may be used as a predictor variable.

Our objectives for this study were to develop a method of defining a core collection from the AMGRC's 20 997 evaluated *Medicago* accessions, assess the proportion of existing variation represented in the core collection, and assess the levels of correlated expression of the traits evaluated.

MATERIALS AND METHODS

Evaluation and passport data was assembled for 20 997 *Medicago* accessions. Evaluations were carried out from 1968 to 1994. The AMGRC grows approximately 1000 accessions each year for characterization. Each accession was grown as spaced plants (15 cm apart) in individual rows (3.6–1.5 m) with up to 25 plants per row. The rows were hand planted with seedlings, which were germinated in petri dishes and then grown in seedling trays for six weeks in the greenhouse. Up to 40 seedlings were germinated from seed, which ideally has been selected from a different pod in the collected population. Up to 25 seedlings were transplanted in the field and included with these introduced lines were a range of cultivars and controls used as checks. The plant characteristics which were evaluated are considered to be highly heritable and thus are expressed consistently across environments. Descriptors recorded were based on the International Board for Plant Genetic Resources (IBPGR) list of descriptors for annual *Medicago* species (International Board for Plant Genetic Resources, 1991). Twenty-seven characters were measured and used in assessing relationships. These characters were (i) days to first flower, (ii) days to first mature pod, (iii) pod spininess, (iv) seedling vigor, (v) average score for herbage yield over winter, (vi) average spring herbage yield score, (vii) grams of pods per plant, (viii) number of pods per plant, (ix) percentage seed in the pod, (x) 1000 pod weight, (xi) 1000 seed weight, (xii) seeds per gram, (xiii) grams of seeds per plant, (xiv) number of seeds per plant, (xv) number of seeds per pod, (xvi) growth habit, (xvii) internode length, (xviii) primary branching, (xix) foliage, (xx) leaflet marker position, (xxi) leaflet marker shape, (xxii) leaflet marker color, (xxiii) leaf surface hairiness, (xxiv) florets per peduncle, (xxv) pods per peduncle, (xxvi) pod coil direction, and (xxvii) pods per petiole. The data set consisted of 20 997 entries representing 34 species and 42 subspecies (Table 1). When species, subspecies, and geographic origin were considered, a total of 793 distinct groups were represented. This number of groups resulted when lack of subspecies identification was considered as a valid distinction from identified subspecies. Control cultivars, grown in the field plantings in multiple years, were replaced in the data set with one entry comprised of the mean across all evaluations for that accession. Accessions that were present in the data set but had not been evaluated were deleted from the analysis. The resulting data set consisted of 16 306 entries.

Means and standard errors of each character were calculated with PROC MEANS of SAS (SAS, 1990). A computer program (available from the first author) was written in ANSI C to compute the Euclidean distance of each accession to all other accessions; then the mean Euclidean distance for that accession was calculated and written to a file. This approach avoids the generation of a 16 306 by 16 306 matrix of distances

Table 1. The number and percentages of evaluated accessions of annual *Medicago* species in the Australian *Medicago* Genetics Resource Center collection and in a proposed core collection.

<i>Medicago</i> species	Subspecies	Australian collection		Core collection	
		Number	Percent	Number	Percent
<i>M. aculeata</i> Gaertn.	–	24	0.1	2	0.1
<i>M. aculeata</i> Gaertn.	<i>aculeata</i>	754	3.6	65	3.8
<i>M. aculeata</i> Gaertn.	<i>inermis</i>	106	0.5	33	1.9
<i>M. arabica</i> (L.) Huds.	–	400	1.9	51	3.0
<i>M. blanchiana</i> Boiss.	<i>blanchiana</i>	17	0.1	8	0.5
<i>M. blanchiana</i> Boiss.	<i>bonarotiana</i>	61	0.3	17	1.0
<i>M. ciliaris</i> (L.) Krockner	–	154	0.7	25	1.5
<i>M. constricta</i> Durieu	–	161	0.8	26	1.5
<i>M. coronata</i> (L.) Bart.	–	32	0.2	16	0.9
<i>M. disciformis</i> DC.	–	100	0.5	15	0.9
<i>M. doliata</i> Carmign.	–	14	0.1	6	0.4
<i>M. doliata</i> Carmign.	<i>doliata</i>	18	0.1	4	0.2
<i>M. doliata</i> Carmign.	<i>muricata</i>	80	0.4	7	0.4
<i>M. granadensis</i> Willd.	–	22	0.1	7	0.4
<i>M. heyneana</i> Greuter	–	2	0.0	1	0.1
<i>M. intertexta</i> (L.) Miller	–	162	0.8	26	1.5
<i>M. intertexta</i> (L.) Miller	<i>ciliaris</i>	12	0.1	6	0.4
<i>M. intertexta</i> (L.) Miller	<i>intertexta</i>	82	0.4	24	1.4
<i>M. laciniata</i> (L.) Miller	<i>brachyacantha</i>	100	0.5	9	0.5
<i>M. laciniata</i> (L.) Miller	<i>laciniata</i>	334	1.6	31	1.8
<i>M. littoralis</i> Rohde ex Lois.	–	425	2.0	2	0.1
<i>M. littoralis</i> Rohde ex Lois.	<i>inermis</i>	67	0.3	16	0.9
<i>M. littoralis</i> Rohde ex Lois.	<i>littoralis</i>	1 964	9.4	85	5.0
<i>M. littoralis</i> Rohde ex Lois.	<i>striata</i>	1	0.0	1	0.1
<i>M. minima</i> (L.) Bart.	<i>brevispina</i>	17	0.1	8	0.5
<i>M. minima</i> (L.) Bart.	<i>minima</i>	354	1.7	41	2.4
<i>M. murex</i> Willd.	–	21	0.1	2	0.1
<i>M. murex</i> Willd.	<i>intermis</i>	6	0.0	4	0.2
<i>M. murex</i> Willd.	<i>murex</i>	185	0.9	24	1.4
<i>M. muricoleptis</i> Tin.	–	9	0.0	5	0.3
<i>M. noeana</i> Bioiss.	–	62	0.3	7	0.4
<i>M. orbicularis</i> (L.) Bart.	–	106	0.5	5	0.3
<i>M. orbicularis</i> (L.) Bart.	<i>biancae</i>	459	2.2	55	3.2
<i>M. orbicularis</i> (L.) Bart.	<i>marginata</i>	1 038	4.9	78	4.6
<i>M. polymorpha</i> L.	–	654	3.1	3	0.2
<i>M. polymorpha</i> L.	<i>brevispina</i>	773	3.7	42	2.5
<i>M. polymorpha</i> L.	<i>polymorpha</i>	1 893	9.0	103	6.0
<i>M. polymorpha</i> L.	<i>vulgaris</i>	1 337	6.4	90	5.3
<i>M. praecox</i> DC.	–	71	0.3	16	0.9
<i>M. radiata</i> L.	–	124	0.6	12	0.7
<i>M. rigidula</i> (L.) All.	–	63	0.3	4	0.2
<i>M. rigidula</i> (L.) All.	<i>agrestis</i>	211	1.0	27	1.6
<i>M. rigidula</i> (L.) All.	<i>cinerascens</i>	262	1.2	40	2.3
<i>M. rigidula</i> (L.) All.	<i>rigidula</i>	553	2.6	58	3.4
<i>M. rigidula</i> (L.) All.	<i>submitis</i>	179	0.9	31	1.8
<i>M. rotata</i> Boiss.	–	27	0.1	6	0.4
<i>M. rotata</i> Boiss.	<i>eliezerei</i>	67	0.3	11	0.6
<i>M. rotata</i> Boiss.	<i>rotata</i>	79	0.4	14	0.8
<i>M. rugosa</i> Desr.	–	291	1.4	33	1.9
<i>M. rugosa</i> Desr.	<i>vulgaris</i>	1	0.0	1	0.1
<i>M. sauvagei</i> Negre	–	6	0.0	2	0.1
<i>M. scutella</i> (L.) Miller	–	425	2.0	55	3.2
<i>M. shepardii</i> Post	–	6	0.0	2	0.1
<i>M. soleirolii</i> Duby	–	21	0.1	8	0.5
<i>M. sphaerocarpos</i> Bertol.	–	545	2.6	38	2.2
<i>M. syriaca</i> L.	–	23	0.1	6	0.4
<i>M. tenoreana</i> Ser.	–	5	0.0	3	0.2
<i>M. tornata</i> (L.) Mill.	<i>aculeata</i>	61	0.3	15	0.9
<i>M. tornata</i> (L.) Mill.	<i>rigidula</i>	6	0.0	2	0.1
<i>M. tornata</i> (L.) Mill.	<i>rugulosa</i>	59	0.3	16	0.9
<i>M. tornata</i> (L.) Mill.	<i>spinulosa</i>	399	1.9	33	1.9
<i>M. tornata</i> (L.) Mill.	<i>striata</i>	13	0.1	2	0.1
<i>M. tornata</i> (L.) Mill.	<i>tornata</i>	135	0.6	35	2.1
<i>M. truncatula</i> Gaerth.	–	821	3.9	3	0.2
<i>M. truncatula</i> Gaerth.	<i>inermis</i>	1	0.0	1	0.1
<i>M. truncatula</i> Gaerth.	<i>longeaculeata</i>	298	1.4	27	1.6
<i>M. truncatula</i> Gaerth.	<i>tricycla</i>	398	1.9	31	1.8
<i>M. truncatula</i> Gaerth.	<i>truncatula</i>	3 602	17.2	169	9.9
<i>M. turbinata</i> (L.) All.	–	14	0.1	1	0.1
<i>M. turbinata</i> (L.) All.	<i>aculeata</i>	24	0.1	11	0.6
<i>M. turbinata</i> (L.) All.	<i>spinulosa</i>	37	0.2	11	0.6
<i>M. turbinata</i> (L.) All.	<i>turbinata</i>	164	0.8	31	1.8
Total		20 997	100	1705	100

Table 2. Ranges and standard deviation of observed measurements of 27 characters in the Australian annual *Medicago* collection and in a proposed core collection.

	Australian collection			Core collection		
	Min	Max	Std†	Min	Max	Std†
Days to first flower	37.0	179.0	17.7	60.0	172.6	19.4
Days to first mature pod	114.0	799.0	13.3	114.0	799.0	19.4
Pod spininess	0.0	20.0	5.9	0.0	20.0	6.3
Seedling vigor	1.0	25.0	2.8	1.0	25.0	3.3
Average score for herbage yield over winter	0.0	41.0	3.8	0.0	37.5	4.4
Average spring herbage yield score	0.0	63.3	6.6	0.0	53.3	7.3
Grams of pods per plant	0.0	149.2	19.3	0.0	149.2	24.2
Number of pods per plant	0.0	8 333.3	483.2	0.0	8 333.3	729.0
Percentage of seed in pod	0.0	78.7	8.7	0.0	75.0	9.3
100 pod weight	1.8	615.0	61.6	1.8	615.0	79.2
100 seed weight	0.5	63.8	3.1	0.5	63.8	4.0
Seeds per gram	15.7	2 202.6	145.5	15.7	2 202.6	206.3
Grams of seeds per plant	0.0	62.1	7.4	0.0	62.1	9.1
Number of seeds per plant	0.0	28 891.8	1 938.5	0.0	28 891.8	2558.6
Number of seeds per pod	0.0	30.9	3.2	0.0	30.9	3.3
Growth habit	1.0	4.0	0.6	1.0	4.0	0.7
Internode length	1.0	3.0	0.7	1.0	3.0	0.8
Primary branching	1.0	3.0	0.5	1.0	3.0	0.5
Foliage	1.0	3.0	0.6	1.0	3.0	0.7
Leaflet marker position	0.0	21.0	6.7	1.0	21.0	7.0
Leaflet marker shape	0.0	30.0	7.6	1.0	29.0	6.8
Leaflet marker color	0.0	21.0	4.7	1.0	17.0	4.2
Leaf surface hairiness	1.0	9.0	3.8	1.0	9.0	3.7
Florets per peduncle	1.0	17.0	1.9	1.0	17.0	2.6
Pods per peduncle	1.0	13.0	1.3	1.0	13.0	1.9
Pod coil direction	1.0	3.0	0.5	1.0	3.0	0.4
Peduncle:petiole ratio	1.0	9.0	114.2	1.0	9.0	2.1

† Std = Standard deviation. A test of equality of variances (Bartlett's test) indicated that all of the standard deviations were significantly different between collections ($P = 0.05$).

in computer memory. The passport data was then used to sort the accessions into 793 groups defined by species, subspecies, and geographical origin. The last step was to construct a subset of data from the groups. This was done with a separate program written in ANSI C. Each group was represented by the extremes of average Euclidean distance found in the group. Within each group, 'typical' accessions, those most like all other accessions in the collection will have low average Euclidean distances, while more unusual accessions will have large average Euclidean distance. To develop a small core collection and still represent most of the variability in the collection, a small percentage of typical accessions, and a larger percentage of unusual accessions were selected from each group. The computer program was written such that at least one accession from the extremes of average Euclidean distance of each of the 793 groups was selected. Groups comprised of one or two accessions were included as a unit in the core collection. This approach ensures that all geographic origins, species, and subspecies are represented. It also ensures that the 'range ratio', the ratio of the range of a trait in the core relative to the entire collection, (Diwan et al., 1995) will be maximized for the mean Euclidean distance in each of the 793 groups. However, the maximum range of each individual character is not necessarily represented by this method. The measured range of each character in the core collection was compared to the range in the original data set to assess the amount of the original range of each character captured in the core by this method.

Associations of trait expression were evaluated by calculating correlation coefficients of all combinations of the evaluated traits. The data set of 16 306 entries was used. Correlations of traits which were scored as ordinal numbers were evaluated with Spearman's rank correlation coefficients (PROC CORR; SAS, 1990) when compared with other ordinal measures or to continuous traits. Correlations of continuous traits were evaluated with Pearson's correlation coefficients (PROC CORR; SAS, 1990). Because 27 traits were evaluated,

351 comparisons were possible. Since these comparisons represent sequential comparisons on the same data set, it is advisable to apply Bonferroni's inequalities to the significance level of the statistical tests (Snedecor and Cochran, 1980). Therefore, a significance level of $(0.05/351) = 0.0001$ was required before a correlation coefficient was considered significant.

The variance-covariance matrix from the core collection was compared with the variance-covariance matrix from the entire collection using the POOL = TEST option of PROC DISCRIM (SAS, 1990). Comparison of variances for individual traits were made with Bartlett's test (Snedecor and Cochran, 1980, page 296). Equality of correlation coefficients in the two collections was tested with Fisher's Z transformation (Snedecor and Cochran, 1980, page 185) and associated test of significance.

RESULTS AND DISCUSSIONS

The mean Euclidean distances of the 16 306 accessions ranged from 5.4 to 48.7. Retaining 0.5% of the low end and 3.0% of the high end of the range of average Euclidean distances within each of the 793 groups defined by species, subspecies, and geographic origin resulted in the retention of 1705 accessions, or 10.4% of the entire collection. To evaluate how well this 1705 accession core collection captured the variation in the original collection, the ranges of each character in the core collection were compared with the corresponding ranges in the entire collection (Table 2). Both extremes were captured in 20 of 27 characters (Table 2). Only one character (days to first flower) had neither extreme represented (Table 2). Thus, 41 of 54 extremes were represented in the core. Additional accessions to represent each of the extremes not included in the core could easily be added to the core collection.

Table 3. Correlation coefficients† of 27 characters measured in the Australian *Medicago* Genetics Resource Center collection.

Trait	Trait								
	1	2	3	4	5	6	7	8	9
1. Days to first flower	1.00	0.62	0.16	-0.16	-0.21	-0.19	-0.23	-0.05	0.15
2. Days to first mature pod	0.62	1.00	-0.01	-0.15	-0.17	-0.17	0.09	0.10	0.06
3. Pod spininess	0.16	-0.01	1.00	0.18	0.22	0.24	0.11	0.08	0.05
4. Seedling vigor	-0.16	-0.15	0.18	1.00	0.74	0.62	0.35	0.05	0.03
5. Winter herbage yield	-0.21	-0.17	0.22	0.74	1.00	0.78	0.35	0.12	0.09
6. Spring herbage yield	-0.19	-0.17	0.24	0.62	0.78	1.00	0.41	0.30	0.20
7. Grams of pods/plant	-0.23	0.09	0.11	0.35	0.35	0.41	1.00	0.44	0.18
8. Number of pods/plant	-0.05	0.10	0.08	0.05	0.12	0.30	0.44	1.00	0.25
9. % seed in pod	0.15	0.06	0.05	0.03	0.09	0.20	0.18	0.25	1.00
10. 100 pod weight	-0.16	-0.08	0.00	0.28	0.26	0.07	0.26	-0.46	-0.07
11. 100 seed weight	-0.16	-0.09	0.00	0.44	0.41	0.24	0.25	-0.25	0.07
12. Seeds per gram	0.09	0.11	0.00	-0.44	-0.42	-0.29	-0.29	0.32	-0.11
13. Grams of seeds/plant	-0.15	0.09	0.11	0.30	0.35	0.45	0.91	0.48	0.52
14. Number of seeds/plant	-0.07	0.14	0.11	0.04	0.10	0.27	0.68	0.74	0.46
15. Number of seeds/pod	0.06	0.03	0.07	0.00	-0.11	-0.13	0.14	-0.32	0.42
16. Growth habit	-0.20	-0.18	0.07	0.22	0.39	0.42	0.16	0.10	0.16
17. Internode length	-0.17	-0.07	0.07	0.21	0.41	0.50	0.38	0.25	0.06
18. Primary branching	0.19	0.19	0.09	0.06	-0.06	-0.10	0.11	0.15	0.03
19. Foliage	-0.07	-0.03	0.20	0.46	0.58	0.64	0.48	0.33	0.21
20. Leaflet marker position	0.16	0.09	-0.09	-0.17	-0.16	-0.20	-0.22	-0.11	0.04
21. Leaflet marker shape	-0.08	-0.13	0.18	0.15	0.19	0.25	0.19	0.15	0.00
22. Leaflet marker color	0.06	0.03	-0.15	-0.11	-0.11	-0.15	-0.13	-0.10	0.04
23. Leaf surface hairiness	-0.27	-0.05	-0.37	-0.18	-0.34	-0.39	-0.12	-0.19	-0.48
24. Florets per peduncle	0.13	0.03	0.07	0.08	0.13	0.18	0.02	0.36	0.12
25. Pods per peduncle	0.11	-0.01	0.10	0.12	0.13	0.22	0.05	0.49	0.12
26. Pod coil direction	0.30	0.13	0.19	-0.01	0.08	0.11	-0.06	0.03	0.31
27. Peduncle:petiole ratio	0.22	0.10	-0.08	-0.09	-0.13	-0.12	-0.21	-0.13	-0.06
Trait	10	11	12	13	14	15	16	17	18
1. Days to first flower	-0.16	-0.16	0.09	-0.15	-0.07	0.06	-0.20	-0.17	0.19
2. Days to first mature pod	-0.08	-0.09	0.11	0.09	0.14	0.03	-0.18	-0.07	0.19
3. Pod spininess	0.00	0.00	0.00	0.11	0.11	0.07	0.07	0.07	0.09
4. Seedling vigor	0.28	0.44	-0.44	0.30	0.04	0.00	0.22	0.21	0.06
5. Winter herbage yield	0.26	0.41	-0.42	0.35	0.10	-0.11	0.39	0.41	-0.06
6. Spring herbage yield	0.07	0.24	-0.29	0.45	0.27	-0.13	0.42	0.50	-0.10
7. Grams of pods/plant	0.26	0.25	-0.29	0.91	0.68	0.14	0.16	0.38	0.11
8. Number of pods/plant	-0.46	-0.25	0.32	0.48	0.74	-0.32	0.10	0.25	0.15
9. % of seed in pod	-0.07	0.07	-0.11	0.52	0.46	0.42	0.16	0.06	0.03
10. 100 pod weight	1.00	0.75	-0.59	0.18	-0.25	0.44	0.01	0.08	-0.05
11. 100 seed weight	0.75	1.00	-0.67	0.23	-0.28	-0.01	0.19	0.20	-0.09
12. Seeds/gram	-0.59	-0.67	1.00	-0.27	0.23	-0.13	-0.19	-0.20	0.09
13. Grams of seeds/plant	0.18	0.23	-0.27	1.00	0.77	0.27	0.19	0.38	0.11
14. Number of seeds/plant	-0.25	-0.28	0.23	0.77	1.00	0.21	0.05	0.24	0.18
15. Number of seeds/pod	0.44	-0.01	-0.13	0.27	0.21	1.00	-0.08	-0.06	0.04
16. Growth habit	0.01	0.19	-0.19	0.19	0.05	-0.08	1.00	0.14	-0.25
17. Internode length	0.08	0.20	-0.20	0.38	0.24	-0.06	0.14	1.00	-0.21
18. Primary branching	-0.05	-0.09	0.09	0.11	0.18	0.04	-0.25	-0.21	1.00
19. Foliage	0.08	0.30	-0.30	0.50	0.30	-0.11	0.28	0.45	0.01
20. Leaflet marker position	-0.07	-0.01	0.01	-0.18	-0.17	-0.10	-0.01	-0.19	0.01
21. Leaflet marker shape	-0.01	0.03	-0.03	0.16	0.16	0.01	0.06	0.18	0.00
22. Leaflet marker color	0.01	0.05	-0.05	-0.09	-0.11	-0.05	-0.05	-0.11	0.01
23. Leaf surface hairiness	0.16	-0.04	0.04	-0.27	-0.23	0.09	-0.19	-0.18	0.01
24. Florets/peduncle	-0.33	-0.14	0.10	0.07	0.16	-0.28	0.07	0.27	-0.07
25. Pods/peduncle	-0.42	-0.19	0.18	0.08	0.22	-0.37	0.11	0.19	-0.03
26. Pod coil direction	0.13	0.05	-0.05	0.04	0.00	-0.12	0.13	0.02	-0.01
27. Peduncle:petiole ratio	-0.07	0.05	-0.05	-0.21	-0.23	-0.19	-0.04	-0.04	-0.06
Trait	19	20	21	22	23	24	25	26	27
1. Days to first flower	-0.07	0.16	-0.08	0.06	-0.27	0.13	0.11	0.30	0.22
2. Days to first mature pod	-0.03	0.09	-0.13	0.03	-0.05	0.03	-0.01	0.13	0.10
3. Pod spininess	0.20	-0.09	0.18	-0.15	-0.37	0.07	0.10	0.19	-0.08
4. Seedling vigor	0.46	-0.17	0.15	-0.11	-0.18	0.08	0.12	-0.01	-0.09
5. Winter herbage yield	0.58	-0.16	0.19	-0.11	-0.34	0.13	0.13	0.08	-0.13
6. Spring herbage yield	0.64	-0.20	0.25	-0.15	-0.39	0.18	0.22	0.11	-0.12
7. Grams of pods/plant	0.48	-0.22	0.19	-0.13	-0.12	0.02	0.05	-0.06	-0.21
8. Number of pods/plant	0.33	-0.11	0.15	-0.10	-0.19	0.36	0.49	0.03	-0.13
9. % of seed in pod	0.21	0.04	0.00	0.04	-0.48	0.12	0.12	0.31	-0.06
10. 100 pod weight	0.08	-0.07	-0.01	0.01	0.16	-0.33	-0.42	-0.13	-0.07
11. 100 seed weight	0.30	-0.01	0.03	0.05	-0.04	-0.14	-0.19	0.05	0.05
12. Seeds/gram	-0.30	0.01	-0.03	-0.05	0.04	0.10	0.18	-0.05	-0.05
13. Grams of seeds/plant	0.50	-0.18	0.16	-0.09	-0.27	0.07	0.08	0.04	-0.21
14. Number of seeds/plant	0.30	-0.17	0.16	-0.11	-0.23	0.16	0.22	0.00	-0.23
15. Number of seeds/pod	-0.11	-0.10	0.01	-0.05	0.09	-0.28	-0.37	-0.12	-0.19
16. Growth habit	0.28	-0.01	0.06	-0.05	-0.19	0.07	0.11	0.13	-0.04
17. Internode length	0.45	-0.19	0.18	-0.11	-0.18	0.27	0.19	0.02	-0.04
18. Primary branching	0.01	0.01	0.00	0.01	0.01	-0.07	-0.03	-0.01	-0.06
19. Foliage	1.00	-0.16	0.20	-0.11	-0.37	0.30	0.31	0.09	-0.16
20. Leaflet marker position	-0.16	1.00	-0.69	0.68	-0.05	-0.12	-0.11	0.13	0.15
21. Leaflet marker shape	0.20	-0.69	1.00	-0.59	-0.07	0.17	0.17	-0.04	-0.05
22. Leaflet marker color	-0.11	0.68	-0.59	1.00	0.00	-0.09	-0.10	0.01	0.07
23. Leaf surface hairiness	-0.37	-0.05	-0.07	0.00	1.00	-0.25	-0.21	-0.46	0.06
24. Florets/peduncle	0.30	-0.12	0.17	-0.09	-0.25	1.00	0.86	0.08	0.08
25. Pods/peduncle	0.31	-0.11	0.17	-0.10	-0.21	0.86	1.00	0.06	0.09
26. Pod coil direction	0.09	0.13	-0.04	0.01	-0.46	0.08	0.06	1.00	0.12
27. Peduncle:petiole ratio	-0.16	0.15	-0.05	0.07	0.06	0.08	0.09	0.12	1.00

† Correlations of traits which were scored as ordinal numbers are represented with Spearman's rank correlation coefficients when compared to other ordinal measures or to continuous traits while correlations of continuous traits were evaluated with Pearson's correlation coefficients. Traits 3, 4, 16, 18-23, and 26 are ordinal.

The core collection comprised slightly more than 10% of the entire collection, and care was taken to represent both diverse and typical genotypes from within each population. Hence, Brown's (1989a) analysis would suggest that at least 70% of the alleles in the entire collection were represented in the core collection of 1705 accessions. The contribution of each species to the Australian and core collections are shown in Table 1.

A global comparison of the variance-covariance matrices based on all traits from the original and core collections indicated a highly significant difference ($\chi^2 > 10,000$, $P < 0.0001$). Pairwise comparisons of the variances of each trait indicated that all variances were significantly ($P < 0.01$) different between the core collection and the entire collection (calculations not shown). The standard deviations of 22 of the 27 traits (Table 2) were greater in the core collection than in the entire collection. Because the core collection deliberately was selected to contain a greater proportion of the extreme types than of the typical types, a larger variance (and therefore, standard error) in the core collection relative to the entire collection was expected. Therefore, measurements made to assess variance within a species, for example, should be based on the entire collection, not the core collection.

We concur with Diwan et al. (1995) that the most representative core collections can be assembled only if all available information is used. Toward that end, the method we propose here systematically selects the most typical, and most diverse, accessions from each group defined by species, subspecies, and geographic origin. It appears that most of the genetic variation in the original data set is represented in the core collection defined by this method. The computer programming necessary to carry out this analysis requires a relatively small amount of computing resources and should be accessible to most researchers. The computer programs are available from the first author.

Correlations were conducted between all 27 traits (Table 3). With more than 16 000 degrees of freedom, any correlation coefficient with an absolute value greater than 0.01 was significant at $P = 0.0001$. However, because the proportion of variance in one trait that can be attributed to its linear relationship to a second trait is indicated by the square of the correlation coefficient (Snedecor and Cochran, 1980), we suggest that meaningful correlation coefficients in this case are those with absolute values greater than 0.71, that is, more than 50% of the variance of one trait is predicted by the other. By this criterion, the significant relationships were seedling vigor and average winter yield ($r = 0.74$), average spring yield and average winter yield ($r = 0.78$), grams of seed per plant and grams of pods per plant ($r = 0.91$), number of seed per plant and number of pods per plant ($r = 0.74$), 1000 seed weight and 1000 pod weight ($r = 0.75$), number of seeds per plant and grams of seeds per plant ($r = 0.77$), and florets

per peduncle and pods per peduncle ($r = 0.86$). None of these relationships were surprising, but they do suggest that it may not be necessary to measure all traits in future germplasm evaluations.

Other relationships, while not meeting the 50% criterion, may be of interest to plant breeders. For example, seeds per gram, an easily measured trait that can be attained for any new plant accession, is significantly related to the traits of herbage and seed yield ($r = 0.4$ – 0.7 ; Table 3), suggesting that seeds per gram may be a useful measure in choosing new accessions for further evaluations.

Although the correlation measurements (Table 3) were based on the entire collection, we also calculated the same correlation coefficients in the core collection and tested their similarity to the coefficients based on the entire collection. Of the coefficients discussed above as being particularly interesting, none were significantly ($P < 0.05$) different between the core and entire collections. We interpret this result to indicate that the associations described are quite robust and tend towards association across a broad range of genotypes and species.

The method presented here should prove useful in the definition of core collections from large germplasm repositories. We suggest that the proposed core collection from the AMGRC collection represents nearly all of the genetic variation in the collection, and should facilitate the development and application of annual medics in Australia and throughout the world. The list of 1705 accessions in the core collection may be obtained from the first author.

REFERENCES

- Brown, A.H.D. 1989a. The case for core collections. p. 136–156. In A.H.D. Brown et al. (ed.) *The use of plant genetic resources*. Cambridge University Press, Cambridge.
- Brown, A.H.D. 1989b. Core collection: A practical approach to genetic resources management. *Genome* 31:818–824.
- Crawford, E.J., A.W.H. Lake, and K.G. Boyce. 1989. Breeding annual *Medicago* species for semiarid conditions in southern Australia. *Adv. Agron.* 42:399–234.
- Diwan, N., M.S. McIntosh, and G.R. Baughan. 1995. Methods of developing a core collection of annual *Medicago* species. *Theor. Appl. Genet.* 90:755–761.
- International Board for Plant Genetic Resources. 1991. *Descriptors for annual Medicago*. International Board for Plant Genetic Resources, Rome.
- Lesins, K., and I. Lesins. 1979. Genus *Medicago* (Leguminosae). A taxonomic study. Junk, The Hague.
- Quinlivan, B.J. 1971. Seed coat impermeability in legumes. *J. Aust. Inst. Agric. Sci.* 37:283–295.
- Rumbaugh, M.D., and D.A. Johnson. 1986. Annual medics and related species as reseeding legumes for Northern Utah pastures. *J. Range Manage.* 39:52–58.
- SAS Institute, Inc. 1990. *SAS procedures guide*, Version 6, 3rd ed. SAS Institute, Inc., Cary, NC.
- Snedecor, G.W., and W.G. Cochran. 1980. *Statistical methods*. 7th ed. Iowa State Univ. Press, Ames.
- Webber, P.H., P.S. Cocks, and B.C. Jeffries. 1976. *Farming systems in South Australia*. South Aust. Dep. of Agric. Adelaide, South Australia.
- Zhu, Y., and C.C. Sheaffer. 1997. Growth analysis of spring and summer seeded annual *Medicago* species. *Crop Sci.* 37:1514–1519.