

# Assessment of Genetic Variation in Hairy Vetch Using Canonical Discriminant Analysis

Kathleen M. Yeater, German A. Bollero,\* Donald G. Bullock, A. Lane Rayburn, and Sandra Rodriguez-Zas

## ABSTRACT

For improvement of hairy vetch (*Vicia villosa* Roth) as a winter cover crop, it is necessary to gain insight into the magnitude of variability present in the species. This study was conducted to assess the sources of genetic and phenotypic variability of *V. villosa* accessions. Thirteen morphological and agronomic traits were measured on 42 populations of *V. villosa* and two populations of *V. pannonica* Crantz in field studies at Urbana, IL, for 2 yr. These measurements include initial seed weight, germination, stem length, stem width, leaf length, and leaf width in the fall and spring, winter survival, biomass, and a ratio of organic C to total N (C:N ratio) of the plants. The multivariate data set was analyzed by canonical discriminant analysis (CDA) in combination with a clustering procedure. In this analysis, the first two canonical variates were significant and accounted for 94% of the among-accession variability. The canonical variates indicated that fall and spring measurements of the leaf and stem, and C:N ratio are the most differentiating traits among the accessions. The canonical variates were used to cluster the accessions into four subgroups on the basis of the differentiating traits. Canonical discriminant analysis was useful in identifying the genetic variation and the traits that better describe the variation among hairy vetch populations. Cluster analysis was successful in differentiating the accessions into similar subgroups on the basis of the measured traits. Plant breeders can use the information on variation among *Vicia* accessions and focus on traits of particular significance.

WINTER COVER CROPS have long been recognized as possible tools for maintaining and improving soil quality and consequently addressing production and environmental concerns in agriculture (Wagger et al., 1998). Leguminous cover crops provide ground cover to reduce erosion potential and synthetic fertilizer N inputs. In the central midwestern USA, hairy vetch is reported to be the most beneficial leguminous winter cover crop (Bollero and Bullock, 1994).

Hairy vetch has performed well under a wide range of growing conditions in several locations throughout the USA. However, Smith et al. (1987) observed that temperature limitations are a likely occurrence north of the Ohio River. Frequent winterkill of fall-seeded plants is likely, and when plants do survive the winter, the accumulated biomass is variable and the subsequent summer crop still requires additional N fertilizer applications. According to Wagger et al. (1998), it is necessary to select for the following important traits in cover-crop-based production systems: chemical composition for enhanced N dynamics and rate of dry matter accumulation in relation to winter hardiness.

For improvement of these traits in hairy vetch, it is imperative to gain insight into the magnitude of variability present in the species, as this will provide the basis for effective selection. Phenotypic variability is the observable physical variation present in a population and includes both genotypic and environmental components. Genotypic variability is the component of variation that is due to the genotypic differences among individuals within a population or among populations within a species, and is the main concern of a plant breeder. The phenotype is based on quantitative and qualitative characteristics by combining the genotypic basis and its interaction with the environment (Loos, 1993). If phenotypic observations are based on adequately large sample sizes and the physical traits measured show significant differences among populations, they can provide a reasonable representation of overall genetic performance (Humphreys, 1991).

There are several methods to measure variation. With univariate analyses, each variable is analyzed separately allowing for substantial overlapping of results to occur. Univariate statistical techniques such as analysis of variance do not explain how accessions differ when all measured variables are considered jointly. In CDA, a multivariate statistical technique, all variables are considered simultaneously in the differentiation of populations. This approach results in a more powerful comparison of populations than can be achieved with univariate analysis, provided the variables are correlated. Canonical discriminant analysis can separate among-population effects from within population effects by maximizing discrimination among populations when tested against the variation within populations (Riggs, 1973). After determination of the among-population variability, the Mahalanobis distance ( $D^2$ ) statistic can be used as an indicator of the difference between populations (Loos, 1993). The information obtained from CDA can then additionally be used to group the accessions-populations into smaller subgroups that are more similar to each other (Khattree and Naik, 2000). Multivariate procedures based on phenotypic, morphologic, and agronomic characteristics have been used in the assessment of genetic diversity in soybean [*Glycine max* (L.) Merr] (Bains and Sood, 1984), perennial ryegrass (*Lolium perenne* L.) (Humphreys, 1991), *Lolium* spp. (Loos, 1993), and tall fescue (*Festuca arundinacea* Schreb) (Vaylay and van Santen, 2002).

Hairy vetch is an annual, predominantly cross-fertilizing species. In recent literature, only two cultivars of hairy vetch have been registered. Plant introduction line (PI) 383803, named 'Americus', is a result of a direct increase from a wild population (Surrency et al., 1995).

K.M. Yeater, G.A. Bollero, D.G. Bullock, and A.L. Rayburn, Dep. of Crop Sciences, S. Rodriguez-Zas, Dep. of Animal Sciences, Univ. of Illinois, 1102 S. Goodwin Ave., Urbana, IL 61801. Received 28 Oct. 2002. \*Corresponding author (gbollero@uiuc.edu).

Published in Crop Sci. 44:185–189 (2004).  
© Crop Science Society of America  
677 S. Segoe Rd., Madison, WI 53711 USA

**Abbreviations:** C:N ratio, carbon to nitrogen ratio; CDA, canonical discriminant analysis; NPGS, National Plant Germplasm System; PI, plant introduction.

Also, PI 575701, 'AU Earlycover', is a selection from accession 9053961, a composite of PI 561945, PI 561946, and PI 561947 (Mosjidis et al., 1995). Little information is available on the genetic and phenotypic variability of hairy vetch accessions available through the National Plant Germplasm System (NPGS). Canonical discriminant analysis is able to distinguish the within-accession variation due to environmental effects and also due to genetic effects of the accessions. This discrimination is obtained by the ratio of variance among accessions to the variance within accessions (Rencher, 1992). The objective of this study was to assess genetic diversity of hairy vetch accessions on the basis of phenotypic characteristics.

## MATERIALS AND METHODS

### Plant Material

The experimental material consisted of 42 populations of *V. villosa* and two populations of *V. pannonica*. All plant introduction (PI) lines were provided by Dr. Clarice Coyne (USDA-ARS Western Regional Plant Introduction Station, Washington State University, Pullman, WA). Common hairy vetch (variety unknown) was obtained from a local feed-supply store.

### Experimental Design

One hundred seventy-six seeds per accession were planted in individual Jiffy pellets (Humert Seed Company, Earth City, Mo) in August 2000 and 2001. They germinated in flats in the greenhouse. Four weeks after planting, the *V. villosa* plant material was transplanted to field plots in Urbana, IL. The *V. pannonica* plant material was transplanted to field plots 6 wk after planting in the greenhouse.

The field design was a randomized complete block with two blocks. In each block, there were 44 plots measuring 2 by 1 m. Each plot represented an individual accession, and there were 50 plants transplanted to each plot. The individual plants were planted in a five columns by 10 rows lattice with 200 mm between each plant. This was conducted in 2 yr.

Initial seed weight for 100 seeds per accession was measured. Percentage germination was calculated for each accession on the basis of the number of plants germinated/total seeds planted in the jiffy pellets. Stem length, stem width, leaf length, and leaf width were measured with a caliper to record fall growth in late October. These measurements were taken on five random plants per accession plot. The leaf measurements were taken near the base of the plant on the third leaflet from the tendril. Stem measurements were taken on the main stems. Ten leaf and stem measurements were taken from each plant. In late April 2001 and April 2002, several additional measurements were recorded in the field for each accession. Winter survival percentages for each accession were determined by number of plants alive/number of plants transplanted. Five plants per plot (if not winter-killed) were randomly selected to measure stem length, stem width, leaf length, and leaf width, with the same methods as the fall measurements. Additionally, a random (30 by 60 cm) section of each plot was cut with hand clippers and weighed for biomass accumulation. Standing residues were cut at ground level with electric shears and the plant material was gathered by hand. Residue samples were dried for 3 d at 65°C, weighed, and ground to pass through a 1-mm mesh. Residue samples were analyzed for total C and N with an automated Dumas instru-

ment (Leco CHN-2000. Leco Corp., St. Joseph, MI). These data were included in the statistical analysis as a ratio of C to N (C:N ratio).

### Data Analysis

The data were analyzed with the combination of CDA and a nonhierarchical clustering method. These analyses were performed with the PROC CANDISC and PROC FASTCLUS procedures of SAS (SAS, 2000). Canonical discriminant analysis is a combination of principal component and canonical correlation analyses (Vaylay and van Santen, 2002). From principal components analysis, linear combinations of the original variables that account for as much of the total variation as possible are constructed. From canonical correlation, a linear association between predictor variables and criterion measures are determined (Dillon and Goldstein, 1984). In this study, a classification variable (accession), and several measured traits are analyzed with CDA to derive canonical variables. Canonical variables are linear combinations of the original quantitative measurements that contain the highest possible multiple correlation with each group and that best summarize among-class variation. Canonical variables are uncorrelated among them even though the measured traits may be highly correlated. In CDA, the differentiation of groups is based on the correlation among the independent variables (measured traits) and their relationships with the dependent variable (accessions) (Vaylay and van Santen, 2002).

The canonical variables determined by CDA were then used to aggregate the accessions into groups that were similar in meaningful ways by cluster analysis (Khattree and Naik, 2000). The goal of this analysis is to arrive at clusters of accessions that display small within-cluster variation relative to the between-cluster variation. In the application of cluster analysis to a set of data, it is necessary to determine a measure of inter-accession similarity and specify a procedure for forming the clusters that is based on the chosen measure of similarity. In this analysis, the similarity measurement comes directly from the calculated canonical variables. The mean values of the canonical variables are referred to as group centroids. The difference between centroid values of two groups is the Mahalanobis distance  $D^2$  and is calculated as:

$$D^2 = (\bar{X}_1 - \bar{X}_2)' S^{-1}(\bar{X}_1 - \bar{X}_2) \quad [1]$$

where  $S^{-1}$  is the inverse of the pooled sample variance-covariance matrix, and  $\bar{X}_1$  and  $\bar{X}_2$  are the respective vectors of measurements on groups 1 and 2. This measure has the distinct advantage of accounting for any correlations that might exist between the variables (Dillon and Goldstein, 1984).

The Mahalanobis distance values were used in a nonhierarchical clustering procedure utilizing a  $k$ -means approach (Khattree and Naik, 2000). This procedure begins by the selection of seeds as the initial centroids for the  $k$  clusters. In this analysis, the initial seeds are the group centroids mentioned previously. Second, after all observations are assigned to the respective clusters, all cluster seeds are updated by replacing old seeds with the new cluster centroid. This step is repeated until the changes in cluster seeds become very small or zero. Finally, when all observations are assigned to the clusters with seeds nearest to the corresponding observations, the final clusters result. In SAS, this procedure is implemented in the FASTCLUS procedure as a refinement of the Nearest Centroid Sorting method (Anderberg, 1973).

## RESULTS AND DISCUSSION

In this analysis, the first two canonical variates were significant ( $P < 0.0001$ ) and accounted for 94% of the among-accession variance (Table 1). Each canonical variate is the linear combination of the independently measured variables and is orthogonal to the others. As suggested by Vaylay and van Santen (2002), CDA measures the strength of the overall relationships between the linear composite of the predictor and criterion sets of variables. In this analysis, the predictor is the canonical variate and the criterion is the *Vicia* accessions. The significant ( $P < 0.0001$ ) canonical correlation between the accessions and the first canonical variate ( $r_c = 0.93$ ), and the accessions and the second canonical variate ( $r_c = 0.89$ ) indicates that the canonical variates explain the differentiation of the accessions.

Canonical loadings measure the simple linear correlation between an original independent variable and the canonical variate. Thus, the canonical loading reflects the variance that the observed variable shares with the canonical variate, and can be interpreted in assessing the relative contribution of each variable to each canonical variate function (Cruz-Castillo et al., 1994). The first canonical discriminant function is dominated by large loadings from the spring measurements of stem length, leaf length, leaf width, and stem width, and followed by C:N ratio (Table 1). The second canonical discriminant function is dominated by a large loading from fall stem length and fall leaf length, followed by the spring measurements of stem length, stem width, and leaf width. These functions indicate that the genetic composition of the accessions differ mostly in their fall and spring growth rates and C:N ratio. The C:N ratio is often used as a tool for predicting the decomposition rate of resi-

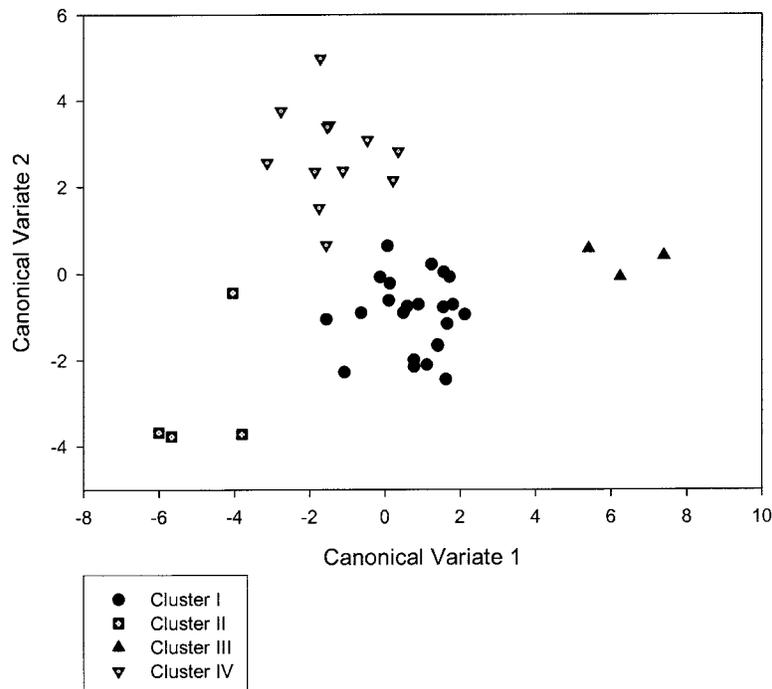
**Table 1.** The canonical loadings of the measured traits on the first two canonical variables of the hairy vetch accessions. The study was conducted for 2 yr at the Crop Science Research and Education Center in Urbana, IL, from 2000 to 2002.

Trait	Canonical variate	
	1	2
Seed weight	-0.31	-0.44
Germination	0.25	0.07
Fall stem length	-0.27	0.86
Fall stem width	0.00	0.26
Fall leaf length	-0.06	0.51
Fall leaf width	-0.02	0.41
Winter survival	0.17	-0.00
Spring stem length	0.85	0.47
Spring stem width	0.63	0.43
Spring leaf length	0.81	0.27
Spring leaf width	0.76	0.45
Biomass	-0.03	0.07
C:N	0.46	0.18
Canonical correlation	0.93	0.89
<i>P</i> level of significance	0.0001	0.0001
Variance accounted for, %	0.60	0.34

dues and subsequent nutrient release (Wagger et al., 1998).

The canonical variates were then used to cluster the accessions into four subgroups (Fig. 1). The centroid values for the first two canonical discriminant functions for the accessions were also plotted (Fig. 1). The separation of the accessions is measured by Mahalanobis distance,  $D^2$  (Table 2). All pairwise distances between cluster groups were significant ( $P < 0.0001$ ). Each cluster group displays small within-cluster genetic variation relative to the between-cluster genetic variation.

Two *V. villosa* accessions, PI 393916 and PI 393922, are not included in any cluster group (Table 3). These two accessions were observed to have no winter survival for either year of the study; and therefore, there were no



**Fig. 1.** Scatterplot of the four cluster groupings on the two canonical discriminant functions. Centroid locations of each cluster are indicated by roman numerals that correspond to a particular cluster group. The study was conducted at Urbana, IL, from 2000 to 2002.

**Table 2. Pairwise squared distances between cluster groups as calculated by Mahalanobis Distance,  $D^2$ . All distances between cluster groups are significant ( $P < 0.0001$ ). Mahalanobis distances measure the extent of genetic diversity between the cluster groups. The study was conducted at Urbana, IL, from 2000 to 2002.**

Cluster	I	II	III	IV
I	0	40.42	39.41	18.93
II		0	138.72	46.50
III			0	69.90
IV				0

plants available to take any of the spring measurements. This lack of data does not allow for the calculation of the canonical variates for these two *V. villosa* accessions.

Cluster I is composed of 23 of the 44 populations included in this analysis (Table 3). This group of accessions includes the common hairy vetch population that is most available to growers in central Illinois. Cluster II is composed of four accessions. Two of these accessions, PI 170008 and PI 515988, are *V. pannonica* species. This indicates that the *V. villosa* accessions in this cluster are more similar to *V. pannonica* than any of the other *V. villosa* accessions in the study. Cluster III contains three accessions, including PI 383808, cultivar Americus (Surrency et al., 1995). Cluster IV includes the remaining 12 accessions. This cluster group contains PI 575701, the second cultivar in the study, AU Earlycover (Mosjidis et al., 1995).

Cluster I accessions differed the most in their fall and spring leaf and stem measurements, as was indicated by their dominant canonical loadings. Cluster II had the smallest plants in the fall and spring as compared with the other accessions in the study. Cluster IV accessions were the largest in the fall and second largest in the spring. Cluster III plants increased the most in size from fall to spring. These accessions were the second smallest in the fall and grew to be the largest plants in the spring.

The next most differentiating trait among the clusters was C:N ratio. The C:N ratio is often used as a tool for predicting the decomposition rate of residues (Wagger et al., 1998). These authors indicated that greater N release occurs from residues with narrower C:N ratios when comparing residues with a range of 36 to 8. As a reference, the common hairy vetch population that was used in this study had a C:N ratio of 10.1 and was included in Cluster I. The mean C:N ratios of Clusters I (10.8) and IV (10.9) were not significantly different. The accessions in Cluster III and Cluster II had a C:N ratio of 12.5 and 9.4 respectively.

The difference in C:N ratio among the clusters in this study does not appear to be significantly correlated to winter survival, a trait that is thought to be very important in the improvement of hairy vetch (Smith et al., 1987; Wagger et al., 1998). The C:N ratio appears to be most influenced by plant growth in the spring, regardless of the actual winter survival percentage of the accession. The smallest plant in the spring (Cluster II) yielded the narrowest C:N ratio, whereas the largest plants in the spring (Cluster III) yielded the widest C:N ratio.

The possible incorporation of this trait from the accessions present in Cluster II would be a positive enhancement for an improved variety of hairy vetch. Cluster II does include two accessions of *V. pannonica*. However, because of their close relationship to *V. villosa* accessions in Cluster II based on the phenotypic measurements of genetic traits, it is conceivable that *V. pannonica* accessions with favorable traits could be incorporated with *V. villosa* accessions to generate varieties with beneficial characteristics such as faster decomposition rates. It is unknown at this time if this can be accomplished in a classical breeding sense, or if genetic transformation is a more viable alternative to accomplish this goal. It is important that proper management techniques need to be incorporated as well, so that traits such as C:N ratio provide the necessary flexibility to manage efficiently N produced by hairy vetch.

## CONCLUSIONS

In conclusion, significant diversity was observed among the *Vicia* accessions. Canonical discriminant analysis is useful in computing a combined measure of variation and identifying the most influential traits associated with variation among populations. Canonical loadings of morphological and agronomic traits of each accession indicate the magnitude of genetic variation. The most influential traits in this study are fall and spring plant size and C:N ratio. Additionally, the use of cluster analysis is successful in differentiating the accessions into similar subgroups on the basis of morphological and agronomic traits. The characterization of the variation of these traits among *Vicia* accessions will permit plant breeders to focus on particular traits that are linked to accession variation, such as plant size and C:N ratio, and to discern that certain traits, such as winter survival, are not close indicators of accession variability. It is then necessary to determine how these influential traits in *Vicia* might lead to an improved hairy vetch cultivar. It is important

**Table 3. Cluster groups as determined by Proc Fastclus using Mahalanobis Distance,  $D^2$  of the centroid values of the first two calculated canonical variables. The *Vicia* plant material was evaluated under field conditions for 2 yr at Urbana, IL, from 2000 to 2002.**

Cluster group	Accession
I	PI 167259 PI 179122 PI 201882 PI 206493 PI 220880 PI 222217 PI 229970 PI 232958 PI 232959 PI 234053 PI 251004 PI 251006 PI 251679 PI 263191 PI 284104 PI 314404 PI 317447 PI 429408 PI 536642 PI 560503 PI 561946 PI 577753 Common hairy vetch
II	PI 170008† PI 515988† PI 206492 PI 393915
III	PI 383803‡ PI 493308 PI 561947
IV	PI 201883 PI 239926 PI 250796 PI 251282 PI 263190 PI 284100 PI 284102 PI 284105 PI 297999 PI 491408 PI 561945 PI 575701§
NC¶	PI 393916 PI 393922

† *V. pannonica* accessions.

‡ Crop Science Registration: Americus Hairy Vetch (Surrency et al., 1995).

§ Crop Science Registration: AU EarlyCover Hairy Vetch (Mosjidis et al., 1995).

¶ Accession does not belong to any cluster because of missing data.

to realize with winter cover crops that proper management techniques need to be incorporated as well, so that traits such as C:N ratio provide the necessary flexibility to manage efficiently N produced by hairy vetch.

### REFERENCES

- Anderberg, M.R. 1973. Cluster analysis for applications. Academic Press Inc., New York.
- Bains, K.S., and K.C. Sood. 1984. Resolution of genetic divergence for choice of parents in soybean breeding. *Crop Improv.* 11:20–24.
- Bollero, G.A., and D.G. Bullock. 1994. Cover cropping systems for the central Corn Belt. *J. Prod. Agric.* 7:55–58.
- Cruz-Castillo, J.G., S. Ganeshanandam, B.R. MacKay, G.S. Lawes, C.R.O. Lawoko, and D.J. Woolley. 1994. Applications of canonical discriminant analysis in horticultural research. *HortScience* 29: 1115–1119.
- Dillon, W.R., and M. Goldstein. 1984. Multivariate analysis methods and applications. John Wiley and Sons, New York.
- Humphreys, M.O. 1991. A genetic approach to the multivariate differentiation of perennial ryegrass (*Lolium perenne* L.) cultivars. *Hereditas* 66:437–443.
- Khattree, R., and D.N. Naik. 2000. Multivariate data reduction and discrimination with SAS software, SAS Institute Inc., Cary, NC.
- Loos, B.P. 1993. Morphological variation in *Lolium* (Poaceae) as a measure of species relationships. *Plant Syst. Evol.* 188:87–99.
- Mosjidis, J.A., C.M. Owsley, M.S. Kirkland, and K.M. Rogers. 1995. AU EarlyCover hairy vetch. *Crop Sci.* 35:1509.
- Rencher, A.C. 1992. Interpretation of canonical discriminant functions, canonical variates, and principal components. *Am. Statist.* 46:217–225.
- Riggs, T.J. 1973. The use of canonical analysis for selection within a cultivar of spring barley. *Ann. Appl. Biol.* 74:249–258.
- SAS Institute, Inc. 2000. SAS user's guide: Statistics. SAS Institute, Cary, NC.
- Smith, M.S., W.W. Frye, and J.J. Varco. 1987. Legume winter cover crops. p. 96–138. *In* Advances in soil science. Vol. 7. Springer-Verlag Inc., New York.
- Surrency, E.D., C.M. Owsley, M.S. Kirkland, D.V. McCracken, P.L. Raymer, W.L. Hargrove, J.L. Day, and J.A. Mosjidis. 1995. American hairy vetch. *Crop Sci.* 35:1222.
- Vaylay, R., and E. van Santen. 2002. Application of canonical discriminant analysis for the assessment of genetic variation in tall fescue. *Crop Sci.* 42:534–539.
- Wagger, M.G., M.L. Cabrera, and N.N. Ranells. 1998. Nitrogen and carbon cycling in relation to cover crop residue quality. *J. Soil Water Conserv.* 53:214–218.